CREDIT Research Paper

No. 21/01

Does the Pay Period Matter in Estimating Returns to Schooling? Evidence from East Africa

by

Livini Donath, Oliver Morrissey and Trudy Owens

Abstract

This paper investigates whether returns to schooling differ according to the choice of the measure of earnings and the different periods in which workers are paid (daily, weekly, and monthly). Using comparable data from the Living Standards and Measurement Study (LSMS) for Malawi, Tanzania and Uganda, accounting for endogeneity using Gaussian Copula and for selection with the Heckman method, we show that converting earnings to common measures and pooling respondents produces different estimates of returns to education. Depending on the common measure chosen, estimates of returns for level of education can differ by up to 100% for Tanzania, up to 50% for Malawi and up to 20% for Uganda. Estimating separately for each pay period, returns also differ significantly. Returns to primary education are 40-70% in Uganda and 20-30% in Malawi and Tanzania. Returns to secondary education are about 80% in Malawi and Tanzania but vary between 50% and 90% in Uganda. Returns to higher education are 130% in Tanzania, 100-150% in Uganda and 120-165% in Malawi. Returns to increase with the level of education completed but estimating separately for different periods is more reliable than pooling.

JEL Classification: I20, J24, O15

Keywords: returns to education, schooling, earnings, pay period, East Africa

Centre for Research in Economic Development and International Trade, University of Nottingham

CREDIT Research Paper

No. 21/01

Does the Pay Period Matter in Estimating Returns to Schooling? Evidence from East Africa

by

Livini Donath, Oliver Morrissey and Trudy Owens

Outline

1. Introduction
2. Literature
3. Empirical Strategy
4. Data and Descriptive Statistics
5. Results and Discussion
6. Conclusion

References

Appendices

The Authors

Livini Donath is a PhD candidate at the School of Economics, University of Nottingham and an assistant lecturer at the University of Dar es Salaam, Tanzania. Oliver Morrissey and Trudy Owens are respectively Professor and Associate Professor at the School of Economics, University of Nottingham.

Corresponding Author: livini.donath@nottingham.ac.uk /livinitesha@gmail.com

Research Papers at www.nottingham.ac.uk/economics/credit/

1. Introduction

It is important to understand the rate of returns to education because it is one of the significant determinants of willingness to invest in education. Part of the justification for public investment in education is that it adds to human capital, skills, and productivity; this should generate a social benefit in addition to the private benefit of increased earnings by more productive workers. People will be willing to pay for education if it increases their earnings (Borjas, 2016) and parents' willingness to invest resources in their children's education depends on how they value future benefits that the children will get after acquiring education (Schultz, 2004). In developing countries expected returns to education may also be an important determinant of child labour—high returns increase school attendance and tend to reduce the likelihood of child labour (Kuepié & Nordman, 2016).

Most studies on returns to education over the past five decades have concentrated on developed countries(Psacharopoulos & Patrinos, 2004, 2018). Nonetheless, there is an emerging body of literature estimating returns to education in Sub-Saharan Africa (SSA) countries, including Malawi, Tanzania and Uganda, with a broad consensus that since around 2000 returns to secondary education have exceeded those of primary education. While coefficient estimates vary, returns are increasing with the level of education (and generally also with years of education). Given limitations in the data, there are weaknesses in the existing evidence. This paper addresses some of these.

Studies on returns to education are mainly based on nationally representative surveys conducted by government statistical agencies and in some cases on surveys by private researchers. These surveys usually collect data on earnings by different pay periods (typically daily, weekly and monthly) which often reflect the type of employment. Most studies then measure earnings by converting these to a common period, normally hourly or monthly (see for example Nikolov & Jimi (2018), Mishra & Smyth (2015), Peet et al. (2015), and Serneels et al. (2017)). For instance, daily and weekly rates are converted to monthly rates by multiplying by a factor of 22 or 4, respectively. A concern of this approach is the possible introduction of measurement error (e.g., a person paid daily may not work a 'normal working' week), leading to biased estimates on the returns to education. To the extent that pay periods indicate different labour markets, pooling can generate misleading estimates of returns to education.

This paper analyses returns to education in Malawi, Tanzania and Uganda by answering two key questions. Firstly, when earnings are aggregated to a common unit, do different units give different estimates of returns to schooling? Secondly, does the pay period matter in estimating returns to schooling in East Africa? Benefiting from relatively large and recent nationally representative datasets, this paper tests the unexplored hypotheses that estimates of returns to schooling depend on (i.e., vary according to) the period of measurement of the earnings and that different conversions may lead to different estimates. To the best of our knowledge, there have been no studies on Africa which have explored this issue. Given the absence of good and comparable instruments for education to account for endogeneity, omitted variable bias and selection, we estimate returns by applying the Gaussian Copula (GC) instrument free method proposed in Park & Gupta (2012), combined with Heckman model for selection into employment categories.

To compare the effect of alternative common measure for estimates of returns to education three different conversions are employed and estimated for the pooled sample. The three common units are: daily (hereafter DailyC), monthly (hereafter MonthlyC) and annualised expressed per month (hereafter MonthlyA). The main difference between the latter two is that MonthlyC allows for information on the number of days worked in a typical month while MonthlyA also accounts for information on the number of months worked in a typical year. To address the effect of pooling the samples for each pay period (daily, weekly and monthly) are estimated separately. Malawi has a unique labour market structure, with a disproportionately large proportion of the labour force in rural areas primarily participating in agriculture and off-own-farm casual (mainly piece rate paid daily) jobs locally known as *ganyu*. This group is excluded from the primary analysis and analysed separately using the same estimation methods.

Our results suggest that pooling earnings converted to different common measures produces different estimates of returns that are generally biased in the direction of the pay period that constitutes the largest proportion of the sample; pooling the periods together may lead to imprecise estimates. Specifically, in Malawi the returns for workers paid daily are the highest, followed by monthly and then weekly (we address how this may be due to a specific feature of the labour market). Generally, converting *ganyu* earnings to monthly yields larger estimates of returns to education than converting to daily or annualised. In Tanzania, the returns for workers paid weekly are not only the highest but also increase at a higher rate than for the other pay periods. In Uganda, returns are highest for weekly earners followed by monthly and then daily. In this regard, our analysis suggests that estimating returns separately for workers paid over different periods is more reliable than pooling.

The rest of the paper is organised as follows: Section 2 provides an overview of the related literature with a focus on studies for sub-Saharan Africa. Section 3 describes the empirical methodology, especially how endogeneity is addressed, followed by Section 4 on data and description. Section 5 presents and discusses the results, and Section 6 concludes with a summary and comments on the possibility that pay periods indicate segmented labour markets.

2. Related Literature

Mincer (1974), based on the human capital investment theory, developed a model for analysing the effect of education on wage earnings called the human capital earnings function (or Mincer wage function). This approach models the logarithm of wage earnings as follows:

$$logW = a + bS + cE + dE^2 + e \qquad (1)$$

Where W is wage earnings, S is completed years of schooling, E is labour market experience, a, b, c and d are parameters, and e is an error term. Since its formulation, this model has become the standard model for analysing returns to education, with many studies extending it to include more variables that affect wage earnings such as gender, race and work-related characteristics (Card, 1999; 2001; Patrinos & Psacharopoulos, 2010; Peet, Fink, & Fawzi, 2015).

One of the challenges with the Mincer model is how to estimate the causal effect of education on earnings with the endogeneity of education given unobserved ability. The consensus in the literature is that without controlling for individual ability, OLS on (1) gives inconsistent estimates (Cameron & Trivedi, 2005; Wooldridge, 2010). Economists have adopted various methods to address this problem. The most widely used solution for addressing this issue is to use instrumental variables (IV) based on either two-stage least squares (2SLS) or a control function (Cameron & Trivedi, 2009; Card, 1999, 2001). Studies have employed different instruments, generally dictated by availability of data.

An alternative solution is using instrument free methods, that is, methods which do not require any external instruments. They include latent instrumental variables (Ebbes et al., 2005); methods that use heteroscedasticity to obtain identification (Farré et al., 2013; Klein & Vella, 2009, 2010; Lewbel, 2012); and Gaussian Copula (Park & Gupta, 2012). These methods are particularly useful when there are no (good) instruments in the data. To date, these methods have not yet been widely applied in the returns to education literature.

Another key challenge with the Mincer model is how to deal with sample selection bias. Sample selection arises because wages are observed only for individuals in employment who report positive values of wages during data collection. The wages of the wage earners might not reflect the wages of the non-wage income earners (for example, the self-employed or casual agricultural workers) had they worked in wage employment. If the exclusion of these individuals from the analysis is not random, without controlling for how individuals select into wage employment the OLS estimator will give inconsistent estimates (Cameron & Trivedi, 2005; Heckman, 1979; Verbeek, 2004). The standard solution for this problem is to use the Heckman Two-step Sample Selection Model formulated by Heckman (1979). The model recovers consistent estimates by running OLS in two steps where the exclusion from the sample is modelled as an omitted variable (see section 3).

Psacharopoulos & Patrinos (2018) document that in developing countries returns to an extra year of schooling averages about 9.2% compared to 8% in developed countries (Appendix A provides a summary of studies for developing countries). These studies differ widely in terms of methods (including OLS, traditional IV, propensity scores matching, and Heckman sample selection models) and data (such as nationally representative, regional or sectoral level data) making it difficult to directly compare the estimates of returns across countries and studies.

Another strand of literature on returns to education in developing countries focuses on examining the possible heterogeneity in returns to education along the earnings distribution and across groups of workers (such as gender, sector of employment and location). Typical results are that the pattern of returns to education differ across the earnings distribution (see for example Chuang & Lai (2017), Stefani & Biderman (2009) and Girma & Kedir (2005)); females have higher returns to education compared to males (Nikolov & Jimi, 2018; Peet et al., 2015; Salisbury, 2016; Schultz, 2004); public sector employees have higher returns than their private counterparts (Lassibille & Tan, 2005); rural workers have higher returns than urban workers and wage employees have higher returns than the self-employed and agricultural workers (Al-Samarrai & Reilly, 2008).

A novel study by Serneels et al. (2017) examined whether the type of questionnaire used in collecting individuals' labour market information matters in estimating returns to education in Tanzania. By using both short and detailed questionnaires, the study found that returns differed by the survey instrument: short module questionnaires led to biased estimates compared to detailed questionnaires. After controlling for endogeneity due to unobserved ability and selection by using a control function, Heckman and Heckman-Hotz methods, the estimated returns ranged between 20-21% for men and 32-49% for women for a year of post-primary school if short modules were used. For the detailed modules, no effects of schooling on wage were found for men, while returns for women were between 29% and 50%.

Whilst much effort has been put into addressing issues like endogeneity of education in estimating returns to education, heterogeneity of returns across the earnings distribution and groups of workers, little attention has been focused on whether the pay period matters in estimating returns to education. What is evident in all the previous studies is the conventional method of aggregating earnings to a common period such as hourly, daily, monthly or annual earnings. However, what is not clear is the impact this has on their findings. In this paper, we demonstrate that the relationship between earnings and education may vary across workers reporting wage earnings over different periods. We argue that the precision of converting the reported wages to the universal unit may be plagued by errors and assumptions made by the researcher, thereby biasing the estimates of the returns. In fact, different common measures may give different estimates of returns to education. Unlike previous studies, this paper considers the implication of alternative ways of converting the reported wages to a common unit/measure and provides separate estimates for each pay period.

3. Empirical Strategy

Recent studies in Tanzania (as discussed in section 2) show increasing returns with levels of education (convex schooling-earning function). We adopt the Mincer equation with quadratic schooling from Söderbom et al. (2014) to ascertain the possible convexity in returns. Thus, our empirical model is specified as follows:

$$Y_{it} = \alpha_1 S_{it} + \alpha_2 S_{it}^2 + \beta X_{it} + \mu_{it} \qquad (2)$$

Where Y_{it} is the log of earnings, S (S^2) is individual's years of schooling (squared), X_{it} is a vector (containing a constant) of individual characteristics (age in years and its square, gender, location, and a dummy variable for individuals observed more than once), i and t index individual and time respectively and μ is a standard error term. The parameters of interest are α_1 and α_2. The sign of α_2 tells us about the shape of the earning function: positive implies convexity, negative implies concavity, and zero implies linearity.

Since the rates of return to schooling may differ by level of education, we also use an alternative specification that uses dummies for completed levels of education to estimate returns to each level of education. Three levels are used for this purpose: Primary, Secondary and Higher. (including). As there are very few observations with above secondary education, tertiary non-university (post-secondary diploma) and university are merged into one group (Higher). The following specification is estimated:

$$lnw_{it} = \delta educ_{it} + \gamma X_{it} + \varepsilon_{it} \qquad (3)$$

Where $educ$ is a vector of dummies for the levels of education with “less than primary education (no education hereafter)” as the reference category, X is as defined earlier, and ε an error term. The returns associated with each level of education with respect to the reference category is given by the vector δ. The returns per additional year of level l with respect to level m can be obtained using the following equation:

$$r_l = \frac{\delta_l - \delta_m}{S_l - S_m} \qquad (4)$$

Where r is the return per year, $\delta_l - \delta_m$ is the difference in returns between the two levels and $S_l - S_m$ the difference in years of schooling between the levels.

In our specifications, we use age and its square in place of experience and its square for two main reasons. Firstly, the surveys did not explicitly ask the years of experience the individual spent in the current job. Therefore, defining experience as age less years of schooling less school starting age as common in the literature might result in accumulation of errors, especially if there were measurement errors in age, years of schooling and/or school starting age. Furthermore, we would have missing values for those who did not report their school starting age or have to choose an arbitrary starting age. Secondly, if schooling happens to be endogenous due to, among other reasons, unobserved ability, by construction experience would also be endogenous. To avoid these issues, we use age as a proxy for experience in our analysis.

While most workers in the sample are paid monthly, significant shares report earnings daily and weekly (also fortnightly and quarterly for Tanzania). The standard method is to convert/aggregate all wages into a common period such as monthly wage or annualised wage (then expressed as monthly) earnings and use their log as the dependent variable. Three common measures for wages are constructed: converting to daily, converting to monthly and converting to annual and expressing as monthly., The analysis examines whether these different conversions give different estimates of returns to education. Then, taking the most common monthly measure, returns to education are estimated for each of the three main pay periods (daily, weekly and monthly) separately to examine if the estimates vary by pay period.

3.1. OLS estimation

As a baseline estimation, we estimate (2) and (3) using OLS. It is well known that OLS will give inconsistent estimates of α_1, α_2 and δ because of omitted variables, measurement errors or if there is sample selection bias. A typical example is when these variables are correlated with the residuals in (2) and (3) due to the presence of other factors that are associated with higher education and higher wages but are not included in the models, such as when more educated individuals possess (unobservable) higher ability correlated with higher wages. Estimating (2) and (3) using OLS without controlling for ability will lead to inconsistent estimates (Cameron & Trivedi, 2009). Furthermore, without controlling for how individuals select into wage employment, OLS will also give biased estimates.

3.2. Gaussian Copula Estimation

One of the standard solutions to recovering consistent estimates for (2) and (3) is using instrumental variables. Several studies, as noted above, have employed different instruments for education in estimating returns to schooling. Frequently used instruments include family background characteristics such as parental education, parents' occupation and spouse education; and school system features such as proximity to school, tuition fees, quality of the school, and (change in) compulsory schooling laws in minimum years of basic education (Card, 2001). To a large extent, the choice of instruments is dictated by the availability of data, and almost every instrument is subject to debate.

LSMS being a general household survey, only family background characteristics were available to use as instruments for education. Neither met the requirements for a good instrument: parental education variables were weak whereas household average education (which combines parental education, siblings' education and spouse education) failed the overidentification test.[1]

Another solution is to use instrument free methods (which do not require external instruments). Heteroscedasticity based methods are good candidates but are only suitable when there is one endogenous regressor - in our case there are three (potential) endogenous regressors (S, S^2 and the log of number of weeks worked in the last 12 months). We therefore employ the Gaussian Copula (GC) method as it can be easily extended to include more than

1 Results available on request. The need to have separate instruments for S and S^2 (at least) is another problem.

one endogenous regressor. The GC approach models the correlation between the suspected endogenous variable and the error term through copulas.[2] By including the copula term(s) of the endogenous regressor(s) as additional regressor(s) in the regression model, this method recovers the estimates of the endogenous regressor(s) which are free from endogeneity (Park & Gupta, 2012).

Significant copula terms in the GC regression imply endogeneity (if not OLS is consistent) and the sign of the copula terms shows the direction of the correlation between the endogenous variables and the errors. However, although the model can recover the true effect of the endogenous regressors, it does not tell anything about the source of the endogeneity (see Hult, Proksch, Sarstedt, Pinkwart, & Ringle, 2018). The method is not suitable when the endogenous regressor is binary (Park & Gupta, 2012). Consequently, rather than the more flexible Mincer specification (3) that uses dummies for education levels we rely on quadratic schooling in (2) to infer whether higher levels of education have higher returns than lower levels and report the results for the levels of education corrected for selection bias.

Following Park & Gupta (2012) and Rutz & Watson (2019) our model is derived recalling (2) where both S and S^2 are endogenous (omitting individual and time subscripts for convenience):

$$Y = \alpha_1 S + \alpha_2 S^2 + \beta X + \mu \tag{5}$$

The relationship between the endogenous variables and the error term is modelled as:

$$\begin{pmatrix} S^* \\ S^{2*} \\ \mu^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{\mu 1} \\ \rho_{12} & 1 & \rho_{\mu 2} \\ \rho_{\mu 1} & \rho_{\mu 2} & 1 \end{bmatrix} \right)$$

Where $S^* = \Phi^{-1}(F_1(S))$ and $S^{2*} = \Phi^{-1}(F_2(S^2))$ are GC functions; $F_1(.)$ and $F_2(.)$ are cumulative distribution functions for S and S^2 respectively; ρ_{12} is the correlation between S and S^2; $\rho_{\mu 1}$ the correlation between S and μ; and $\rho_{\mu 2}$ the correlation between S^2 and μ.

The expression can then be written as

2 Cherubini et al. (2004) define Copulas as functions expressing the joint probability distribution as a function of the marginal distribution.

$$\begin{pmatrix} S^* \\ S^{2*} \\ \mu^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{12} & \sqrt{1-\rho_{12}^2} & 0 \\ \rho_{\mu 1} & \frac{\rho_{\mu 2} - \rho_{12}\rho_{\mu 1}}{\sqrt{1-\rho_{12}^2}} & \sqrt{1-\rho_{\mu 1}^2 - \frac{(\rho_{\mu 2} - \rho_{12}\rho_{\mu 1})^2}{1-\rho_{12}^2}} \end{pmatrix} . \begin{pmatrix} \varpi_1 \\ \varpi_2 \\ \varpi_3 \end{pmatrix}$$

Where

$$\begin{pmatrix} \varpi_1 \\ \varpi_2 \\ \varpi_3 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

It then follows that,

$$\mu = \sigma_\mu . \mu^* = \sigma_\mu . \frac{\rho_{\mu 1} - \rho_{12}\rho_{\mu 2}}{\sqrt{1-\rho_{12}^2}} . S^* + \sigma_\mu . \frac{\rho_{\mu 2} - \rho_{12}\rho_{\mu 1}}{\sqrt{1-\rho_{12}^2}} . S^{2*} + \sigma_\mu . \sqrt{1-\rho_{\mu 1}^2 - \frac{(\rho_{\mu 2} - \rho_{12}\rho_{\mu 1})^2}{1-\rho_{12}^2}} . \varpi_3 \tag{5}$$

Where σ_μ^2 is the variance of the error term. Combining (4) and (5) we get

$$Y = \alpha_1 S + \alpha_2 S^2 + \beta X + \sigma_\mu . \frac{\rho_{\mu 1} - \rho_{12}\rho_{\mu 2}}{\sqrt{1-\rho_{12}^2}} . S^* + \sigma_\mu . \frac{\rho_{\mu 2} - \rho_{12}\rho_{\mu 1}}{\sqrt{1-\rho_{12}^2}} . S^{2*} + \sigma_\mu . \sqrt{1-\rho_{\mu 1}^2 - \frac{(\rho_{\mu 2} - \rho_{12}\rho_{\mu 1})^2}{1-\rho_{12}^2}} . \varpi_3 \tag{6}$$

Equation (6) is a linear regression model with the error term given by its last three component. The model disaggregates the endogenous regressors into two components; one is the part not correlated with the error term (S and S^2) and the other is the part which is correlated with the error term (S^* and S^{2*}). By including the copula functions as additional regressors, OLS on model (6) gives consistent estimates for S and S^2 (Park & Gupta, 2012).

Let $\theta_1 = \sigma_\mu . \frac{\rho_{\mu 1} - \rho_{12}\rho_{\mu 2}}{\sqrt{1-\rho_{12}^2}}$, $\theta_2 = \sigma_\mu . \frac{\rho_{\mu 2} - \rho_{12}\rho_{\mu 1}}{\sqrt{1-\rho_{12}^2}}$ and $\xi = \sigma_\mu . \sqrt{1-\rho_{\mu 1}^2 - \frac{(\rho_{\mu 2} - \rho_{12}\rho_{\mu 1})^2}{1-\rho_{12}^2}} . \varpi_3$

Equation (6) can be rewritten as:

$$Y = \alpha_1 S + \alpha_2 S^2 + \beta X + \theta_1 S^* + \theta_2 S^{2*} + \xi \tag{7}$$

Given the discrete nature of our endogenous regressors, the distribution functions $F_1(.)$ and $F_2(.)$ are step functions lying between two values, such that:

$$F(t-1) < U_t < F(t)$$

for any discrete endogenous regressor t; where U_t follows uniform distribution on [0,1]. It follows, therefore, that:

$$\Phi^{-1}\big(F_1(S-1)\big) < S^* < \Phi^{-1}\big(F_1(S)\big); \text{ and}$$

$$\Phi^{-1}\big(F_2(S^2-1)\big) < S^{2*} < \Phi^{-1}\big(F_2(S^2)\big)$$

Since $F_1(.)$ and $F_2(.)$ are estimable from the data, model (7) can be estimated using OLS.

Equation (7) can also be extended to include more endogenous regressors. For example, because the number of weeks worked in the last 12 months may be endogenous due to a bidirectional relationship between total annual wages and the number of weeks worked, we include the variable "W" (for weeks) when using annualised wages. In this case, the empirical model becomes

$$Y = \alpha_1 S + \alpha_2 S^2 + \alpha_3 W + \beta X + \theta_1 S^* + \theta_2 S^{2*} + \theta_3 W^* + \xi \tag{8}$$

3.3. Heckman Selectivity Correction Model

Bias can arise from non-random missingness in earnings data and selection into periods of employment. Some individuals in the dataset do not have values of wage, either because they were unemployed, self-employed at the time of survey or did not respond. The exclusion of these individuals may not be random, so OLS is likely to give inconsistent estimates due to sample selection bias (Cameron & Trivedi, 2005; Verbeek, 2004). The Heckman (1979) selectivity correction is employed to deal with selection bias. To correct the selection bias after controlling for other sources of endogeneity, GC terms are included in the two-step Heckman selection model. The first stage (8) estimates the probit model for selection into periods of payment and wage employment, the regressors being the exogenous variables, GC terms and exclusion restrictions:

$$P_{it} = \vartheta_1 S + \vartheta_2 S^2 + \Psi_1 X + \Psi_2 \Sigma + \phi_1 S^* + \phi_2 S^{2*} + e \tag{9}$$

Where P is the probability of being in wage employment, Σ is the vector of exclusion characteristics (dummy for the household head (*head*), marital status (*married*), the proportion of children under 5 (*kids5*), and proportion of children between 6 and 14 years (*kids14*) in the household). P is defined as follows:

$$P = \begin{cases} 1 \;\; if \;\; Y \geq 0 \\ 0 \;\; if \;\; Y = . \end{cases} \tag{10}$$

We obtain the inverse mills ratio (λ) from (9) and then include it as a regressor in the estimation of (7) and (8) (again omitting individual and time subscripts for convenience). The selection corrected equations for (7) and (8) are respectively given by (11) and (12):

$$(Y|P=1) = \alpha_{11}S + \alpha_{21}S^2 + \beta_1 X + \theta_{11}S^* + \theta_{21}S^{2*} + \pi_1\lambda_1 + \xi_1 \quad (11)$$

$$(Y|P=1) = \alpha_{12}S + \alpha_{22}S^2 + \alpha_{32}W + \beta_2 X + \theta_{12}S^* + \theta_{22}S^{2*} + \theta_{42}W^* + \pi_2\lambda_2 + \xi_2 \quad (12)$$

The obtained estimates of the returns to schooling from (11) and (12) using OLS are consistent and efficient if π_1 and π_2 are significantly different from 0; otherwise, there is no selection problem, and thus GC is more efficient.

4. Data and Descriptive Statistics

The datasets for the analysis come from the Malawi Integrated Household Surveys (IHS), the Tanzania National Panel Surveys (TNPS) and the Uganda National Panel Surveys (UNPS). The surveys are part of the Living Standards and Measurement Study (LSMS), a World Bank program aimed at facilitating the design and implementation of multi-topic household surveys in developing countries since 1980s (World Bank, 2020a). Malawi, Tanzania, and Uganda remain the main beneficiaries of the LSMS program in the Eastern Africa region.

We limit the scope of our study to the period from 2008 through 2017. The IHS, TNPS, and UNPS used similar questionnaires, allowing comparability across the countries. For Malawi we use IHS3 conducted from March 2010 to March 2011 and IHS4 that was conducted from April 2016 to April 2017; for Tanzania we use four waves of TNPS from 2008 to 2016; and for Uganda we use five waves of the UNPS conducted between 2009 and 2016. All surveys are based on two-step stratified sampling from the respective countries' population and housing censuses and the samples are representative at the national, regional, and urban/rural levels. The LSMS data are accessible and freely downloadable from the websites[3] of the World Bank and the countries' statistical offices.

After cleaning the data, we obtained samples of labour force of 45,494, for Malawi, 38,857 for Tanzania and 29,188 for Uganda. Of these samples 5,8161;[4] 11,215; and 4,631

3 Data accessed on 10th May 2019 from http://surveys.worldbank.org/lsms

4 Not including off-own-farm labour (*ganyu)*, see section 5.3.

individuals had valid values of earnings for Malawi, Tanzania and Uganda respectively on which we focus our analysis. TNPS and UNPS being panels of households, the samples for Tanzania and Uganda have some individuals who were surveyed more than once. Therefore, the 11,215 observations for Tanzania consist of a total of 8,210 individuals of which 6,016 were surveyed only once, 1,458 twice, 661 three times and 75 four times. Likewise, the 4,631 observations for Uganda consist of a total of 2,491 individuals of which 1,929 were surveyed once, 704 twice, 207 three times, 98 four times and 58 five times. Thus, although the TNPS and UNPS were intended as panels, due to attrition and refreshing very few individuals (only 27% for TNPS and 23% for UNPS) are observed at least twice in the sample. As a result, in our analyses we include a variable for individuals with repeat observations but otherwise pool and treat observations as independent.

4.1 Variable Construction

The surveys collected information about individuals wage earnings and the frequency of payment. The payment periods for Malawi were daily, weekly, and monthly; for Tanzania hourly, daily, weekly, fortnightly, monthly, quarterly, semi-annually, and annually; and for Uganda hourly, daily, weekly, and monthly. Note that the payment periods may not necessarily imply the same period/duration of employment, i.e., being paid daily or weekly does not always mean that employment lasts only for a day or a week.

Using this information in combination with the available information from the surveys such as the number of days, weeks and months worked, the reported earnings are converted to three different units: daily (hereafter DailyC), monthly (hereafter MonthlyC) and to annualised expressed per month (hereafter MonthlyA). The aim is subsequently to compare these measures to see if they lead to different estimates of returns to education; and then use one measure to examine whether the returns to education vary across workers reporting earnings daily, weekly and monthly. Data availability and the literature guides the independent variables included. Table 1 shows the names and definitions of each variable used in the analysis. A detailed description of how variables were constructed is provided in Appendix B.

Table 1:Definition of Variables Used in the Analysis

Variable	Description
Wage regression function:	
Log(wage)	the logarithm of wage earnings.
sch	individual's total number of years of schooling.
noeduc	educational dummy, 1 if less than primary education and 0 otherwise.
primary	educational dummy, 1 if completed primary and 0 otherwise.
secondary	educational dummy, 1 if completed ordinary/advanced secondary education and 0 otherwise.
higher	educational dummy, 1 if completed post-secondary (diploma/university) education and 0 otherwise.
age	individual's age in years. Its square is included to capture the non-linear relationship between earning and age.
female	a gender dummy, 1 for females, included to capture the effects of gender on wages.
rural	location dummy, 1 for employment in rural areas, is used to control for rural-urban wage differentials.
panel	for Tanzania and Uganda, a dummy, 1 for individuals observed more than once since we are using imperfect panel surveys.
year	only for Malawi, a year dummy, 1 for 2016 and 0 for 2010.
weeks	Log of number of weeks worked in the past 12 months.
Selection equation	
married	dummy for marital status, 1 if married or living together and 0 otherwise.
head	dummy equals 1 if head of the household and 0 otherwise.
kids5	proportion of children under 6 years of age in the household.
kids14	proportion of children aged 6 to 14 years of age in the household.

4.2 Descriptive Statistics

Table 2 shows the distribution of earnings by payment period using our three measures of earnings. We present the earnings in both US dollars ($US) and local currency units (LCU).

The columns for MonthlyC and MonthlyA are directly comparable since both present earnings per month (calculated in different ways). For DailyC column to be comparable to the other columns, they need to be multiplied by a factor of 22 (assuming 22 working days in a month).

Table 2 shows that DailyC and MonthlyC give larger average monthly earnings compared to MonthlyA, consistent with the latter allowing for workers reporting less than 12 months in a typical year. Importantly, DailyC will overestimate monthly earnings because it does not consider the number of days the worker worked in a week, and the number of weeks worked in a month. Regardless of the measures of earning used, the earnings differ by payment period and which period has the highest/lowest earnings depends on the measure of earnings used. Figure 1 shows, using kernel plots, the distribution of earnings by measure of earnings while Figure 2 shows the distribution of monthly earnings (as measured by MonthlyA) by pay period. In both cases the distribution of earnings follows a pattern similar to that reported in Table 2.

Table 3 shows the distribution of the explanatory variables to be used in the wage equation. For Tanzania and Uganda, the variable 'panel' shows the proportion of workers with at least two repeated (panel) observations. The small proportion (45%) for Tanzania reflects relatively high attrition (a major reason why the sample was almost completely refreshed for the last round of the survey). There are a few issues that could potentially affect our results. Workers paid monthly have more education than their daily and weekly counterparts in all three countries. In Tanzania, there are no workers with higher education reporting earnings daily or weekly. In Malawi, only 12% of the workers reporting daily earnings and 4% of those reporting weekly earnings have higher education, while in Uganda 3% and 8% of the workers reporting daily and weekly earnings respectively have higher education. Overall, workers in Malawi have more years of schooling on average compared to their Tanzania and Uganda counterparts.

Table 2:Distribution of Earnings by Different Measures ($US and LCU)

Country & Period	Obs.	$ DailyC		$ MonthlyC		$ MonthlyA		LCU DailyC		LCU MonthlyC		LCU MonthlyA	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Malawi													
Daily	182	18.77	21.60	258.16	318.11	206.42	251.48	2,650.15	3,049.41	36,443.80	44,907.13	29,140.24	35,501.37
Weekly	503	22.10	50.94	223.54	309.10	174.26	289.04	3,120.01	7,190.90	31,557.13	43,635.62	24,599.68	40,804.40
Monthly	5,129	12.59	18.01	264.82	377.90	226.84	339.20	1,777.72	2,542.08	37,384.43	53,347.49	32,022.80	47,884.38
Pooled	5,816	13.72	23.62	260.58	370.04	221.04	332.38	1,937.44	3,334.45	36,786.58	52,237.98	31,204.69	46,921.56
Ganyu	16,528	5.76	6.93	69.41	101.72	33.76	53.52	812.51	977.83	9,797.99	1,4360.21	4,766.48	7,555.73
Tanzania													
Daily	3,738	5.18	5.50	59.08	100.24	32.54	85.08	6,587.81	7,664.81	78,580.79	133,314.68	43,283.94	113,156.05
Weekly	1,929	4.95	5.76	60.98	100.97	32.30	84.51	6,287.49	7,481.10	81,109.67	134,287.74	42,962.47	112,393.41
Monthly	4,830	6.82	8.02	150.02	176.51	123.90	161.28	9,069.55	10,670.72	199,530.00	234,755.77	164,793.41	214,501.08
Pooled	11,215	6.00	7.55	98.14	147.47	69.10	126.88	7,973.95	10,045.83	130,519.70	196,133.61	91,905.45	168,748.00
Uganda													
Daily	1,262	4.64	5.00	91.47	109.51	78.92	103.44	9,429.51	10,159.08	185,739.05	222,364.60	160,242.78	210,038.33
Weekly	589	5.53	8.06	91.60	122.62	76.34	109.07	11,221.61	16,358.95	185,987.04	248,982.52	155,016.12	221,470.95
Monthly	2,765	5.88	6.71	129.33	147.56	114.19	139.02	11,936.67	13,618.94	262,606.79	299,616.72	231,865.71	282,275.95
Pooled	4,631	5.53	6.74	114.11	136.35	99.66	127.84	11,227.57	13,688.05	23,1705.66	276,867.30	202,361.64	259,583.03

Source: Author's computations from IHS, TNPS and UNPS. Note: Earnings in $ are accounted for inflation using exchange rates in 2009 (1$= 141.17 Malawi Kwacha; 1$ = 1330 Tsh; and 1$= 2030.49 UGX). Off-own-farm (*ganyu*) labour excluded from the pooled sample for Malawi. Distribution adjusted by survey weights.

Figure 1: Distribution of Earnings by Earnings Measure and Country

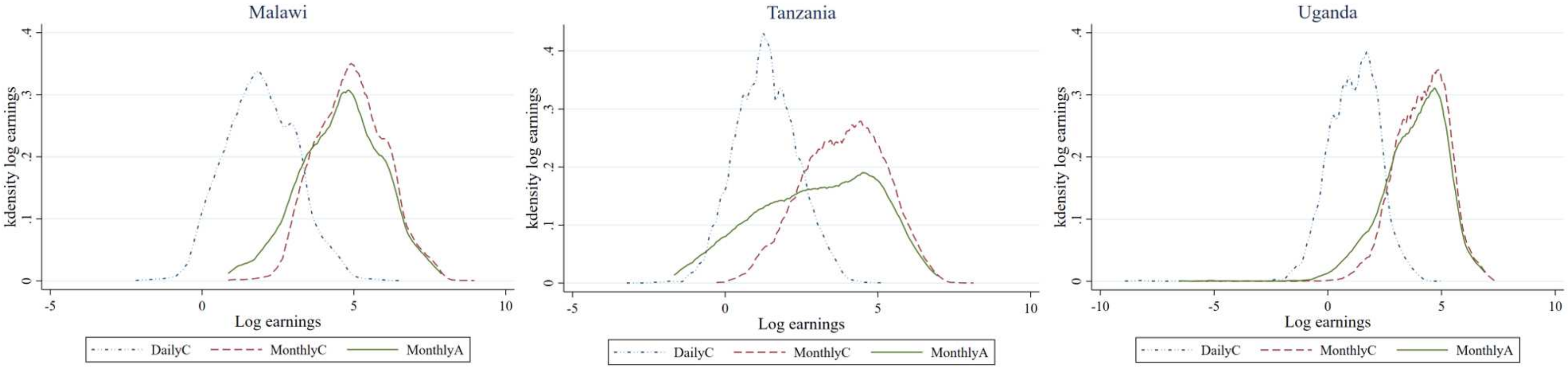


Figure 2: Distribution of Monthly Earnings by Pay Period and Country

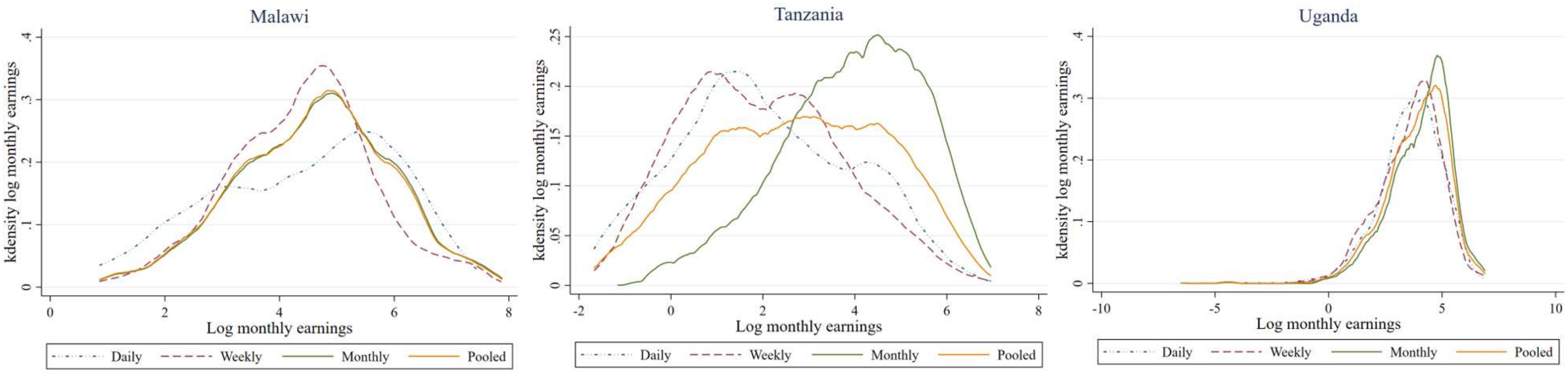


Table 3: Summary Statistics for the Independent Variables Used in Analysis

Country &Sample	Obs.	sch		age		weeks		primary	secondary	higher	female	rural	panel	year
		Mean	SD	Mean	SD	Mean	SD	%	%	%	%	%	%	
Malawi														
Day	182	9.03	4.51	36.80	10.20	36.75	13.99	20	29	12	26	67	NA	74
Week	505	7.15	3.91	34.94	10.35	35.51	14.71	24	10	4	30	66	NA	53
Month	5129	9.26	4.12	35.85	10.71	39.63	12.69	26	24	14	25	49	NA	50
Pooled	5816	9.05	4.16	35.79	10.66	39.13	13.01	26	23	13	25	52	NA	51
Ganyu	16528	4.77	3.52	33.42	11.89	15.64	12.66	14	2	0.0	51	92	NA	64
Tanzania														
Day	3,738	5.26	3.21	33.14	11.92	15.32	15.39	57	4	0	40	78	25	NA
Week	1,929	5.28	3.24	33.64	12.08	13.62	15.16	56	5	0	36	82	19	NA
Month	4,830	8.16	3.69	33.34	11.89	34.95	16.00	51	27	7	38	53	41	NA
Pooled	11,215	6.35	3.67	33.37	11.93	22.10	18.38	55	13	3	38	70	45	NA
Uganda														
Day	1,262	6.35	3.53	30.99	10.84	36.04	14.57	34	11	3	20	64	39	NA
Week	589	6.97	3.81	32.77	11.44	35.27	14.87	32	12	8	29	72	15	NA
Month	2,765	9.85	4.39	34.71	11.06	39.38	12.45	29	16	31	37	56	56	NA
Pooled	4,631	8.51	4.41	33.46	11.18	37.89	13.54	31	14	20	31	61	57	NA

Source: Author's computations from IHS, TNPS and UNPS. *Notes:* weeks is in original scale prior to taking logarithms. The last two columns show % observed multiple times for Tanzania and Uganda (panel) and % in 2016 for Malawi (year) respectively.

5. Results and Discussion

The first sub-section presents the estimates of returns to schooling obtained from pooling all workers together (as previous studies for Africa have done). As noted above, reported wage earnings were converted to three different common periods: daily earnings (DailyC), monthly earnings (MonthlyC) and annualised earnings expressed per month (MonthlyA). Section 5.2 then presents results for the pay period sub-samples estimated separately. Malawi's off-own-farm labour (*ganyu)* is analysed separately in section 5.3.

5.1. Effects of Aggregating Earnings on Estimates of Returns to Schooling

Tables 4-6 compare the estimates of returns to schooling for the three countries from the three measures of earnings. The first three columns present the estimates from the baseline OLS regression (ignoring the possible endogeneity bias). The next three columns (columns 4–6) present estimates corrected for endogeneity due to ability bias using the GC model. The last three columns (columns 7-9) present estimates corrected for both ability bias and selection into employment categories using Heckman sample selection model in combination with GC (HGC). To simplify comparison, the predicted average marginal effects of schooling (AME(sch)) are included in the tables since the quadratic component of years of schooling may complicate the interpretation. As GC and HGC tend to produce unstable estimates when the endogenous regressors are discrete, bootstrap aggregating (bagging) is employed to check the robustness by estimating the regressions many times and averaging the coefficients. The results for bagging in Appendix E indicate that results appear robust.

Irrespective of the measure of earnings or the estimation strategy used, in all countries the coefficient of schooling squared (sch^2) is positive and highly statistically significant implying a strong convex relationship between earnings and years of schooling. While each additional year of education is associated with an increase in earnings, the rate of increase in earnings also increases with years of schooling. That is, the slope of the earnings function increases by some constant amount for each additional year of schooling. For Malawi and Tanzania, the coefficient on schooling (*sch*) is negative (Tables 4 & 5) implying that there is a threshold in the years of schooling (about six years for Malawi and two years for Tanzania) below which the returns are negative.

In line with theory on ability bias but contrary to the IV literature on returns to education, OLS gives upward biased estimates. The predicted marginal effects of schooling including the copula functions for education lowers the returns to education in Malawi by about 50% from 13.8% to 6.0% when the reported earnings are aggregated to daily earnings; from 14.7% to 6.9% when the reported earnings are aggregated to monthly earnings; and from 15.3% to 8.8% when the reported earnings are aggregated to annualised earnings (Table 4). Correcting for selection to employment categories in addition to ability bias lowers the returns even further (daily to 4.1%, monthly to 4.8% and annualised to 7.2%). The coefficient of the inverse mills ratio (IMR) in Tables 4 and 5 is statistically significant, implying that ignoring selection leads to biased results. The negative (positive) sign of IMR implies that there are negative (positive) correlations between the errors in the wage equations and those from the labour force participation equations making OLS results inconsistent. That is, there are unobserved factors that affect participation in wage employment and earnings.

Returns to schooling clearly differ depending on how earnings are converted to a common measure. Table 4 shows that MonthlyA gives larger estimates of returns to schooling in Malawi compared to DailyC or MonthlyC. There is a small (negligible) difference between estimates from DailyC and to MonthlyC (mainly due to the small proportion of the daily earners relative to monthly earners in the sample). The pattern is irrespective of the estimation strategy used. The top panel of Figure 3 plots the HGC (preferred) estimates of returns to schooling from Table 4 for the selected grades and shows how the estimates differ with the measure of earnings.

In Tanzania, MonthlyC gives larger estimates of returns to schooling compared to DailyC or MonthlyA, and the strength of correlation between years of schooling and the error terms in the regressions is significant and stronger for MonthlyC compared to DailyC and MonthlyA (Table 5). The middle panel of Figure 3 plots the HGC estimates of returns to schooling in Tanzania. While the gap between the estimates from DailyC and MonthlyC is generally constant, that between estimates from MonthlyC and MonthlyA increases with education. As for Malawi, MonthlyA gives larger estimates of returns to schooling compared to aggregating to DailyC or MonthlyC in Uganda (Table 6). The correlation between years of schooling and the error terms in the regressions exists and is generally significant. The bottom panel of Figure 3 plots the HGC estimates of returns to schooling in Uganda. The gap between the estimates from DailyC and MonthlyC is also generally constant, while that between estimates from MonthlyC and DailyC as well as between MonthlyC and MonthlyA increases with education. These results support a concern that the choice of the conversion of earnings matters in estimating returns to schooling.

Table 4: Effects of Aggregating Earnings on Estimates of Returns to Schooling in Malawi

Measure	OLS			GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.077***	-0.073***	-0.066***	-0.109***	-0.106***	-0.093***	-0.110***	-0.107***	-0.094***
	(0.009)	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
sch^2	0.011***	0.012***	0.012***	0.009***	0.009***	0.010***	0.008***	0.008***	0.009***
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
age	0.063***	0.059***	0.055***	0.064***	0.060***	0.056***	0.054***	0.049***	0.048***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
$age^2/100$	-0.063***	-0.056***	-0.051***	-0.064***	-0.057***	-0.052***	-0.052***	-0.044***	-0.043***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
female	-0.134***	-0.136***	-0.105***	-0.134***	-0.136***	-0.106***	-0.059*	-0.052	-0.044
	(0.023)	(0.022)	(0.021)	(0.022)	(0.021)	(0.021)	(0.035)	(0.033)	(0.032)
rural	-0.136***	-0.209***	-0.207***	-0.135***	-0.209***	-0.209***	-0.060*	-0.124***	-0.146***
	(0.021)	(0.020)	(0.019)	(0.021)	(0.019)	(0.019)	(0.032)	(0.032)	(0.031)
Copula(sch)				0.146***	0.149***	0.123***	0.151***	0.155***	0.127***
				(0.047)	(0.045)	(0.045)	(0.048)	(0.046)	(0.045)
Copula(sch2)				0.154***	0.154***	0.128***	0.157***	0.157***	0.131***
				(0.049)	(0.047)	(0.046)	(0.050)	(0.049)	(0.050)
Copula(weeks)						1.149***			-0.005
						(0.029)			(0.004)
IMR							-0.164***	-0.184***	-0.136***
							(0.054)	(0.051)	(0.050)
AME(sch)	0.138***	0.147***	0.153***	0.060***	0.069***	0.088***	0.041**	0.048***	0.072***
	(0.003)	(0.003)	(0.003)	(0.015)	(0.014)	(0.014)	(0.016)	(0.015)	(0.015)
Obs.	5,816	5,816	5,816	5,816	5,816	5816	5,816	5,816	5,816
R^2	0.59	0.62	0.73	0.59	0.62	0.74			

Notes: Copula() are Gaussian Copula functions. The Copula functions for schooling are positive and significant implying positive and significant correlation between schooling variables and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. A dummy for survey year, log number of weeks worked (for MonthlyA) and a constant were included in the regressions but are excluded from this table. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Effects of Aggregating Earnings on Estimates of Returns to Schooling in Tanzania

	OLS			GC			HGC		
Measure	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.029***	0.004	-0.017***	-0.033***	-0.031***	-0.033***	-0.028***	-0.010	-0.022**
	(0.006)	(0.007)	(0.006)	(0.010)	(0.012)	(0.010)	(0.009)	(0.012)	(0.010)
sch^2	0.009***	0.010***	0.009***	0.009***	0.010***	0.009***	0.008***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
age	0.075***	0.065***	0.067***	0.075***	0.065***	0.066***	0.064***	0.027***	0.046***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)	(0.005)
$age^2/100$	-0.079***	-0.067***	-0.070***	-0.079***	-0.067***	-0.070***	-0.065***	-0.017**	-0.043***
	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)
female	-0.445***	-0.641***	-0.554***	-0.445***	-0.642***	-0.549***	-0.349***	-0.303***	-0.370***
	(0.017)	(0.021)	(0.018)	(0.017)	(0.022)	(0.019)	(0.028)	(0.039)	(0.031)
rural	-0.171***	-0.664***	-0.339***	-0.171***	-0.662***	-0.330***	-0.136***	-0.538***	-0.270***
	(0.018)	(0.022)	(0.020)	(0.017)	(0.020)	(0.018)	(0.020)	(0.025)	(0.022)
Copula(sch)				-0.011	0.076**	0.039	-0.010	0.080**	0.041
				(0.027)	(0.033)	(0.028)	(0.027)	(0.032)	(0.028)
Copula(sch2)				0.032	0.084***	0.038	0.029	0.076**	0.035
				(0.026)	(0.033)	(0.028)	(0.026)	(0.033)	(0.029)
Copula(weeks)						0.029***			0.030***
						(0.005)			(0.005)
IMR							-0.237***	-0.837***	-0.446***
							(0.052)	(0.075)	(0.059)
AME(sch)	0.089***	0.142***	0.104***	0.084***	0.102***	0.083***	0.079***	0.086***	0.075***
	(0.002)	(0.003)	(0.002)	(0.009)	(0.011)	(0.009)	(0.009)	(0.011)	(0.010)
Obs.	11,215	11,215	11,215	11,215	11,215	11,215	11,215	11,215	11,215
R^2	0.27	0.37	0.78	0.27	0.37	0.78			

Notes: Copula() are Gaussian Copula functions. Significance of Copula functions for a variable implies a significant correlation between variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. A dummy for individual surveyed at least twice, log number of weeks worked (for MonthlyA) and a constant were included in the regressions but are excluded from this table. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Effects of Aggregating Earnings on Estimates of Returns to Schooling in Uganda

	OLS			GC			HGC		
Measure	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	0.061***	0.074***	0.069***	0.020	0.040*	0.030	0.019	0.041**	0.029
	(0.011)	(0.012)	(0.011)	(0.019)	(0.021)	(0.020)	(0.019)	(0.020)	(0.020)
sch^2	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
age	0.080***	0.077***	0.068***	0.081***	0.077***	0.068***	0.082***	0.076***	0.070***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
$age^2/100$	-0.090***	-0.088***	-0.077***	-0.090***	-0.088***	-0.077***	-0.092***	-0.086***	-0.079***
	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)
female	-0.443***	-0.451***	-0.438***	-0.439***	-0.448***	-0.435***	-0.458***	-0.422***	-0.459***
	(0.028)	(0.030)	(0.029)	(0.029)	(0.029)	(0.028)	(0.054)	(0.058)	(0.055)
rural	-0.224***	-0.293***	-0.239***	-0.221***	-0.290***	-0.239***	-0.237***	-0.267***	-0.259***
	(0.027)	(0.029)	(0.028)	(0.027)	(0.029)	(0.029)	(0.050)	(0.053)	(0.051)
Copula(sch)				0.151**	0.129*	0.144*	0.150**	0.129*	0.144*
				(0.073)	(0.076)	(0.074)	(0.072)	(0.076)	(0.074)
Copula(sch2)				0.045	0.034	0.042	0.043	0.036	0.040
				(0.073)	(0.079)	(0.077)	(0.071)	(0.076)	(0.074)
Copula(weeks)						-0.005			-0.005
						(0.006)			(0.006)
IMR							0.045	-0.061	0.055
							(0.113)	(0.119)	(0.114)
AME(sch)	0.113***	0.126***	0.122***	0.069***	0.090***	0.081***	0.072***	0.087***	0.084***
	(0.003)	(0.003)	(0.003)	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)	(0.019)
Obs.	4,631	4,631	4,631	4,631	4,631	4,631	4,631	4,631	4,631
R^2	0.36	0.38	0.60	0.36	0.38	0.60			

Notes: Copula() are Gaussian Copula functions. Significance of Copula functions for a variable implies a significant correlation between variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. A dummy for individual surveyed at least twice, log number of weeks worked (for MonthlyA) and a constant were included in the regressions but are excluded from this table. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3: Effects of Aggregating Earnings on Estimates of Returns to Schooling

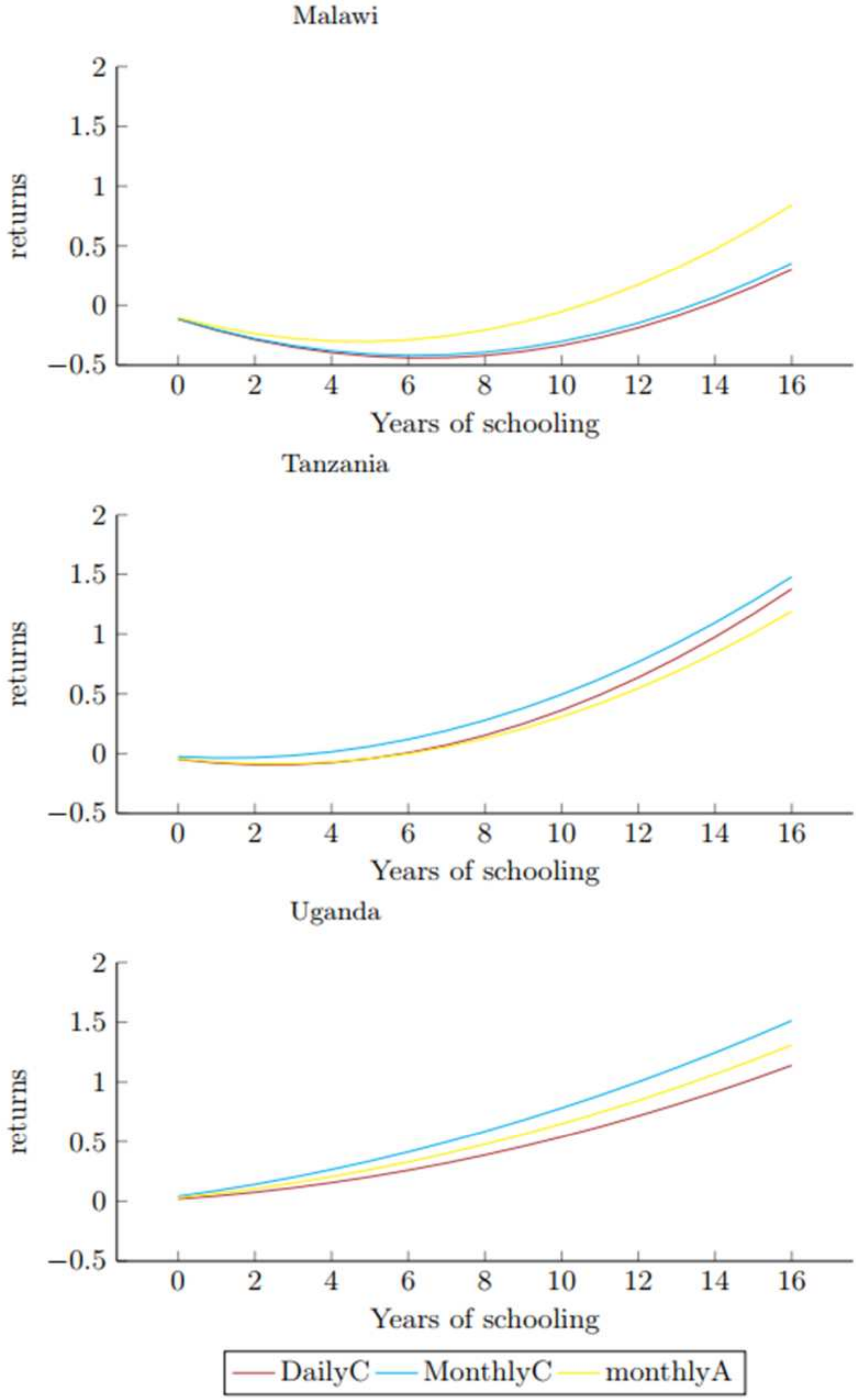


Table 7: Effects of Aggregating Earnings (Levels of Education) - Malawi

	OLS			Heckman		
Measure	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
primary	0.236***	0.274***	0.321***	0.180***	0.218***	0.283***
	(0.026)	(0.024)	(0.024)	(0.032)	(0.030)	(0.029)
secondary	0.693***	0.794***	0.851***	0.563***	0.663***	0.764***
	(0.027)	(0.025)	(0.025)	(0.051)	(0.048)	(0.047)
higher	1.443***	1.549***	1.601***	1.256***	1.359***	1.474***
	(0.032)	(0.030)	(0.030)	(0.070)	(0.066)	(0.065)
age	0.068***	0.063***	0.059***	0.057***	0.053***	0.052***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
age2/100	-0.069***	-0.062***	-0.057***	-0.058***	-0.050***	-0.049***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
female	-0.133***	-0.134***	-0.104***	-0.054	-0.055*	-0.051
	(0.023)	(0.022)	(0.021)	(0.035)	(0.033)	(0.032)
rural	-0.158***	-0.229***	-0.227***	-0.081**	-0.150***	-0.174***
	(0.021)	(0.019)	(0.019)	(0.033)	(0.031)	(0.031)
year	1.297***	1.201***	1.198***	1.316***	1.220***	1.210***
	(0.020)	(0.019)	(0.018)	(0.021)	(0.019)	(0.019)
weeks			1.118***			1.115***
			(0.018)			(0.018)
IMR				-0.168***	-0.170***	-0.115**
				(0.056)	(0.053)	(0.052)
constant	-0.445***	2.625***	-1.672***	-0.042	3.033***	-1.389***
	(0.113)	(0.106)	(0.116)	(0.176)	(0.165)	(0.173)
Obs.	5,816	5,816	5,816	5,816	5,816	5,816
R^2	0.59	0.62	0.73	0.59	0.62	0.74

Notes: IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tables 7-9 present results when using the level of education attained instead of completed years of schooling. GC and HGC are not suitable here, but we present results corrected for endogenous selection to the pay periods. The results are consistent with those from using years of schooling. Importantly, the different measures of earnings yield different estimates of the returns to the levels of education as observed when using years of schooling. MonthlyA gives larger estimates for Malawi and Uganda, while MonthlyC gives larger estimates for Tanzania. Higher levels of education are associated with higher returns, implying a convex relationship between earnings and education. Whether coefficients accounting for selection bias are higher or lower than OLS varies by pay period, level, and country.

Table 8: Effects of Aggregating Earnings (Levels of Education) - Tanzania

	OLS			Heckman		
Measure	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
primary	0.202***	0.402***	0.263***	0.200***	0.394***	0.261***
	(0.020)	(0.025)	(0.021)	(0.020)	(0.024)	(0.021)
secondary	0.797***	1.365***	0.962***	0.755***	1.217***	0.892***
	(0.027)	(0.033)	(0.029)	(0.028)	(0.035)	(0.031)
higher	1.613***	2.239***	1.762***	1.489***	1.798***	1.541***
	(0.048)	(0.059)	(0.051)	(0.055)	(0.068)	(0.059)
age	0.074***	0.063***	0.065***	0.063***	0.026***	0.046***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)	(0.005)
age2/100	-0.077***	-0.063***	-0.067***	-0.063***	-0.014*	-0.042***
	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)
female	-0.460***	-0.675***	-0.572***	-0.365***	-0.341***	-0.401***
	(0.017)	(0.021)	(0.018)	(0.027)	(0.034)	(0.029)
rural	-0.179***	-0.677***	-0.341***	-0.145***	-0.558***	-0.285***
	(0.018)	(0.022)	(0.020)	(0.019)	(0.024)	(0.021)
panel	-0.106***	-0.005	-0.090***	-0.097***	0.027	-0.072***
	(0.017)	(0.021)	(0.018)	(0.017)	(0.021)	(0.018)
weeks			1.132***			1.124***
			(0.008)			(0.008)
IMR				-0.233***	-0.824***	-0.427***
				(0.052)	(0.065)	(0.056)
constant	-0.117*	2.770***	-1.260***	0.260**	4.104***	-0.551***
	(0.071)	(0.089)	(0.078)	(0.111)	(0.137)	(0.122)
Obs.	11,215	11,215	11,215	11,215	11,215	11,215
R^2	0.26	0.36	0.78	0.26	0.37	0.78

Notes: IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Effects of Aggregating Earnings (Levels of Education) - Uganda

	OLS			Heckman		
Measure	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
primary	0.515***	0.570***	0.557***	0.515***	0.551***	0.559***
	(0.033)	(0.035)	(0.034)	(0.036)	(0.038)	(0.037)
secondary	0.634***	0.764***	0.719***	0.635***	0.746***	0.721***
	(0.044)	(0.046)	(0.044)	(0.046)	(0.049)	(0.047)
higher	1.271***	1.415***	1.367***	1.273***	1.339***	1.376***
	(0.039)	(0.041)	(0.040)	(0.076)	(0.081)	(0.078)
age	0.084***	0.082***	0.072***	0.084***	0.079***	0.072***
	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
age2/100	-0.097***	-0.096***	-0.084***	-0.098***	-0.091***	-0.085***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
female	-0.461***	-0.470***	-0.456***	-0.463***	-0.421***	-0.461***
	(0.029)	(0.031)	(0.030)	(0.051)	(0.054)	(0.052)
rural	-0.281***	-0.354***	-0.298***	-0.282***	-0.309***	-0.303***
	(0.028)	(0.029)	(0.028)	(0.047)	(0.050)	(0.048)
panel	0.158***	0.229***	0.167***	0.158***	0.233***	0.167***
	(0.029)	(0.031)	(0.030)	(0.029)	(0.031)	(0.030)
weeks			1.158***			1.158***
			(0.023)			(0.023)
IMR				0.003	-0.114	0.012
				(0.097)	(0.103)	(0.099)
constant	-0.742***	2.246***	-1.920***	-0.747***	2.405***	-1.938***
	(0.124)	(0.132)	(0.141)	(0.185)	(0.195)	(0.201)
Obs.	4,631	4,631	4,631	4631	4631	4631
R^2	0.34	0.36	0.58	0.34	0.36	0.58

Notes: IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.2. Effects of Pay Period on Estimates of Returns to Schooling

The results presented in section 5.1 show that in converting reported earnings to a common unit and pooling, MonthlyC leads to larger estimates of returns to education in Tanzania (compared to DailyC or MonthlyA) while for Malawi and Uganda MonthlyA leads to larger estimates. In this section, we present the results for the samples for each of the pay periods to see whether the returns to education vary depending on the period in which workers are paid. Owing to its ability to allow for seasonal workers who only work some months in a year and some weeks in a month, we choose MonthlyA as our preferred common earnings measure and use it in examining returns by pay period. We focus discussion on the endogenous corrected (both for ability and selection) results (HGC) but include OLS results for comparison. The corresponding GC results and results for the first stage HGC regressions are in Appendixes C and D with corresponding bootstrap aggregating results in Appendix E.

Table 10 shows results by pay period for Malawi. The last three columns show that even after correcting for ability bias and selection, the coefficient of schooling is negative and significant across the pay periods - there is a threshold below which the returns to education are negative. This threshold varies by pay period: four years for workers reporting earnings daily and monthly, and seven years for those reporting earnings weekly.

Comparing different pay periods, reporting earnings daily is associated with higher returns to education than reporting weekly and monthly. The average marginal effects indicate that an extra year of schooling raises earnings by 12% for daily, 5.7% for weekly, and 8.4% for monthly (although the coefficients are not significant for daily and weekly). After disaggregating the sample to pay periods, the coefficients of the Copula functions and the inverse mills ratio are insignificant for daily and weekly suggesting that the correlation between earnings and the error terms observed earlier are associated with only the monthly sample. The implied returns from Table 10 for the selected years of schooling are shown graphically in the top panel of Figure 4. The naïve estimates (pooled) from the last column of Table 4 are also included for comparison. Except for monthly, the pattern and slope of the curves for each pay period are different from that of the pooled curve implying that each period has different returns to education and ignoring this would lead to biased estimates. For monthly, it can be explained by the fact that it constitutes about 88% of the sample and thus pooling the periods together would very likely bias the returns in the direction of monthly.

The coefficient of schooling for Tanzania is also negative throughout (Table 11); again, there is a threshold below which the returns to education are negative (although in this case it is only a few years of education). The correlation between schooling and the error terms is significant only for the monthly sample. When ability and selection biases are accounted for, there are mixed results: the estimates of returns for the monthly decrease while for daily and weekly increase (see the AME(sch) in Table 11). This suggests that the way endogeneity affects OLS results is not homogenous across the pay periods. For instance, unlike OLS, HGC results show that returns for monthly are lower than for daily earners and the difference increases with education (consistent with a particular level of education needed to secure a job paid monthly but does not then affect earnings). This indicates that selection was biasing the returns to schooling downwards for the daily, while for monthly selection was biasing the returns upwards.

Table 10: Effects of Pay Period on Estimates of Returns to Schooling in Malawi

		OLS			HGC	
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	-0.002	-0.095***	-0.066***	-0.018	-0.133**	-0.091***
	(0.076)	(0.030)	(0.008)	(0.088)	(0.052)	(0.010)
sch2	0.006	0.013***	0.012***	0.007	0.012***	0.009***
	(0.005)	(0.002)	(0.000)	(0.007)	(0.002)	(0.001)
age	0.041	0.076***	0.053***	0.060	0.079***	0.045***
	(0.059)	(0.021)	(0.006)	(0.080)	(0.028)	(0.006)
age2/100	-0.035	-0.081***	-0.048***	-0.061	-0.083**	-0.039***
	(0.074)	(0.027)	(0.007)	(0.100)	(0.034)	(0.008)
female	0.014	-0.098	-0.116***	-0.107	-0.129	-0.041
	(0.210)	(0.078)	(0.022)	(0.289)	(0.105)	(0.033)
rural	-0.286	-0.322***	-0.185***	-0.297	-0.308***	-0.106***
	(0.204)	(0.079)	(0.019)	(0.227)	(0.091)	(0.033)
year	0.850***	1.020***	1.278***	0.971***	1.002***	1.301***
	(0.212)	(0.073)	(0.018)	(0.358)	(0.083)	(0.020)
weeks	1.259***	1.037***	1.138***	1.371***	1.226***	1.142***
	(0.166)	(0.052)	(0.019)	(0.221)	(0.081)	(0.029)
Copula(sch)				-0.497	-0.046	0.095**
				(0.463)	(0.235)	(0.048)
Copula(sch2)				0.534	0.277	0.134***
				(0.510)	(0.202)	(0.045)
Copula(weeks)				-0.025	-0.064***	-0.001
				(0.045)	(0.017)	(0.005)
IMR				0.327	0.020	-0.163***
				(0.843)	(0.271)	(0.052)
Constant	-2.010	-1.224***	-1.834***	-3.664	-1.467	-1.003***
	(1.256)	(0.419)	(0.121)	(3.887)	(1.144)	(0.223)
AME(sch)	0.101***	0.113***	0.161***	0.117	0.057	0.084***
	(0.026)	(0.009)	(0.003)	(0.124)	(0.057)	(0.015)
Obs.	182	505	5,129	182	505	5,129
R^2	0.44	0.66	0.77			

Notes: Copula() are Gaussian Copula functions; significance of the functions implies a significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Effects of Pay Period on Estimates of Returns to Schooling in Tanzania

	OLS			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	-0.023	-0.026	0.033***	-0.045**	-0.033	-0.015
	(0.015)	(0.017)	(0.009)	(0.018)	(0.020)	(0.016)
sch2	0.006***	0.009***	0.007***	0.010***	0.010***	0.005***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
age	0.054***	0.046***	0.098***	0.027***	0.040***	0.085***
	(0.008)	(0.009)	(0.006)	(0.010)	(0.014)	(0.006)
age2/100	-0.066***	-0.053***	-0.093***	-0.027*	-0.044**	-0.077***
	(0.011)	(0.012)	(0.008)	(0.015)	(0.019)	(0.008)
female	-0.751***	-0.555***	-0.331***	-0.494***	-0.502***	-0.207***
	(0.036)	(0.040)	(0.024)	(0.063)	(0.089)	(0.031)
rural	-0.537***	-0.228***	-0.258***	-0.547***	-0.211***	-0.088**
	(0.041)	(0.049)	(0.023)	(0.043)	(0.062)	(0.034)
panel	-0.219***	-0.127***	-0.067***	-0.172***	-0.127**	-0.074***
	(0.039)	(0.048)	(0.024)	(0.039)	(0.051)	(0.023)
weeks	1.165***	1.096***	1.073***	1.068***	0.997***	1.036***
	(0.015)	(0.015)	(0.017)	(0.021)	(0.023)	(0.024)
Copula(sch)				-0.016	-0.027	0.038**
				(0.044)	(0.047)	(0.042)
Copula(sch2)				0.036	0.026	0.101**
				(0.045)	(0.046)	(0.042)
Copula(weeks)				0.073***	0.096***	0.009
				(0.013)	(0.018)	(0.006)
IMR				-0.707***	-0.159	-0.517***
				(0.153)	(0.247)	(0.070)
constant	-0.722***	-0.785***	-2.345***	0.894**	-0.250	-0.796***
	(0.148)	(0.165)	(0.113)	(0.364)	(0.648)	(0.230)
AME(sch)	0.050***	0.073***	0.148***	0.069***	0.077**	0.063**
	(0.006)	(0.007)	(0.003)	(0.020)	(0.021)	(0.015)
Obs.	3,738	1,929	4,830	3,738	1,929	4,830
R^2	0.73	0.79	0.71			

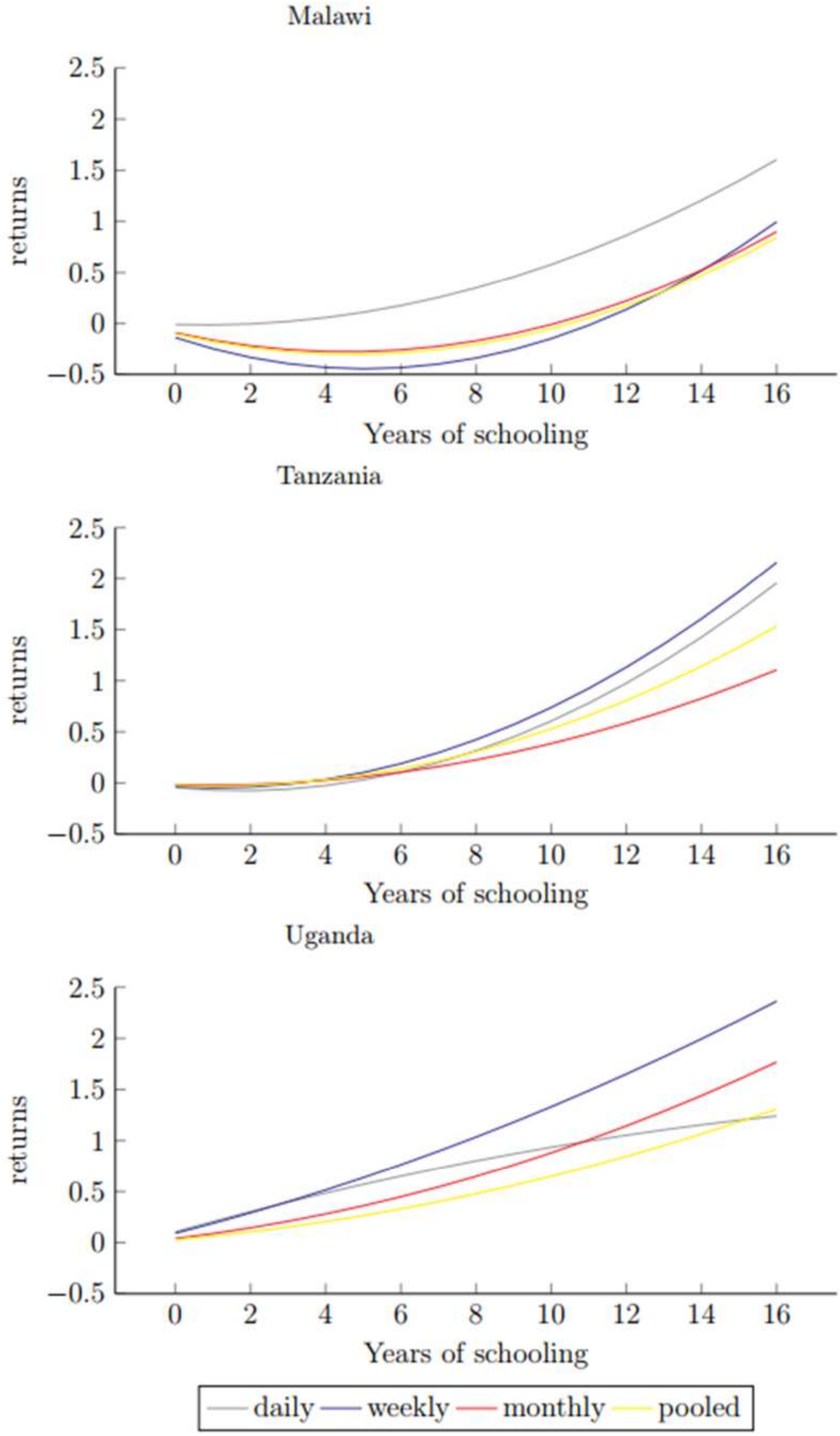
Notes: As for Table 10

Table 12: Effects of Pay Period on Estimates of Returns to Schooling in Uganda

	OLS			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	0.120***	0.058*	0.064***	0.105*	0.091	0.040*
	(0.023)	(0.033)	(0.015)	(0.061)	(0.079)	(0.022)
sch2	-0.002	0.003	0.004***	-0.002	0.003	0.004***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
age	0.063***	0.077***	0.080***	0.065***	0.075***	0.082***
	(0.013)	(0.019)	(0.010)	(0.014)	(0.020)	(0.012)
age2/100	-0.082***	-0.090***	-0.084***	-0.088***	-0.084***	-0.086***
	(0.018)	(0.025)	(0.013)	(0.020)	(0.027)	(0.015)
female	-0.596***	-0.535***	-0.312***	-0.693***	-0.415**	-0.328***
	(0.066)	(0.087)	(0.035)	(0.138)	(0.181)	(0.053)
rural	-0.248***	-0.358***	-0.232***	-0.303***	-0.317***	-0.251***
	(0.055)	(0.087)	(0.034)	(0.101)	(0.101)	(0.059)
panel	0.015	-0.063	0.176***	0.013	-0.067	0.171***
	(0.056)	(0.102)	(0.037)	(0.054)	(0.080)	(0.042)
weeks	1.237***	1.093***	1.133***	1.166***	1.153***	1.164***
	(0.039)	(0.054)	(0.031)	(0.067)	(0.090)	(0.051)
Copula(sch)				0.136	-0.072	0.052
				(0.162)	(0.257)	(0.066)
Copula(sch2)				-0.082	-0.079	0.107
				(0.164)	(0.209)	(0.081)
Copula(weeks)				0.024*	-0.020	-0.007
				(0.013)	(0.020)	(0.008)
IMR				0.173	-0.348	0.069
				(0.203)	(0.416)	(0.181)
constant	-2.106***	-1.790***	-2.655***	-2.084***	-1.448	-2.592***
	(0.260)	(0.396)	(0.191)	(0.569)	(1.110)	(0.482)
AME(sch)	0.096***	0.097***	0.145***	0.076	0.137*	0.116***
	(0.008)	(0.011)	(0.004)	(0.011)	(0.074)	(0.026)
N	1,262	589	2,765	1,262	589	2,765
R^2	0.57	0.57	0.63			

Notes: As for Table 10.

Figure 4:Effects of Pay Period on Estimates of Returns to Schooling



To account for the effect of the sch^2 term, returns over the range of years of education are shown in Figure 4. The middle panel plots the implied returns for Tanzania from Table 11 (marginal effects of schooling on wage, pooled estimates derived from Table 8). Returns for the weekly earners are not only higher but also increase at a higher rate than the other periods (reflecting the higher coefficient on sch^2).

Table 12 shows the effects of the pay period on the estimates of returns to schooling in Uganda. While the coefficients on *sch* are positive across the pay periods, the coefficient on sch^2 for those reporting daily earnings is negative, implying concave returns (returns to an extra year of education decrease as one acquires more schooling). The concavity persists even after accounting for ability bias and selection. The bottom panel of Figure 4 plots the implied returns. The patterns of the curves for each pay period are very different from that of the pooled curve, implying that each pay period has different returns to education.

Tables 13–15 show the corresponding results for levels of education. The pattern of the returns is mixed in Malawi: the returns to primary education are highest if reporting earnings monthly; returns to secondary education are highest if reporting earnings daily; and returns to higher education are highest if reporting earnings weekly (Table 13). Like for the years of schooling, we do not find evidence of significant selection problems for daily and weekly, although it might mean that the sample sizes are too small to detect it. Recall that *ganyu* workers (the majority by far and the lowest paid) are excluded so Malawi is not fully comparable to Tanzania and Uganda. Given the relatively high earnings, it is possible that daily and weekly samples in Malawi include professionals or self-employed with relatively high earnings and education.

The pattern of results for levels of education in Tanzania (Table 14) are similar to Table 11. The returns to the levels of education differ by pay period and weekly have higher returns than their daily and monthly counterparts. Compared to those for daily and weekly, the results for monthly are closer to the results for the pooled sample reported earlier in Table 8. This may suggest that the larger monthly sample biases the pooled results in its direction. Results for the levels of education by pay period in Uganda (Table 15) are inconsistent with those obtained using years of schooling (Table 12). While the latter show a concave relationship between earnings and education, the former show a convex relationship, i.e., returns to education increase with the levels of education. This may be because of relatively few observations at high years of education, so estimates are imprecise, exacerbated by the (negative) sch^2 effect and perhaps some of those with more education are 'waiting' to get into formal monthly wage work. Our data do not provide enough information to investigate this issue further.

Table 13:Effects of Pay Period on Estimates of Returns to Levels of Education in Malawi

	OLS			Heckman		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
primary	0.260	0.135	0.355***	0.315	0.139	0.313***
	(0.233)	(0.083)	(0.024)	(0.298)	(0.089)	(0.030)
secondary	0.684***	0.826***	0.885***	0.795*	0.830***	0.788***
	(0.249)	(0.107)	(0.025)	(0.452)	(0.115)	(0.049)
higher	1.067***	1.643***	1.648***	1.224**	1.651***	1.510***
	(0.299)	(0.139)	(0.029)	(0.612)	(0.158)	(0.067)
age	0.036	0.082***	0.057***	0.049	0.084***	0.050***
	(0.059)	(0.020)	(0.006)	(0.073)	(0.024)	(0.006)
age2/100	-0.032	-0.093***	-0.053***	-0.047	-0.094***	-0.045***
	(0.074)	(0.026)	(0.007)	(0.090)	(0.031)	(0.008)
female	0.006	-0.104	-0.114***	-0.057	-0.112	-0.056*
	(0.210)	(0.077)	(0.021)	(0.301)	(0.109)	(0.033)
rural	-0.307	-0.375***	-0.205***	-0.307	-0.380***	-0.145***
	(0.204)	(0.078)	(0.019)	(0.204)	(0.090)	(0.032)
year	0.866***	1.014***	1.233***	0.938***	1.017***	1.253***
	(0.212)	(0.072)	(0.018)	(0.325)	(0.078)	(0.020)
weeks	1.265***	1.020***	1.136***	1.263***	1.020***	1.132***
	(0.167)	(0.051)	(0.019)	(0.167)	(0.052)	(0.019)
IMR				0.244	0.029	-0.125**
				(0.831)	(0.268)	(0.054)
Constant	-1.708	-1.176***	-1.765***	-2.710	-1.273	-1.456***
	(1.215)	(0.402)	(0.117)	(3.616)	(0.970)	(0.178)
Obs.	182	505	5129	182	505	5129
R^2	0.44	0.68	0.77	0.45	0.64	0.76

Notes: IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Effects of Pay Period on Estimates of Returns to Levels of Education - Tanzania

	OLS			Heckman		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
primary	0.143***	0.295***	0.354***	0.225***	0.311***	0.226***
	(0.036)	(0.040)	(0.036)	(0.040)	(0.047)	(0.040)
secondary	0.425***	0.809***	1.159***	0.753***	0.864***	0.819***
	(0.077)	(0.085)	(0.039)	(0.102)	(0.122)	(0.064)
higher			1.892***			1.337***
			(0.052)			(0.098)
age	0.053***	0.043***	0.095***	0.027***	0.037***	0.085***
	(0.008)	(0.009)	(0.006)	(0.010)	(0.012)	(0.006)
age2/100	-0.065***	-0.049***	-0.087***	-0.028**	-0.041**	-0.076***
	(0.011)	(0.012)	(0.008)	(0.013)	(0.017)	(0.008)
female	-0.765***	-0.569***	-0.352***	-0.534***	-0.525***	-0.240***
	(0.036)	(0.039)	(0.024)	(0.059)	(0.079)	(0.029)
rural	-0.541***	-0.233***	-0.265***	-0.570***	-0.251***	-0.118***
	(0.041)	(0.049)	(0.023)	(0.042)	(0.057)	(0.032)
panel	-0.220***	-0.137***	-0.086***	-0.179***	-0.136***	-0.087***
	(0.040)	(0.049)	(0.024)	(0.040)	(0.049)	(0.024)
weeks	1.170***	1.099***	1.083***	1.162***	1.099***	1.076***
	(0.015)	(0.015)	(0.017)	(0.015)	(0.015)	(0.017)
IMR				-0.681***	-0.135	-0.443***
				(0.139)	(0.214)	(0.066)
constant	-0.677***	-0.702***	-2.117***	0.733**	-0.274	-1.208***
	(0.146)	(0.162)	(0.111)	(0.322)	(0.574)	(0.175)
Obs.	3,738	1,929	4,830	3,738	1,929	4,830
R^2	0.73	0.79	0.71	0.73	0.79	0.71

Notes: IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Effects of Pay Period (Levels of Education) - Uganda

	OLS			Heckman		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly
primary	0.466***	0.439***	0.650***	0.462***	0.425***	0.666***
	(0.060)	(0.088)	(0.047)	(0.061)	(0.089)	(0.063)
secondary	0.628***	0.476***	0.866***	0.609***	0.491***	0.888***
	(0.092)	(0.133)	(0.057)	(0.103)	(0.134)	(0.080)
higher	1.009***	1.230***	1.481***	0.964***	1.251***	1.539***
	(0.161)	(0.157)	(0.049)	(0.198)	(0.159)	(0.157)
age	0.067***	0.075***	0.086***	0.068***	0.072***	0.087***
	(0.014)	(0.019)	(0.010)	(0.014)	(0.020)	(0.011)
age2/100	-0.091***	-0.091***	-0.093***	-0.094***	-0.083***	-0.095***
	(0.019)	(0.025)	(0.013)	(0.019)	(0.027)	(0.014)
female	-0.678***	-0.586***	-0.318***	-0.721***	-0.464***	-0.332***
	(0.065)	(0.087)	(0.036)	(0.126)	(0.158)	(0.051)
rural	-0.297***	-0.460***	-0.292***	-0.327***	-0.413***	-0.310***
	(0.056)	(0.088)	(0.035)	(0.093)	(0.102)	(0.058)
panel	0.001	-0.042	0.209***	0.003	-0.046	0.204***
	(0.057)	(0.103)	(0.038)	(0.057)	(0.103)	(0.041)
weeks	1.238***	1.083***	1.140***	1.239***	1.081***	1.140***
	(0.039)	(0.055)	(0.032)	(0.039)	(0.055)	(0.032)
IMR				0.075	-0.326	0.055
				(0.189)	(0.354)	(0.143)
constant	-1.678***	-1.326***	-2.414***	-1.796***	-0.600	-2.531***
	(0.254)	(0.379)	(0.191)	(0.392)	(0.875)	(0.358)
Obs.	1,262	589	2,765	1,262	589	2,765
R^2	0.55	0.55	0.60	0.55	0.56	0.60

Notes: IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results for the probit estimates (the first stage results for the selection model) for the probabilities of participating in each of the periods of payment are reported in Appendix D. Importantly for Tanzania (Table D.2) the coefficient on sch^2 for daily and weekly is negative and statistically significant, implying that an extra year of schooling reduces participation in daily and weekly employment. For monthly, the coefficient is positive and significant, as might be expected. A minimum level of education seems essential for gaining employment in monthly paid jobs but, conditional on securing such jobs, the marginal effect of schooling on wages is lower than for weekly or daily paid work (which have lower education entry requirements).

Table D.3 (the first stage regression for HGC) shows that an extra year of schooling reduces participation in daily and weekly employment (as indicated by the negative sign of the coefficient on sch^2) but increases participation into monthly employment in Uganda. Again, a minimum level of education is essential for gaining employment in monthly paid jobs; conditional on securing such jobs, the marginal effect of schooling on wages is lower than for weekly but higher than for daily paid work (not quite the same as in Tanzania).

5.3. Measuring Returns for Casual Employment: A Case of *Ganyu* in Malawi

This section presents the results for returns to education for *ganyu* labour in Malawi using the different measures of earnings. After data cleaning, the final sample consists of 16,528 individuals who participated only in *ganyu* labour as their primary source of labour earnings. As expected, a disproportionately large proportion of *ganyu* workers reside in rural areas (91.7% rural vs 8.3% urban), and the distribution has remained relatively stable over the survey years (91.0% rural vs 9.0% urban in 2010; and 92.1% rural vs 7.9% urban in 2016). However, there were relatively more *ganyu* workers in 2016 than 2010 (63.7% of the *ganyu* workers are from the 2016 survey). In terms of educational attainment, about 21% of *ganyu* workers had never gone to school, and less than 1% had more than secondary education (more than 12 years of education). This suggests that generally, more schooling reduces the likelihood of participating in *ganyu* labour.

Constructing the pay period is different for *ganyu* (see Appendix B) such workers report an average daily wage and DailyC is this figure (irrespective of how many days worked). MonthlyC incorporates data on the average number of days in a week and weeks in a month while MonthlyA also accounts for the number of months worked in *ganyu* over the year.Table 16 shows the distribution of *ganyu* earnings by the different earnings measures. Due to the nature of *ganyu* labour supply, because *ganyu* workers rarely work for the whole month or year, it is likely that DailyC and MonthlyC overestimate earnings (the $5.8 DailyC for 22 days implies $127 per month). In contrast, MonthlyA allows for the fact they do not work for the full year. Figure 5 shows the kernel distribution of earnings corresponding to Table 16. It does appear that the smooth distribution for MonthlyA removes possible extremes of DailyC and MonthlyC.

Table 16: Distribution of Ganyu Earnings ($US and LCU)

Measure		DailyC		MonthlyC		MonthlyA	
	Obs.	Mean	SD	Mean	SD	Mean	SD
$	16,528	5.76	6.93	69.41	101.72	33.76	53.52
LCU	16,528	812.51	977.83	9797.99	1,4360.21	4,766.48	7,555.73

Source: Author's computations from IHS. Earnings in $ are in constant 2009 exchange rate (1$= 141.17 Malawi Kwacha)

Figure 5: Distribution of *Ganyu* Earnings by Earnings Measure

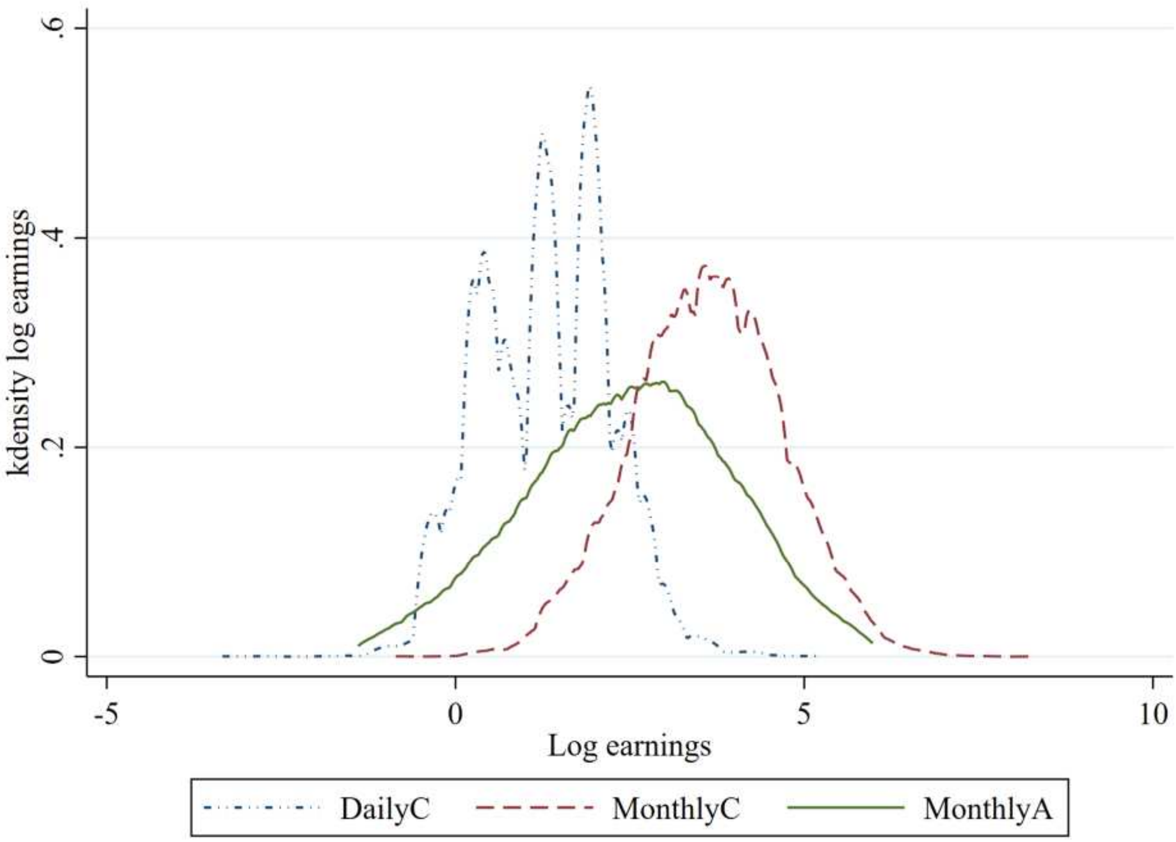


Table 17 shows returns to years of education by the measure of earnings. Generally, MontlyC yields larger estimates of returns to education than DailyC or MonthlyA. Furthermore, the difference between estimates from DailyC and MonthlyA is small, suggesting that converting to MonthlyC gives upward biased estimates. There does not seem to be a critical endogeneity of education, as indicated by the insignificance of the copula function for the education variables. The implied returns to education from Table 17 are shown graphically in Figure 6.

Table 18 shows the corresponding results for levels of education. Because there is a very small proportion of workers with higher education doing *ganyu* labour, we will reserve the discussion on returns to higher education. As can be seen, the results in Table 18 are consistent with those in Table 17 in the sense that the three earnings measures yield different returns to education and MonthlyC results in higher estimates. In addition, the signs, pattern and significance of the inverse mills ratios are similar to those in Table 17, suggesting that the estimates are precise thanks to the large sample.

Table 17:Earnings Measures and Returns to years of Schooling for *Ganyu* Labour

	OLS			HGC		
Measure	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	0.004	-0.018***	-0.000	0.006	-0.014	-0.005
	(0.004)	(0.006)	(0.005)	(0.008)	(0.011)	(0.010)
sch2	0.002***	0.003***	0.002***	0.002***	0.005***	0.003***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
age	0.034***	0.047***	0.040***	0.032***	0.039***	0.036***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)
age2/100	-0.040***	-0.057***	-0.048***	-0.037***	-0.043***	-0.040***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)
female	-0.240***	-0.444***	-0.341***	-0.223***	-0.367***	-0.299***
	(0.010)	(0.014)	(0.012)	(0.012)	(0.017)	(0.014)
rural	-0.284***	-0.420***	-0.331***	-0.330***	-0.624***	-0.440***
	(0.018)	(0.025)	(0.021)	(0.029)	(0.039)	(0.034)
year	1.326***	1.223***	1.218***	1.298***	1.097***	1.150***
	(0.010)	(0.013)	(0.012)	(0.015)	(0.022)	(0.018)
weeks			1.006***			0.981***
			(0.006)			(0.011)
Copula(sch)				-0.011	0.008	0.012
				(0.020)	(0.026)	(0.023)
Copula(sch2)				-0.003	-0.041	-0.003
				(0.019)	(0.027)	(0.023)
Copula(weeks)						0.019**
						(0.008)
IMR				-0.096**	-0.430***	-0.237***
				(0.042)	(0.059)	(0.050)
Constant	0.192***	2.642***	-0.912***	0.317***	3.237***	-0.495***
	(0.047)	(0.065)	(0.056)	(0.091)	(0.130)	(0.111)
AME(sch)	0.020	0.009***	0.018***	0.028***	0.035***	0.025***
	(0.001)	(0.002)	(0.002)	(0.007)	(0.010)	(0.009)
Obs.	16,528	16,528	16,528	16,528	16,528	16,528
R^2	0.60	0.38	0.77			

Note: The Copula() functions for schooling are positive (but insignificant) implying positive (but insignificant) correlation between schooling variables and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 6: Earnings Measures and Returns to Years of Education for *Ganyu* Labour

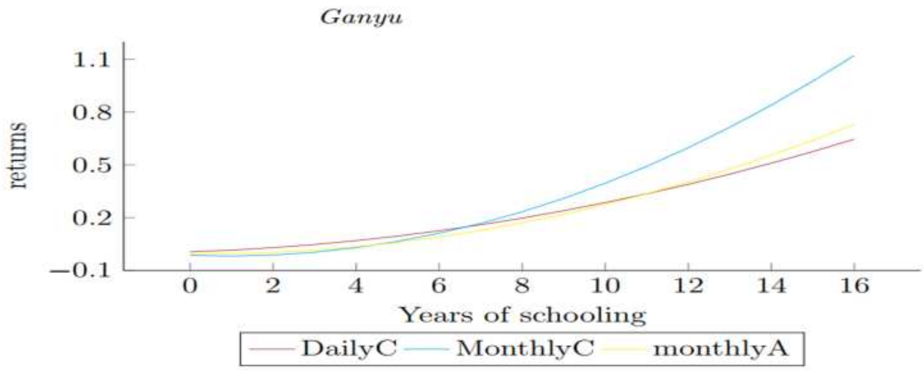


Table 18: Earnings Measures and Returns to Levels of Education for *Ganyu* Labour

	OLS			Heckman		
Measure	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
primary	0.122***	0.095***	0.122***	0.149***	0.207***	0.187***
	(0.014)	(0.019)	(0.016)	(0.017)	(0.023)	(0.020)
secondary	0.148***	0.109**	0.133***	0.219***	0.402***	0.303***
	(0.031)	(0.043)	(0.036)	(0.041)	(0.057)	(0.048)
higher	1.150***	1.282***	1.292***	1.282***	1.834***	1.611***
	(0.107)	(0.148)	(0.124)	(0.118)	(0.163)	(0.138)
age	0.034***	0.047***	0.040***	0.032***	0.039***	0.036***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
age2/100	-0.041***	-0.057***	-0.049***	-0.037***	-0.044***	-0.041***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)
female	-0.256***	-0.447***	-0.355***	-0.238***	-0.373***	-0.313***
	(0.010)	(0.013)	(0.011)	(0.012)	(0.016)	(0.014)
rural	-0.289***	-0.415***	-0.333***	-0.341***	-0.629***	-0.458***
	(0.018)	(0.025)	(0.021)	(0.027)	(0.037)	(0.031)
year	1.336***	1.223***	1.227***	1.306***	1.097***	1.155***
	(0.010)	(0.013)	(0.012)	(0.015)	(0.021)	(0.018)
weeks			1.004***			1.001***
			(0.006)			(0.006)
IMR				-0.106***	-0.443***	-0.257***
				(0.041)	(0.056)	(0.047)
constant	0.277***	2.629***	-0.846***	0.448***	3.339***	-0.429***
	(0.045)	(0.062)	(0.054)	(0.079)	(0.109)	(0.094)
Obs.	16,528	16,528	16,528	16,528	16,528	16,528
R^2	0.56	0.39	0.77	0.56	0.39	0.77

Note: IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6. Conclusions

This paper presents estimates of returns to schooling in Malawi, Tanzania and Uganda using nationally representative and comparable data from the Living Standard Measurement Study. Of interest was whether the relationship between earnings and education varies across workers reporting earnings over different pay periods, and results suggest that this is the case. This is the first comparative study on Africa to examine this issue using large survey datasets. Controlling for endogeneity due to unobserved ability using Gaussian Copula and Heckman method to account for selection, returns to education differ by pay period and pooling the periods together leads to imprecise estimates. Pooling earnings aggregated to different common pay period measures yields different estimates of returns to education and that estimates are generally biased in the direction of the pay period that constitutes the largest proportion of the sample (monthly wages). In this regard, the analysis suggests that estimating returns separately for workers paid over different periods is more reliable than pooling.

Comparing the differences according to a single estimate of the return per year of education is complicated as returns are evidently non-linear (schooling-squared is usually significant). In Tanzania and Malawi the coefficient on schooling is negative which, combined with the positive coefficient on schooling-squared, implies a threshold before returns become positive (two years of school in Tanzania but eight in Malawi). Plotting the non-linear relationship according to years of schooling shows that returns to education do vary according to the period of payment and how they vary differs across the three countries. Specifically, in Malawi the returns for non-*ganyu* workers reporting earnings daily are the highest, followed by monthly and then weekly; in Tanzania, the returns for weekly are not only higher but also increase at a higher rate than for the other pay periods; and in Uganda, returns are highest for weekly followed by monthly and then daily.

The magnitude of the effect of pay periods can be gauged by comparing estimates of returns to the level of education attained (controlling for selection but not endogeneity). For all three countries, of the common conversions DailyC always gives the lowest estimated return to each level of education but there is variation in which monthly measure gives the highest return. Differences in estimates are greatest in Tanzania and lowest in Uganda, and generally decrease with the level of education. In Malawi, returns to primary vary between 0.18 and 0.28 (a 55% difference), to secondary between 0.56 and 0.76 (36% gap) and to higher from 1.25 to 1.47 (18%). Returns are higher and the differences greater in Tanzania: for primary between 0.20 and 0.38 (a 95% difference), to secondary between 0.76 and 1.22 (60% gap) and higher from 1.25 to 1.47 (20%). Differences are much smaller in Uganda: returns to primary are higher than in the other countries and only vary between 0.52 and 0.56 (8%), returns to secondary vary between 0.64 and 0.75 (17%) and to higher from 1.27 to 1.38 (9%).

Estimating by pay period for levels of education shows that returns to higher education are generally highest in Malawi, returns to primary are highest in Uganda, Tanzania generally has the highest return to secondary and there are large differences across pay periods. Those paid weekly have the highest return to primary (30% compared to 23% for daily and monthly) and secondary (86% with 82% for monthly and 75% for daily) in Tanzania and only those paid monthly have returns from higher education (134% - too few observations to estimate for other pay periods). Those paid monthly have the highest returns in Uganda: 76% from primary (compared to 43-46%), 89% for secondary (compared to 49-61%) and 154% for higher (96-125%). In Malawi, estimated returns to primary are only significant for monthly (31%), returns are similar for secondary (79-83%) and weekly have the highest return on higher education (165%, with 151% for monthly and 122% for daily); *ganyu* workers have lower returns except if they have higher education.

Taking the results overall and acknowledging variation, we can be reasonably confident that returns to primary education are 40-70% in Uganda and 20-30% in Malawi and Tanzania. Returns to secondary education are about 80% in Malawi and Tanzania but vary by pay period between 90% (monthly) and 50% (weekly) in Uganda. Returns to higher education are 130% in Tanzania (monthly only), 100-150% in Uganda and 120-165% in Malawi.

The finding that three common measures of earnings give different estimates of returns to education are intuitive. Given the seasonality of casual work, earnings measures that allow for workers who do not work all weeks in the month and for seasonal workers who only work some months in a year are more reliable than measures that do not. There may be different reasons why workers report earnings daily, weekly or monthly. It cannot be assumed that workers who are paid monthly are necessarily in formal sector jobs, or that those paid over shorter periods are not. Unfortunately, the data do not permit a clear distinction between formal sector and informal employment, and many with monthly wages have relatively low earnings. One may expect those paid daily and weekly to be more likely to be in informal jobs but some of these have relatively high earnings (especially weekly). Further research is necessary to relate pay periods to formal or informal employment. It is likely that a minimum level of education is essential for gaining employment in the formal sector, and that such workers are paid monthly. Conditional on securing such jobs the marginal effect of schooling on wages may be lower than for the informal sector.

It is worth pointing out that since in Malawi *ganyu* labour is treated separately, regular employment is likely to constitute the better educated and hence better paid who are more likely to be the formal sector. Many of the workers reporting earnings hourly, daily and weekly in Tanzania and Uganda may have been in *ganyu* labour had they been residing in Malawi, and vice versa. Given this characteristic of Malawi's labour market, comparing the results with those for Tanzania and Uganda needs to be done with caution. This clearly deserves a further and independent investigation but is beyond the scope of this paper.

This paper is the first study in Sub-Saharan Africa to empirically show that the pay period matters, and care should be taken when estimating returns to education for workers paid over different pay periods. Building on the findings of this paper, further exploration on the topic could be carried out using data covering more countries and with more detailed labour market information. This should help to propose standard adjustment factors that could be used by all researchers to convert earnings from one period to another or pooling for comparison, thereby making studies across the region more comparable. Better data on the frequency of work for those paid hourly or daily or weekly would be useful (we incorporate such data as are available).

The findings from this paper suggest that increasing the share of the population in wage employment (especially formal employment) with monthly payments may lead to higher earnings, if not higher returns to education. Future research could extend the current analysis to investigate the differences in returns to education between workers paid monthly who work in the formal sector and those in the informal sector. The main conclusion is that it is challenging and inappropriate to report a single estimate for the return to a year of schooling: returns to education are non-linear and vary according to the period of payment and conversion to a common measure.

REFERENCES

Al-Samarrai, S., & Reilly, B. (2008). Education, employment and earnings of secondary school and university leavers in Tanzania: Evidence from a tracer study. *Journal of Development Studies*, *44*(2), 258–288. https://doi.org/10.1080/00220380701789950

Aslam, M., Bari, F., & Kingdon, G. (2012). Returns to schooling, ability and cognitive skills in Pakistan. *Education Economics*, *20*(2), 139–173. https://doi.org/10.1080/09645292.2010.488470

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to schooling*. Columbia University Press, New York.

Borjas, G. J. (2016). *Labor economics* (7th ed.). McGraw-Hill Education, New York.

Cameron, C. A., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

Cameron, C. A., & Trivedi, P. K. (2009). *Microeconometrics using Stata*. Stata Press, Texas.

Card, D. (1999). The Causal Effect of Education on Earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labour Economics* (Vol.3, pp. 1801–1863). Elsevier B.V. https://doi.org/10.1016/S1573-4463(99)03011-4

Card, D. (2001). Estimating the returns to schooling: progress on some persistent econometric problems. *Econometrica*, *69*(5), 1127–1160.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). Copula Methods in Finance. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9). John Wiley & Sons Ltd.

CHuang, Y., & Lai, W. (2017). Returns to Human Capital and Wage Inequality: The Case of Taiwan. *Journal of Economic Development*, *42*(3), 61–88.

Churchill, S. A., & Mishra, V. (2018). Returns to education in China : a meta-analysis. *Applied Economics*, *50*(54), 5903–5919. https://doi.org/10.1080/00036846.2018.1488074

Ebbes, P., Wedel, M., Böckenholt, U., & Steerneman, T. (2005). Solving and Testing for Regressor-Error (in) Dependence When no Instrumental Variables are Available : With New Evidence for the Effect of Education on Income. *Quantitative Marketing and Economics*, *3*, 365–392.

Farré, L., Klein, R., & Vella, F. (2013). A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: An application to the NLSY. *Empirical Economics*, *44*(1), 111–133. https://doi.org/10.1007/s00181-010-0376-5

Girma, S., & Kedir, A. (2005). Heterogeneity in returns to schooling : Econometric evidence from Ethiopia. *The Journal of Development Studies*, *41*(8), 1405–1416. https://doi.org/10.1080/00220380500187026

Heckman, J. J. (1979). Sample selection bias as a specification error. *Applied Econometrics*, *31*(3), 129–137. https://doi.org/10.2307/1912352

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In E. A. Hanushek, S. J. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 1, Issue 06, pp. 307–458). Elsevier B.V. https://doi.org/10.1016/S1574-0692(06)01007-5

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2008). Earnings Functions and Rates of Return. *Journal of Human Capital*, *2*(1), 1–31.

Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, *73*(3), 669–738. https://doi.org/10.1111/j.1468-0262.2005.00594.x

Heckman, J. J., & Vytlacil, E. J. (2007). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. In *Handbook of Econometrics* (Vol. 6B, pp. 4875–5143). https://doi.org/10.1016/S1573-4412(07)06071-0

Hult, G. T. ., Proksch, D., Sarstedt, M., Pinkwart, A., & Ringle, C. . (2018). Addressing Endogeneity in International Marketing Applications of Partial Least Squares Structural Equation Modeling. *Journal of International Marketing*, *26*(3), 1–21. https://doi.org/https://doi.org/10.1509/jim.17.0151

Klein, R., & Vella, F. (2009). Estimating the Return to Endogenous Schooling Decisions via Conditional Second Moments. *The Journal of Human Resources*, *44*(4), 1047–1065.

Klein, R., & Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, *154*(2), 154–164. https://doi.org/10.1016/j.jeconom.2009.05.005

Kuepié, M., & Nordman, C. J. (2016). Where Does Education Pay Off in Sub-Saharan Africa? Evidence from Two Cities of the Republic of Congo. *Oxford Development Studies*, *44*(1), 1–27. https://doi.org/10.1080/13600818.2015.1110568

Lassibille, G., & Tan, J. P. (2005). The Returns to Education in Rwanda. *Journal of African Economies*, *14*(1), 92–116. https://doi.org/10.1093/jae/ejh035

Lewbel, A. (2012). Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. *Journal of Business & Economic Statistics*, *30*(1), 67–80. https://doi.org/10.1080/07350015.2012.643126

Mincer, J. (1974). *Schooling, experience, and earnings. Human behavior and social institutions*. Columbia University Press, New York.

Mishra, V., & Smyth, R. (2015). Estimating returns to schooling in urban China using conventional and heteroskedasticity-based instruments. *Economic Modelling*, *47*, 166–173. https://doi.org/10.1016/j.econmod.2015.02.002

Nikolov, P., & Jimi, N. (2018). What factors drive individual misperceptions of the returns to schooling in Tanzania ? Some lessons for education policy. *Applied Economics ISSN:*, *50*(44), 4705–4723. https://doi.org/10.1080/00036846.2018.1466991

Park, S., & Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science*, *31*(4), 567–586. https://doi.org/10.1287/mksc.1120.0718

Patrinos, H. A., & Psacharopoulos, G. (2010). Returns to education in developing countries. *International Encyclopedia of Education*, 305–312. https://doi.org/10.1016/B978-0-08-044894-7.01216-1

Peet, E. D., Fink, G., & Fawzi, W. (2015). Returns to education in developing countries: Evidence from the living standards and measurement study surveys. *Economics of Education Review*, *49*, 69–90. https://doi.org/10.1016/j.econedurev.2015.08.002

Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics*, *12*(2), 111–134. https://doi.org/10.1080/0964529042000239140

Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. *Education Economics*, *26*(5), 445–458. https://doi.org/10.1080/09645292.2018.1484426

Rutz, O. J., & Watson, G. F. (2019). Endogeneity and marketing strategy research: an overview. *Journal of the Academy of Marketing Science*, *47*(3), 479–498. https://doi.org/10.1007/s11747-019-00630-4

Salisbury, T. (2016). Education and inequality in South Africa: Returns to schooling in the post-apartheid era. *International Journal of Educational Development*, *46*, 43–52. https://doi.org/10.1016/j.ijedudev.2015.07.004

Schultz, T. P. (1961). Investment in Human Capital. *The American Economic Review*, *51*(1), 1–17.

Schultz, T. P. (2004). Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education. *Journal of African Economies*, *13*(suppl_2), ii95–ii148.

https://doi.org/10.1093/jae/ejh044

Serneels, P., Beegle, K., & Dillon, A. (2017). Do returns to education depend on how and whom you ask? *Economics of Education Review*, *60*, 5–19. https://doi.org/10.1016/j.econedurev.2017.07.010

Söderbom, M., Teal, F., Eberhardt, M., Quinn, S., & Zeitlin, A. (2014). *Emperical Development Economics* (1st ed.). Routledge, London and New York.

Stefani, P. C., & Biderman, C. (2009). The evolution of the returns to education and wage differentials in Brazil : a quantile approach. *Applied Economics*, *41*(11), 1453–1460. https://doi.org/10.1080/00036840601019299

Verbeek, M. (2004). *A Guide to Modern Econometrics* (2nd ed.). John Wiley & Sons Ltd.

Wang, L. (2013). Estimating returns to education when the IV sample is selective. *Labour Economics, 21*, 74–85. https://doi.org/10.1016/j.labeco.2012.12.007

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT, Cambridge, Massachusetts, London.

APPENDICES

APPENDIX A

Table A.1: Selected Studies on Returns to Schooling in Developing Countries

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Nikolov and Jimi (2018)	Log monthly earnings	No	Both linear and non-linear schooling using, three dummies for education levels: primary, O-level and A-level secondary (reference category 'no schooling')	ILFS 2014 and Dar es Salaam Perceived Returns Survey (DPRS) 2014. ILFS 16,817 observations and DPRS 1,211	OLS for non-linear and IV(2SLS) for linear schooling	Convex returns, national returns are 12% and 7% while Dar es Salaam returns are 11% and 9% respectively for OLS and IV. Estimates are insignificant when sample is split into the three levels of education.
Bridges et al. (2017)	Log monthly earnings	No	Non-linear schooling: dummies for completed levels of education: primary, O-level secondary, A-level secondary, vocational/technical.	All three rounds of Tanzania Household Urban Panel Survey (THUPS), a subsample of youth aged 20 to 35 inclusive. A sample of 365 individuals	Fixed Effects	All dummies for education levels are insignificant after controlling for family fixed effects.

Table A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Serneels et al. (2017)	Log daily wages	Yes	Non-linear schooling: primary vs post primary school levels	Survey of Household Welfare and Labour in Tanzania (SHWALITA), 291 sample of wage workers	IV (Control function)	Returns differ by survey instrument but not by type of respondent. Short module questionnaires lead to biased estimates compared to detailed questionnaires. After controlling for endogeneity and selection using Heckman method, returns are about 20% and 49% for a year of post primary school respectively for men and women if short modules are used. Using Heckman-Hotz method, the returns are respectively 21% and 32%. While generally schooling is insignificant for men when detailed modules are used, post primary returns are 50% and 29% for women using Heckman and Heckman-Hotz method, respectively.

Table A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Peet et al. (2015)	Log annual earnings	No	Linear schooling	LSMS 2004, 2008, 2010. Sample 985; 1,807 and 2,716 respectively	OLS	The returns are convex. Returns are 12.1% 9% and 12.2% for the survey years respectively with the period average of 11.1%. Returns are higher for female and urban employees
Barouni and Broecke (2014)		No	Non-linear schooling: dummies for different completed levels of education		OLS	Returns are 5%, 100% and 51% for basic, A level and tertiary education respectively
Kahyarara and Teal (2008)	Log monthly earnings	No	Non-linear schooling: dummies for completed levels of education: primary, O-level secondary, A-level secondary, vocational, technical, professional and university	Fourth and fifth rounds of the Tanzanian Manufacturing Enterprise surveys. Total sample of 2527 employees	IV (control function) with firm fixed effects: parental education and main occupation as instruments	Returns are convex: higher levels of education(academic) have higher returns. Returns to vocation and technical education depend on the level of education(academic) with which one enters vocational/technical college. The higher the entry level the lower the returns.

Table A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Al-Samarrai and Reilly (2008)	Log monthly earnings	No	Non-linear post primary school levels	Tracer survey 2001. A sample of 965 respondents		The returns are convex. The rate of returns for a year of A level of education for the wage employees is 8.8% while the rates for a year of university education is 17.1%. No significant effect of these levels of education on the Self-employed.
Soderbom et al. (2006)	Log monthly earnings	No	Linear schooling	Surveys of employees in the manufacturing sector 1993, 1994, 1999 and 2001. Total sample of 2,738 workers	IV (control function): parental education, main occupation, distance to primary school at age 6 and to secondary at age 12 as instruments	The returns are convex. There has been an increase in returns from early 1990s to 2000. The earning profiles for young and old people are significantly different. After controlling for endogeneity, youth returns are 10.6% and is insignificant for the old

Table A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Agrawal and Agrawal (2018)	Log hourly earnings	No	Both linear and non-linear schooling	India Human Development Survey (IHDS), 2011–2012	OLS corrected for employment selection bias	Returns are higher for females (5.7% compared to 5% for males). Wage employees have highest returns followed by self-employed and agriculture. Returns are convex ranging from 2.2% for primary education to 18.9% for university education.
Chuang and Lai (2017)	Log hourly earnings	No	Linear schooling	Taiwan's 1978-2003 Manpower Utilization Survey	Quantile regression	Returns increased from 5.5% in 1978 to 8.2 % in 2003 with an average of 6.5% The returns are higher for lower quantiles and vice versa.
Salisbury (2016)	Log monthly earnings	No	Both linear and non-linear schooling	National Income Dynamics Study 2008 (south Africa)	OLS	Returns are 18.7%, lowest for Africans (16%) and highest for Asians/Indians (25%). The returns are also higher for females. When allowing for non-linearities in schooling, returns are

						convex: 7%, 13% and 29% respectively for primary, secondary and tertiary education.
Kuepié and Nordman (2016)	Log hourly earnings	No	Non-linear schooling: dummies for different completed levels of education	Employment and Informal Sector Survey (EESIC) 2009 (Republic of Congo)	IV (control function): father's education and job professional status as instruments	Convex returns. Primary education no effect on earning, returns for lower secondary, upper secondary and higher education are respectively 9%, 5% and 12% for Brazzaville and 9%, 14% and 13% for Pointe-Noire.
Peet et al. (2015)	Log annual earnings	No	Linear schooling	LSMS data from 25 developing countries of which 9 countries from Africa: Cote d'Ivoire, Ethiopia, Ghana, Malawi, Niger, Nigeria, SA, Tanzania and Uganda	OLS	Returns differ significantly by countries and within countries by survey years. But generally, they range from 3.2% to 12.5%. The pattern of returns across the levels of education also differs by countries and by survey years. Returns are generally higher for women though the difference is small

Table A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Barouni and Broecke (2014)	Not specified	No	Non-linear schooling: dummies for completed levels of education	Post 2005 household and labour force surveys in Burundi, Egypt, Ghana, Mali, Nigeria, Rwanda, Sudan, South Africa, Tanzania, Togo, Tunisia, and Uganda.	OLS	The average Mincer returns for the 12 countries are 7%, 26% and 26% for basic, upper secondary and tertiary education, respectively. Returns are higher for women except for tertiary education where they are equal. The pattern of returns across the levels of education differs by countries.
Wang (2013)	Log annual wages	No	Linear schooling	urban sample of the China Household Income Project (CHIP) 1995 and 2002	IV (2SLS): parental education vs spouse education	The returns increased over the two survey periods regardless of the instrument used. Returns are higher using parental education as IV relative to spouse education, but the difference is not statistically significant.

Table A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Aslam et al. (2012)	Log daily/regular wages	No	Linear and quadratic schooling	Purposive household survey in Punjab and the North West Frontier Province (NWFP) Pakistan 2006 -2007	OLS with ability proxy, IV (2SLS) and Fixed effects	Males have returns of 10% using IV method, schooling not significant for females. Using Fixed effects, returns are 5% similar to OLS
Stefani and Biderman (2009)	Log hourly wage	No	Linear schooling	Brazil National Household Survey 1988 and 1996	IV Quantile regression: parental education and family size as instruments	Returns are heterogeneous across colour gender and earning distribution, ranging from 6% to 32%
Pietro (2008)	Log hourly wage	No	Linear schooling	The Argentine Permanent Household Survey 1995 - 2003	OLS with selection correction and IV (2SLS): spouse education	Decrease in returns between 1996 and 1999 and increase in returns 1999 to 2002. Returns from OLS corrected for selection average at 8.5% while IV estimates are averaged at 11.5%
Soderbom et al. (2005)	Log monthly earnings	No	Linear schooling	Surveys of manufacturing firms in Ghana and Kenya.	OLS	Returns are 8.3% in Ghana and 10.4% in Kenya

Table A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Lassibille and Tan (2005)	Log hourly wage	No	Both linear and non-linear schooling	Household Living Conditions Survey 1999-2001(Rwanda)	OLS corrected for employment selection bias	Returns are 17.5% and convex: primary - 19%, secondary 29% and tertiary 33%. Generally public sector has higher returns compared to private sector.
Girma and Kedir (2005)	Log hourly wage	No	Linear schooling	Household panel data for Ethiopian seven major cities 1994, 1995, and 1997	IV Quantile regression: parental education as instrument	Returns are 14%. The returns differ across the earning distribution: highest at 25th (20%) quantile and lowest at 90th quantile (%). Lower returns for public sector (12%) relative to private sector (16%).
Schultz (2004)	Log hourly wage	No	Non-linear schooling: dummies for completed levels of education	Various national representative household surveys from 6 African Countries: Burkina Faso, Cote d'Ivoire, Ghana, Kenya, Nigeria and SA in the period 1985 - 1999	OLS	Returns differ significantly by countries and by levels of education. Generally, an extra year is associated with 5 to 20% increase in earnings. Primary school returns range between 3 - 10% while

						tertiary education returns range between 10 - 15%.
Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Moock et al. (2003)	Log monthly earnings	No	Both linear and non-linear schooling	Vietnam Living Standards Survey (VLSS) 1992-1993	OLS	Using linear schooling returns are 8% while using education dummies returns are highest at primary school(13%), followed by university (11%). Returns for secondary and vocational education are respectively 5% and 4%.

APPENDIX B: VARIABLE CONSTRUCTION

The focus is on the survey data cleaning, extraction of key variables from the surveys, construction of the variables for analysis, and how we arrived at the final samples. This is for both the primary samples and the off-own-farm casual labour market (*ganyu*) in Malawi.

Construction of the earnings Measures

The surveys collected information about individuals wage earnings and the frequency of payment[5]. The payment periods for Malawi were daily, weekly, and monthly; for Tanzania hourly, daily, weekly, fortnightly, monthly, quarterly, semi-annually, and annually; and for Uganda hourly, daily, weekly, and monthly. Note that the payment periods may not necessarily imply the same period/duration of employment, i.e., being paid daily or weekly does not always mean that employment last only for a day or a week. Each of the conversion methods is discussed below.

1. Aggregating to Daily Earnings (DailyC)

Wages were converted to daily wages as follows:
(a) Hourly to daily (Tanzania and Uganda)
The hourly wage multiplied by the number of hours assuming nine (9) working hours a day.
(b) Weekly to daily
Weekly wage divided by the total number of days worked per week (unless otherwise stated in the survey, days were inferred from the total weekly hours).
(c) Fortnightly to daily (Tanzania)
Fortnightly wage divided by two and then divided by the total number of days worked per week.
(d) Monthly to daily
Monthly wage divided by 22 (assuming those earning monthly wage worked 22 days in any month).
(e) Quarterly
Quarterly wage divided by 66 (since assumption 22 working days in any month).
(f) Semi-annual
Semi-annual wage divided by 132 working days.
(g) Annual
Annual wage divided by 264 working days.

2. Aggregating to Monthly Earnings (MonthlyC)

Monthly wages were constructed from the reported wages as follows:
(a) Hourly to monthly (Tanzania and Uganda)
Total number of hours worked over the last seven (7) days multiplied by hourly wage and number of weeks worked in the job in a typical month.[6]
(b) Daily to monthly

5 The surveys asked individuals' how much their last (wage) payments were or how much they were expecting to get for those who were not yet paid and the period the payments covered.

6 As stated earlier, TNPS (except for the first wave) asked how many weeks per month did the individual usually work in the job during the last 12 months. For the first wave we replaced it by the median values of the sample for each pay period.

For Uganda, the number of days the individual worked over the last seven days was available from the data. For Malawi and Tanzania, since the number of days was not available, we used the total number of hours in a week to infer days. Assuming nine (9) working hours per days, we obtained the proxy for days by dividing the total hours by nine. We then constructed the monthly wage as a product of the daily wage, days worked, and the number of weeks worked per month.

(c) Weekly to monthly

Weekly wage multiplied by the number of weeks worked per month.

(d) Fortnightly to monthly (Tanzania)

Fortnightly wage multiplied by two (2).

(e) Quarterly to monthly (Tanzania)

Quarterly wage divided by three (3).

(f) Semi-annually to monthly (Tanzania)

Semi-annual wage divided by six (6).

(g) Annually to monthly

Annual wage divided by 12.

3. Aggregating to Annualised Earnings (MonthlyA)

The reported wages were annualised as follows:

(a) Hourly to annual (Tanzania and Uganda)

The product of hourly wage and hours per week, weeks per month and months worked over the last 12 months.

(b) Daily to annual

The product of daily wage and days per week, weeks per month, and months worked over the last 12 months.

(c) Weekly to annual

The product of weekly wage, weeks worked per month and months worked over the last 12 months.

(d) Fortnightly to annual (Tanzania)

Fortnight wage divided by two then multiplied by weeks worked per month and months worked over the last 12 months.

(e) Monthly to annual

Monthly wage multiplied by the number of months worked over the last 12 months.

(f) Quarterly to annual

Quarterly wage divided by three, then multiplied by the number of months worked over the last 12 months.

(g) Semi-annually to annual

Semi-annual wage divided by six, multiplied by the number of months worked over the last 12 months.

Annualised wages are expressed monthly by dividing by 12, which gives the average monthly earnings over the last 12 months. This may not be identical to the constructed measure MonthlyC except for those paid monthly who worked 12 months in the previous year.

After the construction of our wage/earnings variables, we observed a small number of cases with very low or very high MonthlyA, likely due to errors in recording the wage or variables used to construct the aggregated wages. We trimmed the bottom and top one percent of MonthlyA to remove outliers.

For *ganyu*, DailyC is the self-reported average daily wage a worker receives. MonthlyC is obtained by taking into consideration the 'average' number of days in a week and number of weeks in a month the worker participated in *ganyu* (i.e., daily rate x days per week x weeks per month). MonthlyA also allows for the number of months worked in *ganyu* over the last 12 months (i.e. MonthlyC x number of months then divided by 12 to get average per month).

Construction of Explanatory Variables

Years and Levels of Education

In all three countries, each grade requires a year to complete. The IHS, TNPS and UNPS used a closed-ended question to capture the highest grade completed by each member of the household. Therefore, we utilised the information on the grades completed to calculate the respondent's years of schooling assuming that each additional grade corresponds to an additional year of schooling. Note, however, that there was no information on the number of years the individual took to complete their highest grade. Hence, the calculation of years of schooling assumed no repetitions or skipping of grades.

Primary education is compulsory in all three countries, and it runs for eight years in Malawi and seven years in Tanzania[7] and Uganda. In Malawi, secondary education lasts for four years and until 2015 consisted of two sets of two years. The first two years lead to the Junior Certificate of Education (JCE) (which was abolished in 2015) and the second two years to the Malawi Certificate of Secondary Education (MCSE). Admission to (non-university) technical college education such as diplomas in vocational training including nursing, primary teacher training and agriculture requires a minimum of JCE and run for two, three or four years. Admission to university requires the MCSE, with a minimum of three years required to earn a university degree.

In Tanzania and Uganda, secondary education consists of six years in two levels: ordinary level (O-level) and advanced level (A-level) which run for four and two years, respectively. Diploma education is two years for those enrolled after A-level and three or four years for those enrolled after O-level (in our analysis, we use three years for those enrolled after O-level). University education is three to five years, depending on the programme of study. Note that individuals can enrol in technical/vocational education after completing primary or secondary education. This form of education can take less than a year to more than two years. For simplicity, in our calculation, we assume this level does not constitute an additional year of schooling.[8]

Since the surveys reported the highest grade of schooling completed (for each level of education) assigning individuals into dummy variables for the highest completed levels was straight forward. Accordingly, we constructed the following dummy variables:

- **noeduc:** educational dummy, 1 if incomplete primary school education and 0 otherwise.

7 Before 1969 primary education in Tanzania ran for eight years. An education reform act in late 1960s eliminated the 8th grade thereby reducing the primary school years from eight to seven. We used individuals' years of birth to infer whether the individual obtained seven or eight years of primary schooling. We assumed all individuals who completed eight years of primary education started school at age seven and were born before 1956. Any error or misreporting of the birth year would then affect years of schooling, especially those with post-secondary education.

8 For Tanzania and Uganda, a total of 724 and 404 wage employees had vocational education of unspecified duration, accounting for 6.5% and 8.7% of the samples, respectively.

- **primary**: educational dummy, 1 if completed primary school education and 0 otherwise.
- **secondary**: educational dummy, 1 if completed ordinary/advanced secondary school education and 0 otherwise.
- **higher**: educational dummy, 1 if completed diploma/university education and 0 otherwise.

Other Explanatory Variables and Exclusion Restrictions

- **age:** After correcting the inconsistencies in the year of birth[9], we calculated age as the difference between the survey year, and the year birth (taking into consideration the month of birth).
- **female:** The variable female is a gender dummy = 1 for females and 0 otherwise.
- **rural:** The variable rural is a location dummy = 1 for rural areas and 0 for urban. It was readily available in the data sets.
- **panel:** Tanzania and Uganda only, panel is a dummy variable = 1 for the individuals observed multiple times and 0 otherwise.
- **year**: Malawi only, year is a dummy variable = 1 if the year of the survey is 2016 and 0 if 2010.
- **married:** The survey question for marital status consisted of seven responses: monogamous married, polygamous married, living together, separated, divorced, never married, and widow(er). We made a dummy variable = 1 if married or living together and 0 otherwise.
- **kids5:** It is the proportion of children aged five and under in the household calculated as the ratio of the number of children aged five years and younger to the total number of households.
- **kids14**: It is the proportion of children aged 6 to 14 years of age in the household calculated as the ratio of children aged between 6 and 14 years inclusively to the total number of households.

9 For TNPS the year of birth was not available in the fourth wave, instead, we calculated it from the age of the respondents.

APPENDIX C: GC ESTIMATES OF RETURNS TO SCHOOLING BY PERIOD

Table C.1: GC Estimates of Returns to Schooling by Period - Malawi

Period	(1) Day	(2) Week	(3) Month	(4) Pooled
sch	-0.015	-0.140**	-0.091***	-0.101***
	(0.086)	(0.057)	(0.011)	(0.011)
sch2	0.005	0.012***	0.010***	0.009***
	(0.007)	(0.002)	(0.001)	(0.001)
age	0.039	0.078***	0.054***	0.055***
	(0.058)	(0.026)	(0.006)	(0.006)
age2	-0.034	-0.081**	-0.049***	-0.051***
	(0.074)	(0.032)	(0.007)	(0.008)
female	-0.031	-0.118*	-0.119***	-0.106***
	(0.188)	(0.071)	(0.022)	(0.021)
rural	-0.243	-0.305***	-0.188***	-0.211***
	(0.222)	(0.081)	(0.019)	(0.019)
year	0.851***	0.990***	1.277***	1.234***
	(0.214)	(0.078)	(0.017)	(0.019)
weeks	1.366***	1.231***	1.146***	0.168***
	(0.207)	(0.076)	(0.029)	(0.050)
copula(sch)	0.644	0.189	0.129***	0.168***
	(0.400)	(0.261)	(0.046)	(0.049)
copula(sch2)	-0.533	0.074	0.120**	1.151***
	(0.511)	(0.170)	(0.048)	(0.029)
copula(weeks)	-0.038	-0.063***	-0.001	-0.006
	(0.043)	(0.018)	(0.004)	(0.005)
constant	-2.052	-1.360*	-1.401***	-1.199***
	(1.427)	(0.707)	(0.171)	(0.166)
AME(sch)	0.075	0.050	0.096***	0.066***
	(0.113)	(0.643)	(0.015)	(0.014)
Obs.	182	505	5,129	5,816
R^2	0.45	0.67	0.77	0.74

Notes: Copula() are Gaussian Copula functions; significance implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.2: GC Estimates of Returns to Schooling - Tanzania

	Day	Week	Month	Pooled
sch	-0.022	-0.028	-0.006	-0.033***
	(0.017)	(0.019)	(0.016)	(0.010)
sch2	0.005***	0.009***	0.006***	0.009***
	(0.002)	(0.002)	(0.000)	(0.000)
age	0.053***	0.046***	0.098***	0.066***
	(0.008)	(0.009)	(0.006)	(0.004)
age2	-0.065***	-0.053***	-0.093***	-0.070***
	(0.011)	(0.012)	(0.008)	(0.006)
female	-0.738***	-0.553***	-0.329***	-0.549***
	(0.036)	(0.036)	(0.024)	(0.019)
rural	-0.521***	-0.191***	-0.256***	-0.330***
	(0.041)	(0.050)	(0.023)	(0.018)
panel	-0.216***	-0.129***	-0.069***	-0.087***
	(0.039)	(0.049)	(0.022)	(0.018)
weeks	1.075***	0.998***	1.044***	0.039
	(0.023)	(0.024)	(0.026)	(0.028)
copula(sch)	-0.018	-0.027	0.058	0.038
	(0.046)	(0.045)	(0.042)	(0.028)
copula(sch2)	0.037	0.026	0.112***	1.078***
	(0.045)	(0.046)	(0.042)	(0.011)
copula(weeks)	0.074***	0.096***	0.009*	0.029***
	(0.013)	(0.018)	(0.006)	(0.005)
constant	-0.542***	-0.652***	-1.886***	-1.105***
	(0.174)	(0.170)	(0.184)	(0.100)
AME(sch)	0.038*	0.071***	0.103***	0.083***
	(0.019)	(0.020)	(0.014)	(0.009)
Obs	3,738	1,929	4,830	11,215
R^2	0.73	0.80	0.72	0.78

Notes: Copula() are Gaussian Copula functions; significance implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.3: GC Estimates of Returns to Schooling - Uganda

	(1) Day	(2) Week	(3) Month	(4) Pooled
sch	0.105*	0.091	0.040*	0.030
	(0.059)	(0.078)	(0.022)	(0.020)
sch2	-0.002	0.003	0.003***	0.003***
	(0.001)	(0.002)	(0.001)	(0.001)
age	0.063***	0.079***	0.080***	0.068***
	(0.014)	(0.020)	(0.010)	(0.007)
age2/100	-0.082***	-0.092***	-0.083***	-0.077***
	(0.019)	(0.026)	(0.013)	(0.009)
female	-0.592***	-0.544***	-0.313***	-0.435***
	(0.063)	(0.084)	(0.033)	(0.028)
rural	-0.233***	-0.369***	-0.231***	-0.239***
	(0.056)	(0.083)	(0.036)	(0.029)
panel	0.009	-0.062	0.177***	0.155***
	(0.055)	(0.082)	(0.038)	(0.030)
weeks	1.167***	1.158***	1.164***	0.144*
	(0.069)	(0.090)	(0.048)	(0.074)
Copula(sch)	0.137	-0.056	0.055	0.042
	(0.161)	(0.269)	(0.065)	(0.077)
Copula(sch2)	-0.085	-0.076	0.108	1.172***
	(0.169)	(0.200)	(0.081)	(0.038)
Copula(weeks)	0.023*	-0.020	-0.008	-0.005
	(0.014)	(0.020)	(0.007)	(0.006)
constant	-1.830***	-2.204***	-2.426***	-1.933***
	(0.503)	(0.686)	(0.258)	(0.214)
AME(sch)	0.082	0.131*	0.108***	0.081***
	(0.057)	(0.074)	(0.021)	(0.018)
Obs.	1,262	589	2,765	4,631
R^2	0.57	0.57	0.63	0.600

Notes: Copula() are Gaussian Copula functions; significance implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C4: GC Estimates of Returns to Schooling - *Ganyu* Labour

	GC		
	DailyC	MonthlyC	MonthlyA
sch	0.008	-0.007	-0.002
	(0.009)	(0.011)	(0.010)
sch2	0.002***	0.003***	0.002***
	(0.000)	(0.001)	(0.000)
age	0.034***	0.047***	0.040***
	(0.002)	(0.003)	(0.003)
age2/100	-0.040***	-0.057***	-0.048***
	(0.003)	(0.004)	(0.004)
female	-0.240***	-0.444***	-0.341***
	(0.010)	(0.013)	(0.012)
rural	-0.284***	-0.420***	-0.326***
	(0.020)	(0.028)	(0.023)
year	1.326***	1.223***	1.217***
	(0.010)	(0.014)	(0.012)
weeks			0.982***
			(0.011)
Copula(sch)	-0.012	0.004	0.010
	(0.020)	(0.027)	(0.023)
Copula(sch2)	-0.003	-0.043	-0.004
	(0.020)	(0.027)	(0.023)
Copula(weeks)			0.020**
			(0.008)
Constant	0.172***	2.591***	-0.855***
	(0.061)	(0.082)	(0.078)
AME(sch)	0.024***	0.020*	0.016*
	(0.007)	(0.010)	(0.009)
Obs.	16,528	16,528	16,528
R^2	0.56	0.38	0.77

Notes: Copula() are Gaussian Copula functions; significance implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

APPENDIX D: DETERMINANTS OF SELECTION TO EMPLOYMENT

Table D.1: Determinants of Selection to Employment - Malawi

	(1) Day	(2) Week	(3) Month	(4) Pooled	(5) *Ganyu*
sch	-0.025	0.031	-0.096***	-0.105***	0.020*
	(0.043)	(0.030)	(0.016)	(0.015)	(0.011)
sch2	0.004***	0.000	0.009***	0.010***	-0.007***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)
age	0.042**	0.041***	0.070***	0.072***	0.016***
	(0.017)	(0.011)	(0.006)	(0.005)	(0.004)
age2/100	-0.050**	-0.055***	-0.083***	-0.087***	-0.037***
	(0.021)	(0.014)	(0.007)	(0.007)	(0.005)
female	-0.142*	-0.169***	-0.394***	-0.388***	-0.149***
	(0.073)	(0.046)	(0.024)	(0.023)	(0.018)
rural	0.027	-0.191***	-0.677***	-0.659***	0.627***
	(0.075)	(0.044)	(0.021)	(0.020)	(0.019)
year	0.347***	0.101***	-0.216***	-0.152***	0.444***
	(0.059)	(0.035)	(0.019)	(0.018)	(0.012)
kids5	0.113	-0.203	-0.552***	-0.527***	0.423***
	(0.209)	(0.136)	(0.062)	(0.061)	(0.045)
kids14	-0.024	-0.016	-0.408***	-0.367***	0.128***
	(0.177)	(0.100)	(0.049)	(0.046)	(0.037)
married	-0.075	-0.069	-0.040*	-0.049**	-0.190***
	(0.074)	(0.047)	(0.023)	(0.022)	(0.018)
head	0.343***	0.393***	0.573***	0.613***	0.323***
	(0.076)	(0.052)	(0.027)	(0.025)	(0.018)
Copula(sch)	0.183	-0.110	0.137***	0.138***	-0.045
	(0.153)	(0.086)	(0.048)	(0.046)	(0.030)
Copula(sch2)	-0.105	0.098	0.144***	0.158***	-0.013
	(0.132)	(0.093)	(0.048)	(0.046)	(0.030)
Constant	-3.964***	-3.171***	-1.942***	-1.922***	-0.982***
	(0.421)	(0.278)	(0.140)	(0.136)	(0.084)
N	45,494	45,494	45,494	45,494	45,494

Notes: Copula() are Gaussian Copula functions; significance implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D.2: Determinants of Selection to Employment - Tanzania

	Daily	Weekly	Monthly	Pooled
sch	0.036***	0.045***	-0.024**	-0.037***
	(0.009)	(0.011)	(0.010)	(0.008)
sch2	-0.009***	-0.007***	0.007***	0.005***
	(0.001)	(0.001)	(0.001)	(0.000)
age	0.039***	0.039***	0.025***	0.050***
	(0.005)	(0.006)	(0.005)	(0.004)
age2/100	-0.062***	-0.057***	-0.035***	-0.075***
	(0.006)	(0.007)	(0.006)	(0.005)
female	-0.349***	-0.306***	-0.179***	-0.388***
	(0.021)	(0.027)	(0.021)	(0.017)
rural	0.046**	0.150***	-0.410***	-0.171***
	(0.021)	(0.028)	(0.018)	(0.015)
kids5	-0.191***	-0.035	-0.022	-0.064
	(0.020)	(0.025)	(0.020)	(0.049)
kids14	0.182***	0.203***	-0.419***	-0.173***
	(0.066)	(0.078)	(0.065)	(0.043)
panel	0.055	0.079	-0.419***	-0.116***
	(0.057)	(0.069)	(0.053)	(0.015)
married	-0.130***	-0.103***	-0.173***	-0.168***
	(0.022)	(0.027)	(0.021)	(0.017)
head	0.260***	0.217***	0.482***	0.510***
	(0.025)	(0.032)	(0.026)	(0.019)
Copula(sch)	0.030	-0.022	0.035	-0.001
	(0.026)	(0.033)	(0.029)	(0.020)
Copula(sch2)	0.013	-0.054	0.098***	-0.010
	(0.027)	(0.033)	(0.029)	(0.021)
Constant	-1.428***	-2.217***	-1.412***	-0.966***
	(0.097)	(0.121)	(0.100)	(0.074)
Obs.	38,857	38,857	38,857	38,857

Notes: Copula() are Gaussian Copula functions; significance implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: Determinants of Selection to Employment - Uganda

	Daily	Weekly	Monthly	Pooled
sch	0.037	0.024	0.037	-0.010
	(0.037)	(0.045)	(0.024)	(0.023)
sch2	-0.004***	-0.001	0.007***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
age	0.031***	0.021**	0.067***	0.058***
	(0.008)	(0.010)	(0.006)	(0.005)
age2/100	-0.058***	-0.038***	-0.085***	-0.083***
	(0.010)	(0.013)	(0.008)	(0.007)
female	-0.667***	-0.390***	-0.306***	-0.554***
	(0.032)	(0.042)	(0.027)	(0.022)
rural	-0.470***	-0.160***	-0.330***	-0.448***
	(0.033)	(0.039)	(0.026)	(0.023)
kids5	0.024	-0.024	-0.276***	-0.492***
	(0.043)	(0.055)	(0.033)	(0.138)
kids14	-0.236	-0.478*	-0.445**	-0.385***
	(0.201)	(0.279)	(0.185)	(0.046)
panel	-0.413***	-0.208**	-0.254***	-0.160***
	(0.074)	(0.088)	(0.052)	(0.030)
married	-0.299***	-0.192***	-0.208***	-0.297***
	(0.033)	(0.041)	(0.027)	(0.023)
head	0.182***	0.220***	0.090***	0.176***
	(0.039)	(0.048)	(0.030)	(0.026)
Copula(sch)	-0.118	-0.057	-0.202***	-0.152***
	(0.084)	(0.101)	(0.059)	(0.052)
Copula(sch2)	0.001	0.018	0.020	0.049
	(0.102)	(0.102)	(0.067)	(0.062)
Constant	-1.280***	-1.945***	-2.457***	-1.280***
	(0.224)	(0.297)	(0.173)	(0.153)
Obs.	29,188	29,188	29,188	29,188

Notes: Copula() are Gaussian Copula functions; significance implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

APPENDIX E: BOOTSTRAP AGGREGATING (BAGGING) RESULTS

Table E.1: Bootstrap Aggregation Results (Converted Earnings) -Malawi

	GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.110***	-0.105***	-0.095***	-0.111***	-0.106***	-0.096***
	(0.005)	(0.010)	(0.010)	(0.005)	(0.005)	(0.005)
sch2	0.009***	0.009***	0.009***	0.008***	0.008***	0.009***
	(0.0004)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)
age	0.064***	0.060***	0.056***	0.053***	0.048***	0.047***
	(0.0003)	(0.006)	(0.006)	(0.001)	(0.001)	(0.001)
age2/100	-0.064***	-0.057***	-0.052***	-0.052***	-0.044***	-0.042***
	(0.0004)	(0.007)	(0.007)	(0.001)	(0.001)	(0.001)
female	-0.135***	-0.136***	-0.106***	-0.055***	-0.049***	-0.040***
	(0.001)	(0.022)	(0.021)	(0.004)	(0.004)	(0.003)
rural	-0.137***	-0.210***	-0.210***	-0.057***	-0.122***	-0.144***
	(0.001)	(0.020)	(0.019)	(0.004)	(0.003)	(0.003)
year	1.333***	1.239***	1.235***	1.352***	1.259***	1.250***
	(0.001)	(0.019)	(0.018)	(0.001)	(0.001)	(0.001)
weeks			1.149***			1.144***
			(0.026)			(0.001)
Copula(sch)	0.158***	0.149***	0.137***	0.164***	0.156***	0.142***
	(0.046)	(0.047)	(0.046)	(0.045)	(0.043)	(0.042)
Copula(sch2)	0.158***	0.149***	0.138***	0.164***	0.156***	0.142***
	(0.045)	(0.046)	(0.046)	(0.047)	(0.043)	(0.043)
Copula(weeks)			-0.006			-0.005***
			(0.004)			(0.0002)
IMR				-0.173***	-0.191***	-0.144***
				(0.008)	(0.007)	(0.007)
Constant	0.134	3.141***	-1.320***	0.607***	3.663***	-0.914***
	(0.082)	(0.152)	(0.166)	(0.096)	(0.089)	(0.086)
Obs.	5,816	5,816	5,816	5,816	5,816	5,816
R^2	0.59	0.62	0.74			

Notes: The Copula() functions for schooling are positive and significant implying positive and significant correlation between schooling variables and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.2: Bootstrap Aggregation Results (Converted Earnings) -Tanzania

	GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.046***	-0.038***	-0.043***	-0.040***	-0.017**	-0.032***
	(0.006)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)
sch2	0.008***	0.010***	0.008***	0.008***	0.007***	0.007***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
age	0.075***	0.065***	0.066***	0.064***	0.027***	0.046***
	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0004)	(0.0003)
age2/100	-0.079***	-0.067***	-0.070***	-0.065***	-0.017***	-0.043***
	(0.0001)	(0.0003)	(0.0002)	(0.0002)	(0.001)	(0.0003)
female	-0.445***	-0.642***	-0.549***	-0.349***	-0.303***	-0.370***
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)
rural	-0.170***	-0.662***	-0.330***	-0.135***	-0.539***	-0.271***
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
panel	-0.104***	-0.002**	-0.087***	-0.095***	0.030***	-0.068***
	(0.0004)	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)
weeks			1.078***			1.069***
			(0.0004)			(0.0004)
Copula(sch)	0.039	0.097***	0.061**	0.039	0.093***	0.060**
	(0.025)	(0.032)	(0.027)	(0.025)	(0.032)	(0.027)
Copula(sch2)	0.039	0.097***	0.061**	0.038	0.094***	0.060**
	(0.025)	(0.031)	(0.027)	(0.025)	(0.030)	(0.027)
Copula(weeks)			0.029***			0.030***
			(0.0002)			(0.0002)
IMR				-0.237***	-0.835***	-0.446***
				(0.003)	(0.007)	(0.005)
Constant	-0.054	2.902***	-1.029***	0.327***	4.238***	-0.292***
	(0.048)	(0.060)	(0.052)	(0.048)	(0.060)	(0.051)
Obs.	11,215	11,215	11,215	11,215	11,215	11,215
R^2	0.27	0.37	0.78			

Note: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Significant Copula functions implies significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.3: Bootstrap Aggregation Results (Converted Earnings) -Uganda

	GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	0.008	0.027^{**}	0.022^{*}	0.008	0.027^{**}	0.021^{*}
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
sch2	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}
	(0.00005)	(0.00005)	(0.00004)	(0.0001)	(0.0001)	(0.0001)
age	0.081^{***}	0.077^{***}	0.068^{***}	0.082^{***}	0.076^{***}	0.069^{***}
	(0.0003)	(0.0002)	(0.0002)	(0.001)	(0.001)	(0.0005)
age2/100	-0.091^{***}	-0.088^{***}	-0.077^{***}	-0.092^{***}	-0.085^{***}	-0.079^{***}
	(0.0003)	(0.0003)	(0.0003)	(0.001)	(0.001)	(0.001)
female	-0.439^{***}	-0.448^{***}	-0.436^{***}	-0.454^{***}	-0.416^{***}	-0.454^{***}
	(0.001)	(0.001)	(0.001)	(0.008)	(0.009)	(0.008)
rural	-0.220^{***}	-0.289^{***}	-0.239^{***}	-0.234^{***}	-0.262^{***}	-0.255^{***}
	(0.001)	(0.001)	(0.001)	(0.007)	(0.008)	(0.007)
panel	0.147^{***}	0.214^{***}	0.156^{***}	0.145^{***}	0.216^{***}	0.154^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
weeks			1.171^{***}			1.171^{***}
			(0.001)			(0.001)
Copula(sch)	0.125^{**}	0.112^{*}	0.111^{*}	0.122^{**}	0.113^{*}	0.109^{*}
	(0.061)	(0.064)	(0.062)	(0.061)	(0.065)	(0.062)
Copula(sch2)	0.124^{**}	0.111^{*}	0.111^{*}	0.124^{**}	0.115^{*}	0.110^{*}
	(0.062)	(0.065)	(0.063)	(0.061)	(0.065)	(0.062)
Copula(weeks)			-0.005^{***}			-0.005^{***}
			(0.0002)			(0.0003)
IMR				0.036^{*}	-0.074^{***}	0.044^{**}
				(0.019)	(0.020)	(0.019)
Constant	-0.579^{***}	2.318^{***}	-1.864^{***}	-0.637^{***}	2.436^{***}	-1.933^{***}
	(0.098)	(0.104)	(0.100)	(0.112)	(0.120)	(0.116)
Obs.	4,631	4,631	4,631	4,631	4,631	4,631
R^2	0.36	0.39	0.60			

Note: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Significant Copula functions imply significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.4: Bootstrap Aggregation Results (Pay Periods) -Malawi

	GC			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	-0.0002	-0.144***	-0.093***	0.002	-0.143***	-0.094***
	(0.087)	(0.047)	(0.010)	(0.037)	(0.034)	(0.004)
sch2	0.006	0.012***	0.010***	0.006	0.012***	0.009***
	(0.006)	(0.002)	(0.001)	(0.004)	(0.001)	(0.0004)
age	0.039	0.079***	0.054***	0.036***	0.079***	0.044***
	(0.059)	(0.020)	(0.006)	(0.013)	(0.003)	(0.001)
age2	-0.033	-0.082***	-0.049***	-0.029*	-0.083***	-0.038***
	(0.073)	(0.026)	(0.007)	(0.016)	(0.003)	(0.001)
female	-0.006	-0.118	-0.117***	0.011	-0.122***	-0.041***
	(0.212)	(0.078)	(0.021)	(0.068)	(0.014)	(0.004)
rural	-0.288	-0.304***	-0.187***	-0.288***	-0.306***	-0.106***
	(0.202)	(0.078)	(0.019)	(0.020)	(0.009)	(0.004)
year	0.837***	0.990***	1.278***	0.818***	0.991***	1.303***
	(0.211)	(0.072)	(0.018)	(0.076)	(0.008)	(0.002)
weeks	1.386***	1.228***	1.148***	1.385***	1.228***	1.141***
	(0.241)	(0.074)	(0.027)	(0.021)	(0.005)	(0.002)
Copula(sch)	-0.020	0.143	0.137***	-0.008	0.139	0.141***
	(0.402)	(0.178)	(0.046)	(0.402)	(0.198)	(0.043)
Copula(sch2)	-0.013	0.139	0.139***	-0.027	0.140	0.145***
	(0.311)	(0.198)	(0.047)	(0.385)	(0.188)	(0.042)
Copula(weeks)	-0.034	-0.063***	-0.002	-0.034***	-0.063***	-0.001***
	(0.046)	(0.017)	(0.004)	(0.004)	(0.001)	(0.0003)
IMR				-0.063	0.015	-0.166***
				(0.238)	(0.046)	(0.008)
Constant	-2.392	-1.324**	-1.359***	-2.134*	-1.380***	-0.890***
	(1.518)	(0.605)	(0.168)	(1.280)	(0.397)	(0.084)
Obs.	182	505	5,129	182	505	5,129
R^2	0.44	0.66	0.77			

Note: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Significant Copula functions imply significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.5: Bootstrap Aggregation Results (Pay Periods) -Tanzania

	GC			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	-0.028***	-0.035***	-0.010	-0.050***	-0.040***	-0.019*
	(0.007)	(0.008)	(0.010)	(0.007)	(0.008)	(0.010)
sch2	0.005***	0.009***	0.006***	0.010***	0.010***	0.005***
	(0.001)	(0.001)	(0.0001)	(0.001)	(0.001)	(0.0001)
age	0.053***	0.046***	0.099***	0.028***	0.039***	0.085***
	(0.0002)	(0.0003)	(0.0003)	(0.001)	(0.001)	(0.0003)
age2/100	-0.065***	-0.053***	-0.095***	-0.028***	-0.043***	-0.078***
	(0.0003)	(0.0004)	(0.0004)	(0.001)	(0.002)	(0.0004)
female	-0.736***	-0.552***	-0.336***	-0.498***	-0.491***	-0.211***
	(0.001)	(0.001)	(0.001)	(0.005)	(0.011)	(0.003)
rural	-0.520***	-0.191***	-0.257***	-0.547***	-0.215***	-0.093***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.005)	(0.003)
panel	-0.216***	-0.129***	-0.089***	-0.173***	-0.126***	-0.074***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
weeks	1.074***	0.999***	1.044***	1.068***	0.998***	1.036***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Copula(sch)	0.032	0.022	0.094**	0.032	0.023	0.082**
	(0.043)	(0.045)	(0.039)	(0.043)	(0.046)	(0.039)
Copula(sch2)	0.031	0.023	0.094**	0.031	0.020	0.080**
	(0.043)	(0.044)	(0.039)	(0.044)	(0.045)	(0.039)
Copula(weeks)	0.074***	0.096***	0.009***	0.073***	0.096***	0.009***
	(0.0004)	(0.001)	(0.0003)	(0.0004)	(0.001)	(0.0003)
IMR				-0.689***	-0.190***	-0.503***
				(0.016)	(0.035)	(0.008)
constant	-0.481***	-0.597***	-1.846***	0.919***	-0.121	-0.782***
	(0.075)	(0.072)	(0.098)	(0.078)	(0.107)	(0.091)
Obs.	3,738	1,929	4,830	3,738	1,929	4,830
R^2	0.73	0.79	0.71			

Notes: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Significant Copula functions imply significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.6: Bootstrap Aggregation Results (Pay Periods) -Uganda

	GC			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	0.103^{**}	0.106^{*}	0.040^{***}	0.106^{**}	0.100^{*}	0.040^{***}
	(0.051)	(0.058)	(0.009)	(0.051)	(0.059)	(0.009)
sch2	-0.002^{***}	0.003^{***}	0.004^{***}	-0.002^{***}	0.003^{***}	0.004^{***}
	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0004)
age	0.063^{***}	0.078^{***}	0.078^{***}	0.065^{***}	0.074^{***}	0.081^{***}
	(0.001)	(0.001)	(0.0003)	(0.001)	(0.001)	(0.001)
age2	-0.082^{***}	-0.092^{***}	-0.081^{***}	-0.087^{***}	-0.082^{***}	-0.086^{***}
	(0.001)	(0.001)	(0.0004)	(0.001)	(0.002)	(0.002)
female	-0.592^{***}	-0.542^{***}	-0.308^{***}	-0.680^{***}	-0.405^{***}	-0.328^{***}
	(0.003)	(0.005)	(0.001)	(0.019)	(0.025)	(0.010)
rural	-0.233^{***}	-0.368^{***}	-0.229^{***}	-0.294^{***}	-0.313^{***}	-0.250^{***}
	(0.002)	(0.004)	(0.002)	(0.013)	(0.011)	(0.013)
panel	0.009^{***}	-0.062^{***}	0.187^{***}	0.013^{***}	-0.066^{***}	0.172^{***}
	(0.002)	(0.004)	(0.001)	(0.002)	(0.005)	(0.004)
weeks	1.164^{***}	1.158^{***}	1.157^{***}	1.163^{***}	1.154^{***}	1.162^{***}
	(0.002)	(0.004)	(0.001)	(0.002)	(0.004)	(0.002)
Copula(sch)	0.032	-0.094	0.080	0.028	-0.087	0.081
	(0.148)	(0.195)	(0.066)	(0.145)	(0.198)	(0.068)
Copula(sch2)	0.029	-0.092	0.080	0.027	-0.093	0.077
	(0.150)	(0.193)	(0.066)	(0.147)	(0.203)	(0.066)
Copula(weeks)	0.024^{***}	-0.020^{***}	-0.007^{***}	0.024^{***}	-0.020^{***}	-0.008^{***}
	(0.0005)	(0.001)	(0.0003)	(0.0005)	(0.001)	(0.0004)
IMR				0.151^{***}	-0.366^{***}	0.063
				(0.031)	(0.068)	(0.044)
constant	-1.804^{***}	-2.297^{***}	-2.406^{***}	-2.040^{***}	-1.457^{***}	-2.573^{***}
	(0.311)	(0.375)	(0.106)	(0.319)	(0.413)	(0.188)
Obs.	1,262	589	2,765	1,262	589	2,765
R^2	0.57	0.57	0.63			

Note: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Significant Copula functions imply significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.7: Bootstrap Aggregation Results (Pay Periods) - *Ganyu* Labour

Measure	GC DailyC	GC MonthlyC	GC MonthlyA	HGC DailyC	HGC MonthlyC	HGC MonthlyA
sch	-0.001	-0.018**	-0.007	-0.003	-0.024***	-0.010
	(0.006)	(0.008)	(0.007)	(0.006)	(0.008)	(0.007)
sch2	0.002***	0.003***	0.002***	0.002***	0.005***	0.003***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00003)	(0.00003)
age	0.034***	0.047***	0.040***	0.032***	0.039***	0.036***
	(0.00003)	(0.00003)	(0.00003)	(0.0001)	(0.0001)	(0.0001)
age2/100	-0.040***	-0.057***	-0.048***	-0.037***	-0.043***	-0.040***
	(0.00004)	(0.00005)	(0.00004)	(0.0001)	(0.0002)	(0.0001)
female	-0.240***	-0.444***	-0.341***	-0.223***	-0.367***	-0.299***
	(0.0002)	(0.0002)	(0.0002)	(0.001)	(0.001)	(0.001)
rural	-0.284***	-0.420***	-0.326***	-0.329***	-0.623***	-0.439***
	(0.0003)	(0.0004)	(0.0004)	(0.001)	(0.002)	(0.002)
year	1.326***	1.223***	1.217***	1.298***	1.097***	1.150***
	(0.0002)	(0.0003)	(0.0003)	(0.001)	(0.001)	(0.001)
weeks			0.982***			0.981***
			(0.0002)			(0.0002)
Copula(sch)	0.009	0.001	0.012	0.010	0.003	0.013
	(0.018)	(0.025)	(0.021)	(0.018)	(0.025)	(0.021)
Copula(sch2)	0.009	0.0002	0.012	0.010	0.003	0.013
	(0.018)	(0.025)	(0.021)	(0.018)	(0.025)	(0.021)
Copula(weeks)			0.020***			0.019***
			(0.0001)			(0.0001)
IMR				-0.095***	-0.429***	-0.236***
				(0.003)	(0.004)	(0.004)
Constant	0.217***	2.643***	-0.832***	0.361***	3.289***	-0.473***
	(0.031)	(0.042)	(0.034)	(0.032)	(0.042)	(0.036)
Obs.	16,528	16,528	16,528	16,528	16,528	16,528
R^2	0.60	0.38	0.77			

Note: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Significant Copula functions imply significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$