Returns to Education among the Self-Employed: Evidence from Rural Western Uganda*

Katharina Lehmann-Uschner

*German Institute for Economic Research (DIW Berlin), Mohrenstraße 58, 10117 Berlin, klehmann@diw.de

Abstract

In many developing countries, the majority of the workforce is engaged in petty self-employment. Given the low skill intensity of this own account work and a very limited potential for selection into different sectors or types of occupation, potential returns to formal education remain unclear. This paper addresses this gap in the literature. Using a unique sample of 1’048 market vendors in Western Uganda, I provide evidence of 7 percent returns to formal education for the non-agricultural self-employed. Both allocative and productive efficiency gains as well as social capital increases are presented as potential mechanisms underlying the observed returns. Furthermore, I find no differential returns to schooling by education level. I address endogeneity by a synthetic instrumental variable approach. Finally, to avoid biased estimates through confounding factors, I use a double machine learning approach for the selection of additional control variables. Estimates on the returns to education are in line with the baseline specification.

Keywords: Returns to education, Self-employment, Economic Development

JEL: I26, J24, O15

*I greatly acknowledge valuable feedback from Rachael Meager, Fenella Carpena and Rajesh Ramachandran as well as the participants of the AEL PhD Workshop, the Nordic Conference in Development Economics, and the Development Economics Network Berlin Discussion Round. I thank the vendors in Western Uganda for participating in our study as well as all the project members at Mountains of the Moon University, Fort Portal (Uganda), especially Robert Mawenu and Oliver Schmidt. Funding for the data collection received from the Agricultural and Rural Finance Program (AGRUFIN) of the Gesellschaft für internationale Zusammenarbeit (GIZ) in Uganda is greatly acknowledged. Special thanks go to Dirk Steinwand, Julia Kirya, Esther Nanjovu, and Tim Kaiser.

To be submitted - do not cite or circulate 22nd January 2020
1. Introduction

In many developing countries, high rates of survival-driven and thus precarious self-employment lead to widespread poverty among those who work (Gindling and Newhouse, 2014; Quatraro and Vivarelli, 2014). In the absence of a market-clearing amount of adequate employment opportunities, many poor people start working on their own account (ibid.). Business performance and in turn returns from this form of necessity entrepreneurship are generally lower compared to when the entrepreneur acted voluntarily (termed opportunity entrepreneurship in the literature, see f. ex. Calderon et al., 2017). The role of formal schooling in helping the self-employed achieve higher returns in such a static and low skill intensive setting is so far unclear. The existing literature provides tremendous evidence on the returns to education. On average, these studies find the global rate of return for one additional year of schooling to be 10 percent (Psacharopoulos and Patrinos, 2004). Yet, the vast majority of this literature is centered on industrialized countries and focuses on the returns in terms of wage income, based on the Mincerian wage regression (Card, 1999). The promise of education mainly rests on the fact that with more education, people will be able to take up better-paid jobs (see for ex. Schultz, 1988). Direct empirical evidence on the returns to education in developing countries is scarce (Peet et al., 2015). In particular, there is not much evidence on the returns to education for own-account workers when wage jobs are a rare good and the opportunities to select into different sectors constrained.

This paper addresses this gap in the literature by providing evidence on the returns to education for the self-employed in a rural developing country. The focus lies on petty vendors who act in a static environment with very limited room for entrepreneurial innovation and occupational choice. Using a sample of market vendors in Western Uganda, I find that one additional year of schooling increases average daily income from market vending by 7 percent. This is comparable to what other studies have found for settings in which selection into different types of jobs plays a much larger role. To shed light on the causal mechanisms that could explain the observed returns, I combine several strands of the existing literature and focus on three aspects of the schooling-earnings relationship: (1) the selection into the vendor type (food vs. non-food), (2) classical increases in human capital in terms of both general and business-specific skills (the worker effect in Welch’s (1970) notation) and (3) social capital. I find evidence for all three mechanisms. Education increases the selection into the higher return sector, it increases the classic worker productivity through enhanced generalized knowledge as well as business specific skills and higher social capital also leads to productivity increases. In addition, financial constraints signi-
significantly hinder selection into the higher return sector. For women, education helps to alleviate these constraints. Finally, these increases in earnings seem economically meaningful as the higher income translates into higher consumption opportunities.

Furthermore, I analyze return heterogeneity by level of education. There has been some debate as to whether returns are higher for primary schooling (see e.g. Psacharopoulos and Patrinos, 2004) or secondary or tertiary education (Barouni and Broecke, 2014), sparking very different policy implications. Using a spline regression approach I find that within a sample of individuals following very homogenous occupational activities, returns to education do not differ by the level of education.

Finally, studies on the educational returns for entrepreneurs often fail to account for the endogeneity of education (Van der Sluis et al., 2005). I address the potential endogeneity of schooling choices and subsequent occupational outcomes by using the synthetic instruments method developed by Lewbel (2012). To avoid biased estimates due to a selective inclusion of covariates, I implement the debiased machine learning approach developed by Chernozhukov et al. (2018). Results confirm baseline findings.

So far, research on the effect of education on entrepreneurship is still disappointing, despite the large body of evidence on returns for wage employment (Van der Sluis et al., 2005). This paper contributes to the existing literature on the returns to schooling for the self-employed in developing countries in three important ways. First, this paper provides robust evidence for the existence of the returns to education in a static labour market setting with very limited options for occupational choice. It thus extends the existing literature in which returns to education are mainly discussed as sorting device between wage- and self-employment or as enabling individuals to profit from dynamic opportunities (see for ex. Vijverberg, 1986). In particular, Van der Sluis et al. (2005) shows that the more educated workers typically end up in wage employment. This effect is stronger for women and in least-developed countries where agriculture is more dominant. In contrast, the present paper finds evidence for significant returns to education among a group of own-account workers that entered self-employment mainly due to labour market push factors. This is remarkable given that the returns to education have been found to be larger for opportunity compared to necessity entrepreneurs (Fossen and Büttner, 2013).

Second, this paper enhances our understanding of where these returns even within a narrowly defined type of occupation - own-account market-vending - come from. Understanding these mechanisms is important to generalize findings from this pa-
This study shows that education is relevant for different aspects of self-employment. Schooling increases the actual labour productivity through both general education effects and enhanced business-specific knowledge. In addition, it increases earnings from non-agricultural self-employment as it raises the probability to select into a more profitable category or type of self-employment. Social network effects also improve returns.

Finally, this paper provides evidence for educational returns in an economically meaningful setting. Economic growth in Uganda, like other developing countries, has been largely jobless over the past years: the positive economic developments were not matched with increased employment opportunities. Petty self-employment is predicted to persist (Filmer and Fox, 2014) and therefore more research is needed on how skills could benefit the returns from this form of occupation.

The paper proceeds as follows: The next section discusses the theoretical foundations for analyzing the returns to education, followed by a presentation of the data and background information on educational achievements and employment opportunities in the region. Section 4 discusses the empirical strategy used. Section 5 presents the results, robustness tests are discussed in section 6 and section 7 concludes.

2. Theoretical Foundations

In the following, I present the theoretical foundations regarding the ways in which human capital (and thus education) affects earnings from self-employment. In particular, it is not immediately evident that education should have measurable returns in the present setting, which is static and characterized by very limited room for innovation, entrepreneurial dynamics and outside options. I therefore discuss different mechanisms that might explain the observed returns to education.

The existing literature provides different models for analyzing the returns to education. Most popular are approaches based on the human capital hypothesis (see for ex. Mincer’s two seminal papers (Mincer, 1958, 1974) or Becker, 1962) which defines human capital as an investment good that helps to raise individual productivity in the future. In particular, Mincer showed the marginal effect of schooling on wages to be the result of a compensation (for the longer time spent in school), or of a simple accounting identity. Both approaches lead to similar estimation equations. While
the vast majority of studies test the human capital hypothesis for wage employees, the same model has in a few studies also been applied in the estimation of the returns to self-employment (see for ex. Agrawal and Agrawal, 2019), using earnings in place of employee wages. This earnings-based approach permits however only limited insights into the channels and mechanisms leading to the observed earning increases.\footnote{Other criticisms of the Mincer approach are voiced by Heckman et al. (2006): In particular, given the uncertainty about future earning streams and the sequentiality with which schooling decisions are taken, the rate of return estimated here should not be misinterpreted as an internal rate of return, similar to a return on other investments. Rather, it should be seen as a growth rate of market earnings with years of schooling.}

Other studies (see for ex. Welch, 1970) have therefore followed a different approach: Based on a production function, Welch treats human capital as productive factor that not only increases the productive efficiency of a worker (the worker effect) but also the allocation of resources across sectors (the allocative effect). This allocative effect has been tested regarding the cross-sector allocation of fixed inputs (Yang and An, 2002) or of household human capital (Laszlo, 2005).

To fully capture the different mechanisms through which education might affect earnings of own-account workers, I combine the earnings and profit function based approaches to model three aspects of the education-earnings nexus. In a first step, I focus on sectoral selection, then the returns to education conditional on sectoral selection and finally the role of social networks for educational returns.

First, human capital might raise earnings by making resource allocation into different sectors more efficient (Welch, 1970; Yang and An, 2002). Adapting Welch’s model on the role of education in production to the present setting of small-scale market vendors, suppose there are two products - similar to the two sectors in Welch’s model - that market vendors can specialize in. The first refers to food and the second to non-food items. Given market restrictions, vendors can only select into one or the other product category, but not simultaneously in both. Gross sales $Q$ are then

$$Q = \max[p_1q_1, p_2q_2]$$  \hspace{1cm} (1)

with $p_1$ and $p_2$ being the sales prices (exogenous to the market vendors), $q_1$ and $q_2$ the commodities sold, which are a function of the inputs $x_1$ and $x_2$ and human capital $(HC)$: $q_i = f(x_i, HC)$ for $i = 1, 2$. The input vector $X = (x_1, x_2)$ is the capital invested to buy one product or the other. Given the characteristics of the
two products, capital requirements differ between them.\(^2\) While food can be bought in incremental amounts, units of non-food item are not infinitely separable and thus require lumpy upfront investments. Thus, the non-food sector is relatively more capital intensive than the food sector, which implies higher returns. Several studies (see for ex. Hundley, 2001; Klapper and Parker, 2010, for a review) have shown that low capital intensive industries – while requiring less upfront investments – offer lower prospective returns than the capital intensive ones, due to a lower growth and development potential.

Following Welch, the overall role of education for the earnings of market vendors can then be described by the following equation:

\[
\frac{\partial Q}{\partial HC} = \left( \max \left[ p_1 \frac{\partial q_1}{\partial x_1}, p_2 \frac{\partial q_2}{\partial x_2} \right] \right) \frac{dx_1}{dHC} + \left( \max \left[ p_1 \frac{\partial q_1}{\partial HC}, p_2 \frac{\partial q_2}{\partial HC} \right] \right)
\]

The first term represents the marginal role of education in the selection into one or the other sector, assuming that the choice to allocate capital into one or the other sector is a function of human capital (i.e. \(x_1 = x_1(HC)\)). The second term refers to education’s contribution to technical efficiency and will be analyzed in the second step. Utility is derived from the expected profits of the activity which are a function of capital and labour times their respective prices.\(^3\)

In a frictionless market, entrepreneurs should be able to borrow up to their expected profits (\(E(\pi_i) \equiv E(Q)\) when normalizing input prices and abstracting from additional costs). In turn, individual allocative efficiency and thus education should affect sectoral choice while individual physical capital should not. However, if financial frictions are important, sectoral allocation will also depend on individual access to financial means, as Paulson and Townsend (2004) have shown.\(^4\) This provides me with the following testable hypothesis:

**Hypothesis 1 (H1):** More education reduces selection into the low-return sector.

**Hypothesis 2 (H2):** Individual credit constraints are important and therefore impact sectoral choice.

\(^2\)Capital requirements only refer to the working capital required to buy inputs as there is no other productive or manufacturing activity performed.

\(^3\)In the absence of hired labour I will abstract from wage costs.

\(^4\)I abstract from the fact that education could potentially affect the access to credit. To reduce endogeneity concerns, I instrument education in the robustness section.
While the present analysis focuses on selection into two different product types within one specific entrepreneurial activity (market vending), the conclusions drawn from it readily extend to the general entrepreneurial selection when one sector is relatively more capital intensive.

In a second step, the returns to education for self-employment earnings are modeled conditional on the sectoral choice. This worker effect of education (Welch, 1970) is what most studies on the returns to education allude to: In a standard profit function framework, firm productivity (sometimes also termed technical efficiency) augments the returns to physical capital and labour and is itself a function of human capital. Similarly, standard human capital theory (as f.ex. in Mincer, 1958) predicts that education increases the marginal product of labour (and thus also returns).\(^5\)

Yet, there is so far only limited evidence on the precise mechanisms through which education increases labour productivity. Regarding the returns to education for entrepreneurs, Lazear (2004) proposes a jack-of-all-trades theory: Entrepreneurs perform a multitude of tasks. Their overall performance is therefore constrained by the weakest link in their skills.\(^6\) In turn, entrepreneurs profit more from a balanced skill set\(^7\) than specialization. This might explain the observed returns from generalized education. In addition, there are business-specific skills – financial literacy in particular – which also contribute to greater entrepreneurial productivity (see for ex. Lusardi and Mitchell, 2014). Schooling might profit the development of these skills both explicitly if directly addressed in the curriculum or implicitly through enabling individuals to better access and process new information (Rosenzweig, 1995). To evaluate the size of the worker effect of education on productivity and shed light on the mechanisms behind it, I will therefore test the following hypotheses:

**Hypothesis 3 (H3):** General education increases the earnings of own-account workers.

**Hypothesis 4 (H4):** Business-specific skills (financial literacy) increase productivity of the own-account worker and are a function of her schooling.

---

\(^5\)I will abstract from potential signalling gains from education as they have been found to matter primarily for wage workers but not so much for own-account workers (Van der Sluis et al., 2005).

\(^6\)According to Lazear (2004), the income of the entrepreneur when there are two skills \(x_1\) and \(x_2\) is \(Y_i = \lambda \times \min[x_1, x_2]\).

\(^7\)Lazear (2004) proves his model focusing on the course choice in higher education. In the setting of this paper, I transfer his idea of a weakest link to basic reading and writing skills.
Hypothesis 5 (H5): Generalized knowledge and business-specific skills are not perfect substitutes.

In a different strand of the literature, Fafchamps and Minten (2002) as well as Kolstad and Wiig (2013) discuss the importance of social networks for entrepreneurial success. In imperfect markets in particular, social relations can help reduce transaction costs and thereby increase earnings (Berrou and Combarnous, 2011). And this increased social capital can be an additional return to schooling, as longer exposure to a social group enhances both the width as well as the strength of the social connections. This results in the last hypothesis to be tested:

Hypothesis 6 (H6): Stronger social networks lead to increased returns for own-account workers. Schooling fosters this form of social capital.

A final word of caution concerns the interpretation of results. In the absence of a proper control group, returns in the form of higher earnings can be attributed to individual levels of schooling (i.e. a growth rate of earnings with years of schooling, see Heckman et al. (2006) and the discussion in footnote 1) but undoubtedly establishing causality remains difficult. The following three arguments might however support a causal interpretation of the observed returns. First, the human capital hypothesis, supported for example by Card (1999), postulates that the “cognitive skills acquired in school are an important component of human capital and the return to that capital in the labour market leads in turn to higher income” (Glewwe, 2002, p.466). Second, the mechanisms presented in this section and tested empirically in section 5, demonstrate how this transmission of education to skills to earnings might look in practice. Third, returns to education are also observable when using an instrumented measure for education to reduce endogeneity concerns.

3. Background: Education and Employment in Western Uganda and Data Description

The analysis is based on a unique sample of market vendors across the entirety of 83 rural marketplaces in the seven districts of the Rwenzori Region in Western Uganda, collected in Autumn 2015. The dataset covers all relevant permanent and regular marketplaces in the region. From each marketplace, 15-20 vendors were randomly selected, resulting in a total sample of 1’048 individuals. A simple cdf-graph of market vending income by education category shows that more years of education are associated with higher income along the entire distribution (Figure 1). The difference manifests itself in particular for the step from no to some education. The
average number of years of schooling is 5.81 and thus lower than the average at the national level, which is 10.85 for women and 11.29 for men, reflecting the rural setting of the sample. The literacy rate of 67 percent is comparable to the national average.

Figure 1: CDF for average daily market income by education category

Disaggregating education by gender and types of item sold shows three things (see Figure 2): Average education of women is lower than that of men. Females are more likely to have no education at all and the density mass is concentrated towards fewer years of education compared to men. Furthermore, there is a large overlap in the years of completed education between food and non-food vendors,\(^8\) indicating there is no prerequisite level of schooling to sell a particular item. And finally, men are more likely to engage in non-food vending than women.

Summary statistics are presented in Table 1. The sample is predominantly com-

\(^8\)An exception is female higher education. But note that this concerns only very few individuals in our sample.
Note: There are 91 male food and 102 male non-food vendors. There are 694 female food and 108 female non-food vendors.

Figure 2: Years of education by type of item sold and gender

prised of women (81 percent), reflecting the fact that women are overrepresented in the services and sales sector in Uganda (Uganda Bureau of Statistics, 2018). The average age is 36 years. The market vending activities are crucial to the household's income: 70 percent of the respondents report being the main contributor to household income and market income constitutes on average 90 percent of total household income. The average household respondents live in counts 6.5 people and respondents support on average 4.2 children.

Despite the sample being representative for the population of market vendors only, their employment characteristics are comparable to the average rural household in Uganda. Among our sample, only 1 percent of the respondents earned additional wage income. A high prevalence of self-employment is typical for many developing country economies, particularly in rural areas. Only one in ten smallholder households have occasional wage earning jobs (Anderson et al., 2016) with the overall self-employment rate among the active workforce in rural Uganda being 70 percent (Uganda Bureau of Statistics, 2018). Lack of a wage job is by far the most import-
Table 1: Descriptive statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Education and experience</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td>5.81</td>
<td>3.7</td>
<td>0</td>
<td>17</td>
<td>1042</td>
</tr>
<tr>
<td>Able to read and write</td>
<td>0.67</td>
<td>0.47</td>
<td>0</td>
<td>1</td>
<td>1008</td>
</tr>
<tr>
<td>Numeracy</td>
<td>-0.02</td>
<td>0.83</td>
<td>-1.62</td>
<td>0.94</td>
<td>1048</td>
</tr>
<tr>
<td>Financial literacy</td>
<td>-0.01</td>
<td>0.78</td>
<td>-1.3</td>
<td>1.26</td>
<td>1048</td>
</tr>
<tr>
<td>Years of experience (market vending)</td>
<td>7.2</td>
<td>7.14</td>
<td>0</td>
<td>52</td>
<td>1030</td>
</tr>
<tr>
<td><em>Market vending</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. daily income from market vending activities (UGX)</td>
<td>8092.63</td>
<td>14186.04</td>
<td>0</td>
<td>100000</td>
<td>1048</td>
</tr>
<tr>
<td>Item sold = food (fresh or cooked)</td>
<td>0.78</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
<td>1017</td>
</tr>
<tr>
<td>Ever received fin lit training</td>
<td>0.23</td>
<td>0.42</td>
<td>0</td>
<td>1</td>
<td>1046</td>
</tr>
<tr>
<td>Keeps business log</td>
<td>0.3</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
<td>1032</td>
</tr>
<tr>
<td>Business formally registered</td>
<td>0.25</td>
<td>0.43</td>
<td>0</td>
<td>1</td>
<td>1017</td>
</tr>
<tr>
<td>Respondent pays business taxes</td>
<td>0.69</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
<td>1018</td>
</tr>
<tr>
<td>&gt;= one job (past 3 months)</td>
<td>0.21</td>
<td>0.40</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Subsistence or commercial farming</td>
<td>0.15</td>
<td>0.36</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Share of market income in total income</td>
<td>0.9</td>
<td>0.23</td>
<td>0.01</td>
<td>1</td>
<td>1046</td>
</tr>
<tr>
<td>Vending in other markets as well</td>
<td>0.25</td>
<td>0.43</td>
<td>0</td>
<td>1</td>
<td>916</td>
</tr>
<tr>
<td>Dist to market &gt; 30min on foot</td>
<td>0.15</td>
<td>0.36</td>
<td>0</td>
<td>1</td>
<td>1015</td>
</tr>
<tr>
<td><em>Financial constraints, wealth and social capital</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial constraints</td>
<td>0.57</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>1028</td>
</tr>
<tr>
<td>Total number of assets owned</td>
<td>46.4</td>
<td>27.35</td>
<td>0</td>
<td>164</td>
<td>943</td>
</tr>
<tr>
<td>Total current savings (‘000 UGX)</td>
<td>416.65</td>
<td>1055.81</td>
<td>0</td>
<td>20000</td>
<td>1042</td>
</tr>
<tr>
<td>Business investments (‘000 UGX, past 6 months)</td>
<td>413.27</td>
<td>788.54</td>
<td>0</td>
<td>5000</td>
<td>958</td>
</tr>
<tr>
<td>Relative wealth (standardized)</td>
<td>0</td>
<td>1.02</td>
<td>-2.13</td>
<td>3.78</td>
<td>1040</td>
</tr>
<tr>
<td>Main material of wall: burnt/unburnt bricks or cement</td>
<td>0.46</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>1035</td>
</tr>
<tr>
<td>No. of community groups</td>
<td>1.25</td>
<td>1.18</td>
<td>0</td>
<td>8</td>
<td>1040</td>
</tr>
<tr>
<td>Some friends are vendors</td>
<td>0.9</td>
<td>0.3</td>
<td>0</td>
<td>1</td>
<td>1029</td>
</tr>
<tr>
<td><em>Individual and household characteristics</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH size (count)</td>
<td>6.5</td>
<td>3.97</td>
<td>0</td>
<td>34</td>
<td>1044</td>
</tr>
<tr>
<td>Number of children respondent supports</td>
<td>4.2</td>
<td>2.99</td>
<td>0</td>
<td>28</td>
<td>1036</td>
</tr>
<tr>
<td>Respondent contributes the most to HH income</td>
<td>0.69</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
<td>1047</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>35.94</td>
<td>11.77</td>
<td>12</td>
<td>78</td>
<td>1037</td>
</tr>
<tr>
<td>Female</td>
<td>0.81</td>
<td>0.4</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Risk aversion (standardized)</td>
<td>0.04</td>
<td>1.05</td>
<td>-1.49</td>
<td>2.64</td>
<td>1020</td>
</tr>
<tr>
<td>Patience (standardized)</td>
<td>0.01</td>
<td>1</td>
<td>-1.4</td>
<td>0.99</td>
<td>1038</td>
</tr>
<tr>
<td><em>District</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyeggoa</td>
<td>0.11</td>
<td>0.31</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Kyenjojo</td>
<td>0.13</td>
<td>0.34</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Kamwenge</td>
<td>0.12</td>
<td>0.33</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Kasese</td>
<td>0.11</td>
<td>0.31</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Kabarole</td>
<td>0.19</td>
<td>0.39</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Bundibugyo</td>
<td>0.29</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
<tr>
<td>Ntoroko</td>
<td>0.05</td>
<td>0.22</td>
<td>0</td>
<td>1</td>
<td>1048</td>
</tr>
</tbody>
</table>
ant reason for being self-employed, according to the National Labor Force Survey of the Ugandan Bureau of Statistics. Furthermore, the total median monthly income in rural Uganda is 132,000 UGX for males and 88,000 UGX for females (ibid.). This corresponds well to the median daily income from market vending activities of approximately 3’300 UGX we find in our sample.

4. Empirical Strategy

To estimate the role of education for the selection into selling the low-return product, I estimate the following regression using a standard probit model:

$$\text{food}_i = \alpha_0 + \beta_1 \text{schooling}_i + \beta_2 \text{experience}_i + \beta_3 \text{experience}_i^2 + \beta_4 \text{female}_i + \beta_5 \text{creditconstraint}_i + \beta_6 \text{female}_i \ast \text{creditconstraint}_i + \beta_7 X + \epsilon_i \quad (3)$$

where the choice of covariates in the baseline specification follows the standard Min-
cerian wage regression model with additional controls. I measure schooling as years of completed education following standard practice in the literature. More experienced vendors might be better able to observe the higher returns in the non-food sector. experience therefore measures the number of years the individual has been working as a market vendor. Similarly, the individual’s gender might influence both push and pull factors related to the type of item sold (access to inputs or family responsibilities for example). In addition, financial constraints might play an important role in shaping entrepreneurial activity, as Paulson and Townsend (2004) have shown. The creditconstraint variable therefore measures whether the individual would not be able to obtain a sum of 1 Million UGX (equal to 300 USD in 2015) if they needed to. In addition, household wealth - which I proxy by the standardized value of the household’s total asset holdings - might affect sectoral selection for credit-constrained individuals and is therefore included as additional control. To capture heterogeneous effects of financial constraints by gender, I also interact the constraints variable with gender. Finally, the vector $X_i$ contains additional covariates. For example, a larger household size might provide the market vendor with a larger labour force, potentially supporting the selection into the labour-intense but lower-return sector.

---

9 An exception are tertiary institutes and university, which are counted as one year each.
10 Recall that returns in the capital-intensive industry are typically higher compared to the less capital-intensive industry, see the discussion in section 2.
11 As a robustness test, I use a more restrictive measure in which I classify all individuals as being credit constrained who are unable to access 100’000 UGX (30USD) in case of an emergency.
To ensure results do not depend on the particular measure of education employed, I categorize education into primary and secondary or tertiary education and implement a spline regression approach (see for ex. Kazianga, 2006). Furthermore, current credit constraints might be endogeneous to the stream of past realized earnings, which might bias estimated coefficients. I therefore instrument whether the individual would be able to access 1 million UGX with the household’s land holdings. Land has been shown to serve as collateral for obtaining a loan (Kolstad and Wiig, 2015). And land markets are typically static in many developing countries in Africa with land titles being inherited from one generation to the next (ibid.). Land holdings are thus unlikely to respond easily to changes in household income.  

For the estimation of the returns to education in terms of productive efficiency (i.e. the worker effect) conditional on the selection into the category of item sold, I start with the Mincerian wage regression (Mincer, 1958, 1974). Despite several criticisms regarding mainly the interpretation of estimated coefficients (see for ex. Heckman et al., 2006 or Iversen et al., 2010), Mincer type specifications are still standard for estimating returns to education for both wage labourers as well as self-employed (see for example Van der Sluis et al., 2005, 2008).

\[
\begin{align*}
\log(earnings_i) = & \alpha_0 + \beta_1 schooling_i + \beta_2 experience_i + \beta_3 experience_i^2 + \beta_4 food \\
& + \beta_5 female + \beta_6 X + \epsilon_i
\end{align*}
\]

(4)

Entrepreneurial performance can be measured in different ways, for example as earnings, profits, survival and firm growth (for a discussion, see Van der Sluis et al., 2008). The majority of the existing literature has focused on earnings. I therefore also use the average daily market earnings as outcome variable. This is the total income from market vending activities reported for the past 3 months divided by the number of

\[\text{One could however also perceive a direct effect of land holdings on the type of item sold which would render land holdings to be invalid as instrument. In particular, if own land holdings make the market vendor more likely to pursue agricultural activities and if this results in agricultural products to be sold on the market, land holdings might no longer be exogenous to the selection into the type of item sold. In a regression of vending food items on land holdings, I do however not find any significant effect. As a further robustness test, I estimate whether pursuing any agricultural activities in addition to market vending increases the likelihood to sell food items as opposed to non-food items. I find no significant effect in the full sample, but pursuing agricultural activities decreases (and not increases) the likelihood to be a food vendor for the female subsample. Alternative income could therefore also help to reduce credit constraints. Put together, these findings support the claim that land holdings are linked to alleviating credit constraints but do not directly affect the choice of item sold.}\]
days for which the respondent reported doing these activities.\(^\text{13}\) To correct potential reporting errors resulting in extreme values of average daily market income, I winsorize income data at the 99th percentile. Furthermore, following standard practice in the returns to education literature, I log-transform individual earnings. This eases interpretation of the estimated coefficients and allows to account for concavities or decreasing returns in the schooling-earnings relationship. Potentially heterogeneous effects of education by gender are addressed by including interaction terms.

The role of general education for increased earnings could manifest itself not only through a higher capacity to access and process new information (captured by the general measure of education, see for ex. Rosenzweig, 1995) but also through improved basic numeracy or literacy. I measure numeracy based on five questions of varying difficulty evaluating the respondent’s mathematical ability. I then use Item Response Theory (IRT) to aggregate these measures into a reliable scale of the latent trait numeracy. Instead of summing correct responses only (see for ex. Carpena et al., 2011), the IRT approach takes a question’s difficulty and discriminatory power into account (see for ex. Rasch, 1960 or Lord, 2012 for a discussion, Kaiser and Menkhoff, 2018 present an application). Following Kaiser and Menkhoff (2018), I employ the two-parameter logistic model (see for example Birnbaum, 1968), which is widely used in the construction of psychological measures. Literacy is a dummy for whether the individual is able to read and write. Furthermore, I interact the numeracy and literacy measures to investigate to what Lazear’s jack-of-all-trades theory (Lazear, 2004) applies also to market vendors. In particular, a significant interaction term between reading and writing will indicate that vendors are constrained by the weakest link in their skill set.

Schooling might also help develop a better understanding of economic interrelations and skills necessary to become more productive. Financial literacy might be the most important aspect of this business-specific knowledge. I measure financial literacy based on six questions eliciting the respondent’s financial knowledge, which have been widely used in the literature (see for example Lusardi and Mitchell, 2014, p.10). Similar to the construction of the latent numeracy trait discussed before, I then apply Item Response Theory to construct a measure of the market vendor’s financial literacy. To account for actual financial behavior, I include savings and business investments as a robustness test. This should capture the extent to which financial knowledge already translates into good financial behavior. Savings are

\(^{13}\)I thus abstract from the decision of how much to work.
measured by total current savings (summed over all saving locations) and investments as total reported business investments over the past 6 months. To explore to what extent generalized education and business-specific knowledge are substitutes or complements (Schultz, 1988), I compare the estimated coefficients from a regression in which all measures have been included jointly to the estimates from regressions in which the measures have been included separately.

To evaluate the role of social networks in increased earnings from self-employment (through for example a larger pool of potential customers or access to better or more information), I additionally control in equation 4 for the number of community groups an individual is member of and for whether she has friends who are also market vendors.

To ensure these proposed mechanisms are indeed relevant for the observed returns to education, I estimate in separate regressions the effect of education on the selection of items sold, literacy, numeracy, financial literacy and social networks.

In addition, Iversen et al. (2010) show in the context of the Danish labour market that returns are heterogeneous and non-linear. I therefore also explore whether there are differential returns to education based on the type of school in which this additional year of education was completed, employing a spline regression approach (see for example Kazianga, 2006) and by categorizing education into completed primary, secondary and tertiary education.

A widely discussed problem in properly estimating the returns to education is endogeneity: people might select into different sectors entailing different wages for (unobservable) reasons that also affect their educational choices. In terms of this paper, individual labour productivity and thus earnings as a market vendor might be affected by unobservable characteristics such as grit, ability or motivation that also influence the schooling decision. This paper addresses endogeneity concerns in two ways. First, I analyze returns to education within a homogenous sample, enabling me to abstract from concerns relating to occupational or sectoral choice. Second, I use an instrumental variable regression approach that relies on synthetic instruments as developed by Lewbel (2012). In addition to the heteroskedasticity-based instruments this estimator constructs from within the system, this approach allows for additional external instruments. I am therefore also using Uganda’s universal primary education reform from 1996 as additional instrument.
Finally, there might be other factors affecting self-employment outcomes. Omitting these aspects from the analysis might lead to biased estimates, in particular if these factors are linked to educational outcomes. I therefore pay particular attention to accounting for potential confounders to be included into the control vector $X$. Existing research on the determinants of entrepreneurship selection and performance for example shows that self-employment outcomes are a function of risk attitude, access to capital, labor market experience, business acumen, family background, psychological traits and education (see for ex. Le, 1999; Van der Sluis et al., 2008, both focusing on developed economies though). First, I include two measures of personality traits (risk aversion and patience) in the analysis. Patience is measured by the standardized z-score of the answers to the question how willing respondents would be to take a 30 minute walk instead of taking a taxi for that move (4-point scale). And risk attitudes are measured by the standardized z-score of a widely used non-incentivized survey item asking for the participant’s willingness to take risks on a 0 to 10 scale (see Dohmen et al., 2011). Second, I use the debiased (or double) machine learning approach proposed by Chernozhukov et al. (2018) to select additional confounders. This method partials out the effects of additional $x$’s from the estimator of interest (education). To best reduce the impact of covariate selection on the estimated education effect, this approach uses the least absolute shrinkage or selection operator (LASSO) for covariate selection in both the regression of market income as well as of education on the control variables. In a next step, it applies sample splitting as well as cross-validation to the partialling-out estimator.\footnote{For example, to control for a potential “necessity” channel according to which respondents with a larger dependency share in the household need to earn more, I include household size, whether the respondent is the main contributor to household income and the number of children supported by the respondent. In addition, a “formalization” dividend could also affect market vending income. I therefore also include whether the respondent pays business taxes as covariate. And additional income sources such as farming might be linked to the choice of item sold and could affect earnings. Farming is therefore also included as covariate. The full list of covariates included in the LASSO estimation as well as in the partialling-out is presented in the appendix.}

5. Results

5.1. Allocative efficiency - Selection into type of item sold

The first part of the analysis focuses on the role of human capital for increasing the allocative efficiency of individual market vendors. Results from the estimation of equation 3 with a probit regression controlling for district fixed effects and with standard errors clustered at the market level show that education indeed significantly
affects the choice of item sold (Table 2). In particular, schooling is associated with less selection into the labour-intensive but lower return sector. With each additional year of schooling, the likelihood of selling food instead of non-food products declines by 1 percentage point (column 1). Results from a regression that uses education categories instead of continuous years shows a similar picture: Individuals who have attended primary school are 8 percentage points less likely to be food vendors, for those with secondary or higher education, the probability of vending food decreases by 11 percentage points (column 4).\textsuperscript{15}

In addition to human capital, there is an important gender dimension to the sectoral allocation. Women are 25 percentage points more likely to be active in the lower-return segment of food vending instead of the higher-return non-food vending. The efficiency-enhancing effect of education seems however not to differ between the sexes: Interaction terms between the education and the gender of the market vendor remain insignificant (Table 2, columns 3 and 5). Several reasons for these gender differences in allocative efficiency are conceivable. While the data do not allow to test for the potential existence of norms or taste-based selection into food vending, financial constraints are expected to affect the choice of item sold and are likely to differ by gender. As discussed in the theoretical foundations before, vending non-food items yields higher returns but also requires larger upfront investments. Being unable to obtain a credit therefore constitutes a significant barrier to selecting into selling this type of item. And the risk of being credit constrained is 61 percent for a women and thus 1.5 times higher than for a men. In a next step, I therefore also control for reported financial constraints of the individual (Table 2, column 2). Indeed, facing credit constraints makes individuals 6 percentage points more likely to enter the low-return item category. When interacting gender with credit constraints (Table 2, column 5), I find that this result is primarily driven by female market vendors.

To further explore the extent to which human capital and financial constraints affect the individual’s allocative efficiency, I include a triple interaction between gender, credit constraints and education into the estimation of the sectoral allocation. Predictive margins at different combinations of the vendor’s gender, schooling and reported credit constraints are presented in Figure 3. Results suggest a double role of education: First, confirming baseline findings, education reduces the selection

\textsuperscript{15}Note that education patterns between food and non-food vendors are very similar (see Figure 2).
Table 2: Selection into type of item sold - Baseline specification

<table>
<thead>
<tr>
<th></th>
<th>Outcome: Type of item sold = food</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Years of education</td>
<td>-0.01***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>Highest educ level: Some or completed primary</td>
<td>-0.08**</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
</tr>
<tr>
<td>Highest educ level: &gt;= secondary</td>
<td>-0.11**</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
</tr>
<tr>
<td>female</td>
<td>0.26***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
</tr>
<tr>
<td>educ yrs.*female</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>Primary educ*female</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Sec or tert educ*female</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of experience (market vending)</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>Experience squared</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>Credit constraints</td>
<td>0.06*</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
</tr>
<tr>
<td>Credit constraints*female</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>HH size</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>Observations</td>
<td>995</td>
</tr>
<tr>
<td>Credit constr. for females</td>
<td>0.06**</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
</tr>
<tr>
<td>Yrs. of primary educ for females</td>
<td>-0.07*</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
</tr>
<tr>
<td>Yrs. of sec. or tert. educ for females</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
</tr>
<tr>
<td>Educ for females</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
</tbody>
</table>

Table reports marginal effects. Standard errors (clustered at the market level) in parentheses; District FE included; *** p<0.01, ** p<0.05, * p<0.1

into vending the low-return product type and therefore increases the allocative efficiency of the market vendor. Second, for women a higher level of education seems to alleviate the constraining role played by a restricted access to credit. Women without education have a predictive margin of 0.96 to be a food vendor as opposed to vending non-food items when they are credit-constrained whereas those who do not face credit constraints have a predictive margin of 0.83. At secondary or higher education, the predictive margin to be a food vendor for credit-constrained women drops to 0.84 whereas those of the non-credit-constrained women remains unaffected. It is important to note however that the interaction terms are only significant in the
regression for the female subsample (see Table 3, column 4) but not in the full sample. I therefore present these results only as suggestive evidence of the double role of education. More research is needed in this respect.

Table 3 presents robustness tests. First, I use individual land holdings as instrument for financial constraints to address a potential endogeneity of reported credit constraints to previous income streams (Table 3, column 1). To account for the fact that credit constraints matter primarily for females, this regression is estimated for the female subsample only. Furthermore, I additionally control for household wealth - proxied by the standardized value for the total number of assets owned by the household - to fully capture the effect of access to financial means (column 2 and 3). Results confirm the baseline finding that financial constraints diminish selection into the higher-return non-food vending activities.

Overall, the data provide support for hypotheses 1 and 2: Human capital increases the allocative efficiency of market vendors. Individuals with a higher level of
are more likely to invest into vending the higher-return non-food products instead of the lower return food products. This might be driven by enhanced capacities to access and process new information (Rosenzweig, 1995). At the same time, financial constraints restrict selection into the high-return sector, but this restricting effect loses importance with higher levels of education.

5.2. Returns to education conditional on sectoral selection (worker effect) - Baseline specification

Results provide evidence for significant returns to education: One year of additional completed education is associated with increases in average daily income of 7 percent in the baseline regression (Table 4). This is comparable to the global average rate of return for wage employees (Psacharopoulos and Patrinos, 2004; Ashenfelter et al., 1999) but slightly lower than educational returns found for developing economies – in particular Sub-Saharan African countries – where average rates of return are closer to 12 percent per year (ibid.). A comparison with results from the entrepreneurship literature shows that these baseline returns are in line with those found for opportunity entrepreneurs in a developed economy (Fossen and Büttner, 2013). They are larger by 2 percentage points for average returns to entrepreneurs found in a meta-study for developed economies (Van der Sluis et al., 2008) and larger by 3 percentage points than those found for necessity entrepreneurs (Fossen and Büttner, 2013).

An additional year of experience as vendor on the market increases income by 4 percent. Additionally controlling for age (results not shown) leaves results unaffected: The coefficient on age is very close to zero and insignificant and point estimates for experience and schooling are not significantly different from the baseline specification. Furthermore, average daily income is substantially lower for women compared to men (by 46 percent). In line with other studies that provide evidence for higher returns to education for women than men (Van der Sluis et al., 2008), I find that the returns to primary and secondary education in the spline regression specification accrue to females only (Table 4 columns 3 and 5). The type of item sold significantly affects income from vending activities, as already discussed in the theoretical foundations section. Compared to non-food vendors, the income of food vendors

\[ \text{I hypothesize this could be driven by the lower number of hours worked per day by women due to family obligations. Unfortunately, the data at hand do not allow to test this hypothesis.} \]

\[ \text{Yet, when using the continuous education measure, the education-income nexus is unaffected by gender, as can be seen by the insignificant interaction terms between gender and schooling (Table 4 column 4).} \]
is 41 percent lower (Table 4, column 1). Part of this effect stems from the gender differences in sectoral selection: When additionally controlling for the vendor’s sex, the effect of the item choice on income drops to 24 percent (Table 4, column 2).

In the present setting, the returns to one additional year of completed education do not differ between primary, secondary and tertiary education. In a spline regression following Kazianga (2006), point estimates for the effect of an additional year of schooling on income do not differ significantly between different education levels (Table 4, column 3).\(^{18}\) This has two implications. First, even for occupations such as market vending that do not seem very skill intensive, pursuing schooling up to secondary or higher education has positive returns.\(^{19}\) Second, the differences in returns by levels of education found for example by Psacharopoulos and Patrinos (2004) and Barouni and Broecke (2014) might be context specific, depending on the schooling system or earning environment considered.

5.3. Generalized education vs. business-specific skills

The next section explores the mechanisms behind the observed productivity increasing effect of education once sectoral choice is accounted for. First, the productivity of the market vendor might grow due to enhanced business-specific skills as a result of more education. For market vendors, such knowledge salient to good business practice could be financial literacy. As expected, I find that financial literacy is associated with increases in income (Table 5, column 1). A one standard-deviation increase in financial literacy can be attributed to an income increase of 18 percent. To evaluate whether business-specific knowledge translates into “good financial behaviour”, I use business investments and total savings as proxy for business-specific knowledge as a robustness test (Table 8, column 3). Again, baseline results are confirmed. A 10 percent increase in business investments is associated with a 1.1 percent higher income from market vending activities and increasing total savings by 10 percent with an income increase of 0.3 percent.\(^{20}\) The point estimate for the effect of education reduces in size but still remains significant.

---

\(^{18}\)Using an alternative spline regression specification \((\text{educyears} + (\text{educyears} - 7) \times 1(\text{educyears} \geq 7) + (\text{educyears} - 13) \times 1(\text{educyears} \geq 13))\) as shown in Greene (2000, p.322), I fail to reject the null hypothesis that the slope of the function is constant (i.e. the coefficients on additional years of secondary and on tertiary education are jointly zero).

\(^{19}\)Note that positive returns to additional years of schooling in secondary or higher education are also observable among the subsample of food vendors only.

\(^{20}\)As there are other factors potentially influencing business investments and savings, I am cautious to not claim a causal relationship here.
Second, the generalized knowledge obtained through schooling also seems to profit the self-employed as it improves their capacity to access and process new information (Rosenzweig, 1995). This generalized knowledge could be measured as the residual returns to education once business-specific knowledge is accounted for. Indeed, the dividend to this general knowledge seems to materialize in addition to the returns on business-specific knowledge (financial literacy). The point estimate for years of schooling reduces only by 1 percentage point when a measure of financial literacy is included (Table 5, columns 1 and 5). In the reasoning of Schultz (1988), generalized knowledge is therefore not a perfect substitute for the business-specific skills.

One could think of an alternative explanation for the relative stability of the education coefficient even when a measure of (financial) ability is included. In a wage employment setting, this finding would provide support for the sorting hypothesis: Firms choose their workers based on observables (their education level) and not unobservables (their ability) (see Weiss, 1995). In self-employment however, income and productivity are much closer linked and sorting or sheepskin effects should play a much smaller role (see Glewwe, 2002). I therefore use numeracy and literacy as more specific measures of generalized knowledge (Table 5, column 3-5). While the level effects of numeracy and literacy are not significant when included in addition to the general schooling measure\textsuperscript{21} (Table 5, column 3), the interaction term between numeracy and literacy is significant. This suggests basic skills (such as numeracy and literacy) are not conducive to increased earnings when only one of them is present. This becomes even more apparent when numeracy and literacy are used as proxy for generalized education and are thus included instead of the education measure (and not in addition to it, Table 5 column 4 and 5). For very low levels of numeracy, I do not find any significant effect of being able to read and write on market income. Yet, once numeracy gets large enough (i.e. once a market vendor has a numeracy level above the 25\textsuperscript{th} percentile), literacy significantly increases earnings. And this effect gets larger with increasing numeracy: At a numeracy level above the 75\textsuperscript{th} percentile, literacy leads to income increases of 45 percent. The same holds true for numeracy. While there is no significant income effect of numeracy for those unable to read and write, a one standard-deviation increase in numeracy increases income for the literate by 12 percent. Additionally controlling for business-specific skills in the form of financial literacy does not alter the results (Table 5, column 5).

\textsuperscript{21}Note however that the effect of years of education remains highly significant in this regression. This suggests, the effect of schooling encompasses more than the acquisition of basic numeracy and literacy skills.
Both generalized knowledge as well as business-specific skills lead to an increased income of market vendors, which shows that the two measures pick up different aspects of the education-earnings nexus. This provides supportive evidence for Lazear’s jack-of-all-trades theory (Lazear, 2004): Market vendors perform a multitude of tasks and therefore profit from a balanced skill set. As evidenced by the positive and significant interaction term between literacy and numeracy, a skill will only deploy its true potential when the self-employed disposes of complementing skills at a matching level.

5.4. Social capital

Finally, productivity increases might accrue due to an enlarged social network. The social network, as measured by the number of community groups an individual is a member of, is associated with significant increases in income (Table 5, columns 6 and 7). Being active in one community group leads the market vendor to receive an income higher by 28 percent compared to those not engaged in any community groups and the income rise amounts to 72 percent if the market vendor is active in more than three community groups. Intuitively, a larger social network could improve the market vendor’s access to information or enlarge the vendor’s client base. In addition, the effect of social capital does not differ by the achieved education level. Interaction terms between years of schooling and community group activities remain insignificant (Table 5, column 7). Furthermore, the network effect does not seem to work through knowing more people on the market: whether or not any friends or relatives are market vendors as well has no effect on income (Table 5, column 8).

Overall, results provide evidence for hypotheses 3 - 6. Both generalized knowledge and business-specific skills increase productivity of the self-employed and thus lead to higher income once selection into the type of item sold is accounted for. In addition, enhanced social capital also leads to increases in income for the self-employed. To what extent literacy, numeracy, financial literacy as well as community group engagement are causally linked to schooling is explored further down in section 6.1 (instrumenting education).

Furthermore, the different skill sets are not perfect substitutes for each other. When controlling for business-specific knowledge or the vendor’s social capital, the point estimate on the years of schooling reduces by 1 to 2 percentage points but still remains highly significant. The resulting returns to education are still comparable to the average returns for entrepreneurs found by Van der Sluis et al. (2008).
This provides further supportive evidence for the role of education in enhancing individual generalized knowledge, when defined as this residual return to schooling once other mechanisms are accounted for. Improving individual capacities to access and process new information (Rosenzweig, 1995) might lead to productivity increases for example through better procurement strategies, improved (business) planning or different (more successful) vending strategies. Given the type of market vending activities analyzed, business investments as discussed in the literature on entrepreneurial returns to education are not widely applicable in the setting of this study.

Finally, my findings suggest that even in a setting in which formal wage earning jobs are a rare good, there are returns to education that are in line with results found for environments that allow for selection. Yet, the exact magnitude of returns, in particular also at different educational levels, might depend on specific circumstances and the type of self-employment. For example, large business owners who also employ workers are expected to profit more from tertiary education.

6. Robustness Tests

The estimated positive returns to education do not depend on the specific measure of education used. Returns to education are robust to measuring education in an alternative way (categorize education into primary, secondary and tertiary education). Estimated effects are in line with the number of years per primary and secondary education. Note that it does not make a difference whether the individual just completed some or all years of primary or secondary education. In a separate regression (results not shown), in which some primary/secondary and completed primary/secondary education are defined as separate categories, there is no statistically significant difference between the coefficients on some and completed education of one school type. Returns to education are also unaffected by previous financial literacy trainings, even if having received financial literacy training increases income by 30 percent.

6.1. Instrumenting education

Endogeneity of education is mostly discussed in terms of an ability bias (although Ashenfelter et al. 1999 find it to be small). In our sample, the socioeconomic situation of the respondent’s parents might have a significant impact on schooling decisions of the individual and occupational outcomes in terms of the financial means disposable for investment as well as the choice of item sold. The parental socioeconomic background is however unobservable here. To make sure endogeneity is not the driver
of the results, I employ a novel approach to estimating endogenous regressor models proposed by Lewbel (2012). This method takes the form of a modified 2sls estimator that uses heteroscedasticity in the errors of a linear projection of the endogeneous variable on all other regressors (like the first stage in a traditional IV approach).\textsuperscript{22} Identification then stems from having some regressors that are uncorrelated with the product of these heteroskedastic errors. In addition, the approach also allows for the inclusion of external instruments to improve efficiency. I use Uganda’s universal primary education reform from 1996. Specifically, the additional instrument used is whether the individual could have profited from at least one year of free primary education under the UPE reform. Results show that the positive effect of education on income remains under instrumentation and is thus not driven by unobserved factors (Table 8 column 5).

\textsuperscript{22}The chi-squared test statistic for this linear projection in our baseline model is 160.79 (or 316.51 when cluster fixed effects are included). I thus reject the null of constant error variance.
Table 3: Selection into type of item sold - Robustness tests

<table>
<thead>
<tr>
<th></th>
<th>Outcome: Type of item sold = food</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>IV</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>Highest educ level: Some/completed primary</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
</tr>
<tr>
<td>Female</td>
<td>0.23***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
</tr>
<tr>
<td>Educ yrs.*female</td>
<td>0.05*</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
</tr>
<tr>
<td>Experience squared</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>Credit constraints</td>
<td>0.05*</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
</tr>
<tr>
<td>Primary educ*credit constraints</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
</tr>
<tr>
<td>Secondary educ*credit constraints</td>
<td>-0.16*</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
</tr>
<tr>
<td>Total no. of assets owned (std.)</td>
<td>-0.03*</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
</tr>
<tr>
<td>Assets*female</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>HH size</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>Assetholdings for females</td>
<td>-0.04**</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
</tr>
<tr>
<td>Educ for females</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>Observations</td>
<td>775</td>
</tr>
</tbody>
</table>

Table reports marginal effects. Standard errors (clustered at the market level) in parentheses; District FE included; *** p<0.01, ** p<0.05, * p<0.1
In columns 1 and 5, the model has been estimated for the subsample of female market vendors only.
In column 1, credit constraints have been instrumented with access to land; standard errors for this regression are obtained through bootstrapping with 50 replications.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average daily income (log)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td>0.07***</td>
<td>0.06***</td>
<td>0.06***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yrs of primary educ</td>
<td>0.05***</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yrs of secondary educ</td>
<td>0.05***</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yrs of tertiary educ</td>
<td>0.07***</td>
<td>0.06***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.021)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>-0.47***</td>
<td>-0.46**</td>
<td>-0.46**</td>
<td>-0.70***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.185)</td>
<td>(0.257)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of experience</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.04***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Experience squared</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Item sold = food</td>
<td>-0.41***</td>
<td>-0.26***</td>
<td>-0.26***</td>
<td>-0.26***</td>
<td>-0.26***</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.086)</td>
<td>(0.087)</td>
<td>(0.087)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>Female*primary years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*secondary years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*tertiary years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*education (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>7.99***</td>
<td>8.35***</td>
<td>8.38***</td>
<td>8.34***</td>
<td>8.60***</td>
</tr>
<tr>
<td></td>
<td>(0.177)</td>
<td>(0.185)</td>
<td>(0.200)</td>
<td>(0.223)</td>
<td>(0.307)</td>
</tr>
<tr>
<td>Observations</td>
<td>995</td>
<td>995</td>
<td>995</td>
<td>995</td>
<td>995</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.087</td>
<td>0.104</td>
<td>0.105</td>
<td>0.104</td>
<td>0.106</td>
</tr>
<tr>
<td>Yrs of primary educ &amp; female = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yrs of secondary educ &amp; female = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yrs of tert educ &amp; female = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educ (years) &amp; female = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary - secondary educ (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary - tertiary educ (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors (clustered at the market level) in parentheses; District FE included; *** p<0.01, ** p<0.05, * p<0.1
Table 5: Returns to education: Worker effect - Business specific knowledge

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average daily income (log)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td>0.05***</td>
<td>0.05***</td>
<td>0.05***</td>
<td>0.05***</td>
<td>0.06***</td>
<td>0.06***</td>
<td>0.06***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.013)</td>
<td>(0.010)</td>
<td>(0.018)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Literacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.27***</td>
<td>0.24***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.101)</td>
<td>(0.085)</td>
<td>(0.085)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numeracy</td>
<td></td>
<td>-0.08</td>
<td>-0.08</td>
<td>-0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.074)</td>
<td>(0.076)</td>
<td>(0.077)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literacy*Numeracy</td>
<td></td>
<td>0.17**</td>
<td>0.19**</td>
<td>0.19**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.086)</td>
<td>(0.086)</td>
<td>(0.086)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of experience</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.03**</td>
<td>0.03**</td>
<td>0.03**</td>
<td>0.04***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Experience squared</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00*</td>
<td>-0.00*</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Item sold = food</td>
<td>-0.26***</td>
<td>-0.26***</td>
<td>-0.28***</td>
<td>-0.30***</td>
<td>-0.29***</td>
<td>-0.28***</td>
<td>-0.28***</td>
<td>-0.26***</td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
<td>(0.088)</td>
<td>(0.091)</td>
<td>(0.091)</td>
<td>(0.091)</td>
<td>(0.087)</td>
<td>(0.088)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.44***</td>
<td>-0.44***</td>
<td>-0.43***</td>
<td>-0.51***</td>
<td>-0.48***</td>
<td>-0.43***</td>
<td>-0.43***</td>
<td>-0.47***</td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.103)</td>
<td>(0.105)</td>
<td>(0.105)</td>
<td>(0.109)</td>
<td>(0.108)</td>
<td>(0.108)</td>
<td>(0.100)</td>
</tr>
<tr>
<td>Financial literacy</td>
<td>0.18***</td>
<td>0.15*</td>
<td></td>
<td></td>
<td>0.18***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.088)</td>
<td></td>
<td></td>
<td>(0.057)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education*financial literacy</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of community groups: 1</td>
<td></td>
<td></td>
<td></td>
<td>0.28***</td>
<td>0.27***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.086)</td>
<td>(0.087)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of community groups: 2</td>
<td></td>
<td></td>
<td></td>
<td>0.56***</td>
<td>0.55***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.100)</td>
<td>(0.101)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of community groups: &gt;= 3</td>
<td></td>
<td></td>
<td></td>
<td>0.73***</td>
<td>0.73***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.123)</td>
<td>(0.127)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community group(1)*Educ</td>
<td></td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.132)</td>
</tr>
<tr>
<td>Community groups(2)*Educ</td>
<td></td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.027)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community groups(&gt;= 3)*Educ</td>
<td></td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.036)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some friends are vendors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.132)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>8.37***</td>
<td>8.37***</td>
<td>8.31***</td>
<td>8.51***</td>
<td>8.51***</td>
<td>8.12***</td>
<td>8.07***</td>
<td>8.38***</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.192)</td>
<td>(0.203)</td>
<td>(0.213)</td>
<td>(0.219)</td>
<td>(0.196)</td>
<td>(0.218)</td>
<td>(0.224)</td>
</tr>
<tr>
<td>Observations</td>
<td>995</td>
<td>995</td>
<td>959</td>
<td>964</td>
<td>964</td>
<td>988</td>
<td>988</td>
<td>990</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.115</td>
<td>0.115</td>
<td>0.105</td>
<td>0.091</td>
<td>0.102</td>
<td>0.142</td>
<td>0.142</td>
<td>0.105</td>
</tr>
</tbody>
</table>

Standard errors (clustered at the market level) in parentheses; District FE included; *** p<0.01, ** p<0.05, * p<0.1
<table>
<thead>
<tr>
<th></th>
<th>Fin Lit IV Reg</th>
<th>Numeracy IV Reg</th>
<th>Literacy IV Reg</th>
<th>Com. Groups IV Reg</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Years of education</strong></td>
<td>0.04***</td>
<td>0.05***</td>
<td>0.07***</td>
<td>0.05***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.012)</td>
</tr>
<tr>
<td><strong>Years of experience (market vending)</strong></td>
<td>0.01</td>
<td>0.00</td>
<td>0.03***</td>
<td>0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.013)</td>
</tr>
<tr>
<td><strong>Experience squared</strong></td>
<td>-0.00</td>
<td>-0.00*</td>
<td>-0.00**</td>
<td>-0.00***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td><strong>female</strong></td>
<td>-0.17**</td>
<td>-0.24***</td>
<td>-0.19***</td>
<td>-0.18***</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.064)</td>
<td>(0.066)</td>
<td>(0.030)</td>
</tr>
<tr>
<td><strong>Item sold = food</strong></td>
<td>-0.03</td>
<td>-0.05</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.060)</td>
<td>(0.056)</td>
<td>(0.101)</td>
</tr>
<tr>
<td><strong>At least 1 yr. of completed education</strong></td>
<td>0.43***</td>
<td>0.44***</td>
<td>0.50***</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.133)</td>
<td>(0.068)</td>
<td>(0.166)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-0.14</td>
<td>-0.28*</td>
<td>-0.31**</td>
<td>0.39***</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.126)</td>
<td>(0.168)</td>
<td>(0.058)</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>995</td>
<td>991</td>
<td>995</td>
<td>988</td>
</tr>
<tr>
<td><strong>R-squared</strong></td>
<td>0.072</td>
<td>0.063</td>
<td>0.113</td>
<td>0.091</td>
</tr>
<tr>
<td><strong>Hansen J stat</strong></td>
<td>9.382</td>
<td>10.67</td>
<td>21.27</td>
<td>16.70</td>
</tr>
<tr>
<td><strong>Hansen p-value</strong></td>
<td>0.496</td>
<td>0.384</td>
<td>0.0193</td>
<td>0.0812</td>
</tr>
</tbody>
</table>

Standard errors (clustered at the market level) in parentheses; District FE included; *** p<0.01, ** p<0.05, * p<0.1
IV-regressions estimated using Lewbel’s (2012) method with Uganda’s Primary Education reform as additional instrument.
To further substantiate whether the proposed mechanisms are indeed linked to education, I estimate a basic first stage for the effect of education on financial literacy, literacy, numeracy and participation in community groups (Table 6). The association between years of schooling and these mechanisms points in the expected direction: Higher levels of education are associated with higher levels of business specific-knowledge, generalized knowledge and social capital. Given the potential endogeneity of education not only to unobserved individual skills and talent but also to parental socioeconomic background, I use again an instrumented measure of education to be able to establish a causal relationship (see Lewbel, 2012). The point estimates of instrumented education on the different mechanisms still point in the right direction and are significant for business specific knowledge (financial literacy), numeracy and literacy, but not social capital. However, I reject the null hypothesis of joint instrument validity in the regression for literacy and community group engagement. Conclusions on the effect of schooling on the market vendor’s literacy and social capital are therefore less straightforward to establish: The insignificance of instrumented education (or potentially invalid instruments) might be indicative of the fact that social capital and literacy are relevant for earning outcomes in their own right but are not a transmission belt from education to earnings. At the same time, I cannot reject the hypotheses that the instruments only pick up variation in education that is irrelevant for the channels’ effect on earnings.

6.2. Confounders

Confounding factors – particularly if they are linked to both education as well as productivity outcomes – might lead to biased estimates. I am therefore including additional controls in the estimation of equation 4. I start by including personality traits (patience and risk aversion) and then also implement Chernozhukov et al.’s (2018) double machine-learning partialling-out approach.

Patience has a marginally significant negative effect on income but leaves the education effect unchanged. While this does not rule out that having been to school longer makes people more patient (or more patient individuals go to school longer), there is no evidence that more patient people profit more from education in terms of their income (Table 7 columns 1 and 2). And risk aversion decreases income. A possible explanation could be that more risk averse individuals make fewer investments or it is the result of the stylized fact that females are more risk averse (and earn less in our sample). Risk attitudes have however no effect on the returns to education. Education does not moderate risk aversion effects on income and risk aversion does not moderate returns to education (see Table 7 columns 3 and 4).
<table>
<thead>
<tr>
<th></th>
<th>Outcome: Average daily income (log)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Years of education</td>
<td>0.06*** (0.010)</td>
<td>0.06*** (0.010)</td>
<td>0.06*** (0.010)</td>
<td>0.06*** (0.010)</td>
<td>0.04*** (0.015)</td>
</tr>
<tr>
<td>Years of experience</td>
<td>0.04*** (0.015)</td>
<td>0.04*** (0.015)</td>
<td>0.05*** (0.015)</td>
<td>0.05*** (0.015)</td>
<td>0.05*** (0.015)</td>
</tr>
<tr>
<td>Experience squared</td>
<td>-0.00 (0.001)</td>
<td>-0.00 (0.001)</td>
<td>-0.00* (0.001)</td>
<td>-0.00* (0.001)</td>
<td>-0.00* (0.001)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.46*** (0.103)</td>
<td>-0.46*** (0.103)</td>
<td>-0.50*** (0.106)</td>
<td>-0.50*** (0.106)</td>
<td>-0.50*** (0.106)</td>
</tr>
<tr>
<td>Item sold = food</td>
<td>-0.25*** (0.087)</td>
<td>-0.26*** (0.088)</td>
<td>-0.24*** (0.089)</td>
<td>-0.24*** (0.089)</td>
<td>-0.24*** (0.089)</td>
</tr>
<tr>
<td>Patience (standardized)</td>
<td>-0.08* (0.046)</td>
<td>-0.12 (0.079)</td>
<td>-0.24*** (0.051)</td>
<td>-0.24*** (0.051)</td>
<td>-0.24*** (0.051)</td>
</tr>
<tr>
<td>Education*Patience</td>
<td>0.01 (0.011)</td>
<td>-0.13*** (0.047)</td>
<td>-0.14*** (0.065)</td>
<td>-0.14*** (0.065)</td>
<td>-0.14*** (0.065)</td>
</tr>
<tr>
<td>Risk aversion (standardized)</td>
<td>-0.08 (0.046)</td>
<td>-0.12 (0.079)</td>
<td>-0.24*** (0.051)</td>
<td>-0.24*** (0.051)</td>
<td>-0.24*** (0.051)</td>
</tr>
<tr>
<td>Education*Risk aversion</td>
<td>0.00 (0.008)</td>
<td>0.00 (0.087)</td>
<td>0.00 (0.087)</td>
<td>0.00 (0.087)</td>
<td>0.00 (0.087)</td>
</tr>
<tr>
<td>Constant</td>
<td>8.32*** (0.179)</td>
<td>8.32*** (0.180)</td>
<td>8.34*** (0.182)</td>
<td>8.34*** (0.182)</td>
<td>8.34*** (0.182)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.109</td>
<td>0.109</td>
<td>0.116</td>
<td>0.116</td>
<td>0.099</td>
</tr>
<tr>
<td>Observations</td>
<td>987</td>
<td>987</td>
<td>968</td>
<td>968</td>
<td>752</td>
</tr>
</tbody>
</table>

Standard errors (clustered at the market level) in parentheses; District FE included; *** p<0.01, ** p<0.05, * p<0.1

Column 3 presents the education coefficient obtained from Chernozhukov et al.’s (2018) double machine-learning partialling-out estimator.

Chernozhukov et al.’s (2018) double machine-learning approach to partial out the potentially confounding effects of other covariates from the effect of education on income from self-employment confirms baseline findings. The “pure” effect of an additional year of schooling from which the effects of other covariates have been partialled out amounts to an increase in income of 4 percent (Table 7, column 5). This provides further evidence for the claim that education increases worker productivity and thus earnings even within a narrowly defined type of self-employment in a rural developing country setting. In particular, this finding suggests the existence of a true education effect that is not driven by some underlying confounders.

Furthermore, I show that the separate discussion of the channels does not drive the observed returns to education. Even in a regression in which all potential mechanisms are included, I find returns to education still amount to 5 percent per additional years of schooling (Table 8 column 6).
In addition, the increased income due to more years of completed education even translates into higher per capita consumption. Coefficients match closely (see Table 8 column 5).

6.3. Additional considerations - Selection into self-employment

In the sample of market vendors studied in this paper, I do not observe those individuals that follow different activities or do not earn anything at all.\textsuperscript{23} If returns to education are different in these other activities and if selection into market vending is influenced by education, the estimated returns to education for market vendors might be biased. While the choice to sample market vendors only does not allow me to formally address selection concerns using the Heckman selection model, I am confident the overall selection into self-employment is not a major source of concern in the present setting for the following reasons. First, while several scholars argue that better educated individuals are more likely to select into off-farm work since education increases entrepreneurial ability (see for ex. Le, 1999; Tao Yang, 1997), this form of selection requires the existence of choice options. Yet, alternative income earning opportunities are often limited in rural agricultural contexts (see for ex. World Bank, 2013). In addition, if self-employment is started due to labour market push factors, individual characteristics and in particular human capital play a much smaller role in the selection into entrepreneurship. Second, in their meta study on entrepreneurship selection and performance, Van der Sluis et al. (2008) do not find an effect of education on selection into entrepreneurship.

7. Discussion and Conclusion

This paper presents robust evidence for the existence of returns to education for the self-employed in a rural developing country setting. In contrast to much of the existing literature, these returns materialize within one particular self-employment activity and do not rely on dynamic labour market opportunities and occupational choice options. Using a unique sample of market vendors in Western Uganda, I find that one additional year of schooling can be attributed to average daily income increases from market vending by 7 percent. These returns are comparable to what

\textsuperscript{23}Note that I also do not observe earnings for individuals who migrated. Yet, I argue this will not affect estimated returns for two reasons. First, the rural population studied in this paper provides only very limited opportunities for migration and second, migration generally leads to increased returns due to the higher real wages in urban areas (Schultz, 1988). Hence, estimated returns could be seen as lower bound.
Table 8: Returns to education: Worker effect - Robustness tests

<table>
<thead>
<tr>
<th></th>
<th>avg. daily inc (1)</th>
<th>avg. daily inc (2)</th>
<th>avg. daily inc (3)</th>
<th>avg. daily inc IV Reg (4)</th>
<th>avg. daily cons p.c. inc (5)</th>
<th>avg. daily inc (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of education</td>
<td>0.04*** (0.010)</td>
<td>0.02* (0.010)</td>
<td>0.03*** (0.006)</td>
<td>0.05*** (0.013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest educ level: some/completed primary</td>
<td>0.36*** (0.109)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest educ level: some/completed secondary</td>
<td>0.59*** (0.130)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest educ level: Tertiary or University</td>
<td>1.18*** (0.220)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least 1 yr. of completed education</td>
<td>0.72*** (0.230)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of experience</td>
<td>0.04*** (0.015)</td>
<td>0.04*** (0.014)</td>
<td>0.04*** (0.012)</td>
<td>0.03** (0.012)</td>
<td>-0.00 (0.008)</td>
<td>0.04*** (0.015)</td>
</tr>
<tr>
<td>Experience squared</td>
<td>-0.00 (0.001)</td>
<td>-0.00 (0.001)</td>
<td>-0.00 (0.000)</td>
<td>-0.00 (0.000)</td>
<td>-0.00 (0.000)</td>
<td>-0.00 (0.000)</td>
</tr>
<tr>
<td>female</td>
<td>-0.48*** (0.101)</td>
<td>-0.44*** (0.102)</td>
<td>-0.30*** (0.099)</td>
<td>-0.48*** (0.094)</td>
<td>-0.10 (0.068)</td>
<td>-0.39*** (0.113)</td>
</tr>
<tr>
<td>Item sold = food</td>
<td>-0.25*** (0.086)</td>
<td>-0.26*** (0.089)</td>
<td>-0.22*** (0.090)</td>
<td>-0.28*** (0.086)</td>
<td>-0.16** (0.064)</td>
<td>-0.29*** (0.093)</td>
</tr>
<tr>
<td>Ever received fin lit training</td>
<td>0.28*** (0.093)</td>
<td>0.23** (0.099)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial literacy</td>
<td>0.17*** (0.052)</td>
<td>0.14*** (0.052)</td>
<td></td>
<td></td>
<td>0.16*** (0.053)</td>
<td></td>
</tr>
<tr>
<td>Numeracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.05 (0.057)</td>
<td></td>
</tr>
<tr>
<td>Literacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.01 (0.098)</td>
<td></td>
</tr>
<tr>
<td>Business investments (log)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.11*** (0.031)</td>
<td></td>
</tr>
<tr>
<td>Total current savings (log)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.03*** (0.008)</td>
<td></td>
</tr>
<tr>
<td># of community groups: 1</td>
<td>8.28*** (0.207)</td>
<td>8.37*** (0.192)</td>
<td>6.78*** (0.429)</td>
<td>8.20*** (0.267)</td>
<td>11.38*** (0.109)</td>
<td>8.14*** (0.221)</td>
</tr>
<tr>
<td># of community groups: 2</td>
<td>0.51*** (0.097)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of community groups: &gt;= 3</td>
<td>0.63*** (0.128)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>995</td>
<td>993</td>
<td>911</td>
<td>991</td>
<td>1.201</td>
<td>952</td>
</tr>
<tr>
<td>Observations</td>
<td>15.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hansen J stat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.116</td>
</tr>
</tbody>
</table>

Standard errors (clustered at the market level) in parentheses; District FE included; *** p<0.01, ** p<0.05, * p<0.1
IV-regressions estimated using Lewbel’s (2012) method with Uganda’s Primary Education reform as additional instrument.
other studies have found for settings in which selection into different types of jobs or sheep skin effects (Kolstad et al., 2014) play a much larger role. In addition, this paper combines several theoretical approaches to investigate the precise mechanisms through which educational returns might arise. It shows that educational returns are the result of an effect on the choice of the higher return sector, credit constraints and the development of different skill sets. Applying Welch’s (1970) concept of allocative and productive efficiency effect of education, this paper finds that education stimulates the selection into the higher return vending category. And results from an investigation of the interactions between financial constraints and human capital suggest that higher levels of education help to overcome the barrier posed by financial constraints for choosing the higher return vending activity. Furthermore, this paper finds that both generalized knowledge and business specific skills acquired through schooling raise the market vendor’s productive efficiency once sectoral selection is accounted for. These two skill sets are not perfect substitutes for each other. In addition, individuals with larger social capital also receive a higher income from self-employment.

This has two implications. First, even in a context in which schooling standards are not optimal (see for ex. Bold et al., 2017), individuals active in self-employment profit from education. It is not only business specific skills but in particular also the generalized education from which those self-employed seem to profit as they need to perform a multitude of tasks (Lazear, 2004). Similarly, schooling might help the development of the capacity to access and process new information, which in turn raises their productivity (Rosenzweig, 1995).

Second, the insights from this study are relevant for other developing country contexts as well. In the absence of a market-clearing amount of decent (wage) jobs, many poor individuals enter self-employment (Quatraro and Vivarelli, 2014). And this situation is unlikely to change soon. Despite the recent economic growth, self-employment is expected to stay important in many developing country economies (Filmer and Fox, 2014). Understanding which factors can help improve earnings for these self-employed is therefore important to untighten the link between poverty and labour market outcomes (Fields, 2011). This paper has shown that there are economically meaningful returns to education for own account workers even if selection into self-employment is driven by lack of alternatives and not by increased entrepreneurial abilities. Yet, the net effect of schooling on earnings in the entire economy remains to be established. In particular, the analysis based on the present sample is unable to make reliable claims as to whether the observed returns are net gains
that would accrue to the entire economy or rather distribution effect. If schooling indeed also increases income for the petty self-employed, then more investments in general human capital (through improved curricula or access to education) might be one promising route towards poverty alleviation. More research on other small-scale self-employment activities will be very informative.

References


Appendix – List of variables included in the double machine-learning partialling-out estimation

All variables presented in the descriptive statistics (Table 1) have been included in the LASSO estimation of market income and education. For the estimation, I transformed assets into a standardized asset value and used the dependency ratio instead of the number of supported children. The particular estimation method is the LARS (least angle regression), the LASSO approach for linear models (Efron et al., 2004). I included the set of covariates that minimized the $C_p$ statistic (similar to an AIC) for the partialling-out regressions. For the cross-validation, the sample was split into two groups.

Based on the results from the LASSO estimation, the coefficient vector $X$ for the market income partialling-out regression contains:

$$X = \{\text{credit constraints, standardized asset value, female, patience (standardized), financial literacy, selling food, risk aversion (standardized), also vending on other markets, community groups, district FE, business log, business formally registered, market income share in total income, main material of the wall is bricks, farming, whether some friends are vendors, distance to the market }\geq 30\text{ minutes, relative wealth (standardized), literacy, numeracy, experience, experience squared, paying taxes}\}$$

Similarly, the coefficient vector $X$ for the education partialling-out regression contains:

$$X = \{\text{literacy, business log, age, dependency ratio, credit constraints, female, business formally registered, experience, experience squared, also vending on other markets, district FE, community groups, respondent is main contributor to HH income, household size, distance to the market }\geq 30\text{ minutes, farming, standardized asset value, numeracy, more than one job, main material of the wall is bricks, financial literacy, risk aversion (standardized), farming}\}$$