MULTIPLE-JOB-HOLDING IN AFRICA: THE CASE OF TANZANIA*

Theis Theisen*

02.02.2006

Abstract

In a sample of Tanzanian formal sector workers the vast majority have a desire for working longer hours in their main job, and supplement earnings through participation in informal production. Determinants of participation in informal production are examined through estimation of structural-form Logit models. A new way around the problem of measuring incomes from informal production is suggested and evaluated. The results show that participation in informal production is inversely related to income, and positively related to age. Age seems to play a very different role in the transition from work to retirement in Tanzania compared to industrialised countries.

JEL classification: D13; J22

Keywords: Multiple-job-holding; Home production; Informal sector; Participation; Labour supply

Address: Department of Economics and Business Administration, Agder University College, Serviceboks 422, 4604 Kristiansand S, Norway, tel. +47-38141526, fax +47 38141061. E-mail: Theis.Theisen@hia.no.

* This paper derives from a collaborative project with the late Ram Jogi. I am indebted to Ram Jogi and Elia W. Mwakagali for their efforts in data collection. Otto Andersen, S. Quamrul Ahsan, Jonathan Baker, Kristin Dale, Magnus Hatlebakk, Jochen Jungeilges and Ingrid Leiprecht are thanked for comments to earlier versions of the paper.

MULTIPLE-JOB-HOLDING IN AFRICA: THE CASE OF TANZANIA

1. Introduction

Individuals in developing countries often rely on various sources of monetary incomes. Leibbrand, Woolard and Woolard (2000) provide empirical evidence that this is the case in South Africa. Moreover, Glick (1999) documents that, in the case of Guinea, monetary earnings often are supplemented through extensive engagement in home production. The situation is likely to be similar in other countries of sub-Saharan Africa, and in other developing countries as well. This suggests that studies of labour market participation and labour supply in developing countries should start from the presumption that multiple-job-holding is the norm rather than the exception. Except for the work of Glick (1999) and Joliffe (2004) this is, however, not the state of the art in the economics literature. Hence, the main aim of the present paper is to examine the extent of multiple-job-holding in the African context, and to examine which factors drive individuals into multiple-job-holding.

My attention to the phenomenon of multiple-job-holding in developing countries was triggered by the observation that real wages of Tanzanian formal sector workers declined in much of the period from the early 1970s to the commencement of the restructuring period in the late 1980s. Such large and long-term declines in real wages bring up a number of questions. How are individuals able to cope in their daily lives? Has their material standard of living deteriorated correspondingly, or has lower real wages induced workers to increase their working hours, either in their main job or in additional jobs? In particular, has formal sector workers' participation in informal production increased?[1]

Glick (1999) and Joliffe (2004) address some of these questions, but due to data limitations they both had to use reduced form models. By contrast, since the appearance of the seminal article of Shishko and Rostker (1976) studies of multiple-job-holding in industrial

[1] The concept of the informal sector is by now fairly well established. In brief, it is usually taken to include small-scale enterprises, largely operating without a business permit, without paying taxes, and without adhering to a number of other regulations, cf. Feige (1990). The informal sector is also economically important. For instance, Schneider and Enste (2002) estimate that 54 % of the African labour-force hold a job in the informal sector.

countries have largely been based on structural form models, which usually are more informative than reduced form models. This methodological difference is largely due to the problems of observing what people in developing countries earn when working in the informal sector. Deaton (1997, p29) emphasises strongly the difficulties involved in income measurement in developing countries. In the present paper I therefore suggest a new way around this problem. The suggested method permits the estimation of a structural form multiple-job-holding model. To my knowledge, structural models of multiple-job-holding have not previously been estimated on data from developing countries.

The structural form approach chosen in the present paper places it in the tradition of Shisko and Rostker (1976). Their approach is, however, generalised by drawing on Gronau (1977) to include home production in the analysis. On the other hand, while Shisko and Rostker (1976) and most of the multiple-job-holding literature based on data from developing countries estimate labour supply equations, the present paper is limited to the study of participation in multiple-job-holding. This limitation is mainly due to the available data, and is related to the measurement problems pointed out above. The question of participation is, however, sufficiently complex to warrant a separate study.

Notice also that the present study focuses exclusively on whether workers holding a job in the formal sector also hold additional jobs in the so-called informal sector. Previous papers of Gindling (1991), Pradhan and van Soest (1995) and Funkhauser (1996) have examined factors that determine whether individuals in Brazil, Bolivia and Central America chose to work in the formal or the informal sector. These contributions are, however, all based on the assumption that an individual will work either exclusively in the formal sector, or exclusively in the informal sector, and they abstract from home production. Consequently, the present paper complements this vein of the literature.

In Section 2 we establish a simple theoretical model, which is used as a vehicle for deriving conditions for participation in home production, as well as labour supply functions for time allocated to such production, conditional on the individual holding a job in the formal sector. The basic model is subsequently extended to the case where there is no restriction on the number of jobs that can be held simultaneously. The econometric model developed in

Section 3 consists of two alternative specifications of the participation condition. In addition, in Section 3 the approach used to come around the problem of measuring earnings from informal production is presented and discussed. Section 4 contains a descriptive analysis of participation in informal production. Two main forms of participation are distinguished: home production for own consumption and income generating informal activities like selling products or own labour on the market. In Section 5 we estimate an earnings function that subsequently is used for predicting earnings from informal production. In Section 6 main results from estimating our two alternative participation models are presented, and in Section 7 we discuss some of the results in more detail. Section 8 summarises and provides some suggestions for further research.

2. Theoretical model

2.1 The basic model: Formal sector employment supplemented by home production

Consider a single individual whose behaviour is guided by the utility function $U(X,L)$, where X is the volume of a Hicksian aggregate good, and L is leisure. Utility is maximised subject to two constraints. First, there is a goods constraint. When the unit price of the aggregate good is normalised to 1, the goods constraint takes the form $X = V + W\overline{N} + ZF(H)$, where V is non-labour income, W is the formal sector wage rate, $\overline{N}$ is hours worked in the formal sector, Z is a productivity index, and F(H) is a function for which we assume that $F(0) = 0$ and that $F(H)$ is increasing and concave in H ($F_H > 0$ and $F_{HH} \leq 0$, with subscripts to functions denoting partial derivates). The goods constraint lies out that goods consumption can be "financed" from three sources: (I) non-labour income (V), (II) formal sector wage earnings ($W\overline{N}$), and (III) home production ($ZF(H)$). Notice that our home production function corresponds closely to that of Gronau (1977), and that the model implies the assumption that home produced goods and goods bought in the market are perfect substitutes in consumption. Finally, notice that other factors of production than household time, like capital, land, skills, etc., may affect the Z-term in the home production function, but are not specified in the theoretical model.

The second restriction under which our individual maximises his or her utility is the time constraint, $L + H + \overline{N} = T$, where T is total time available, with L, H, $\overline{N}$ and T all non-negative. Formal sector working-time is assumed to be determined from the demand side of the labour market; hence it is exogenous to the individual, as indicated by the bar above N. In addition, we assume that an individual who wants a formal sector job typically must accept to work full time. This assumption is in accordance with the observation that part-time jobs in the formal sector are practically non-existent in Tanzania.[2]

When combined, the goods constraint and the time constraint yield the budget constraint ABCDE in Figure 1. With the indifference curves drawn in the figure utility is maximised at Q. In the present paper we focus, however, on participation in home production (and other forms of informal production). Consider therefore the situation at point D in Figure 1, where the individual is working $\overline{N}$ hours in the formal sector. The crucial question is whether he or she will participate in home production. The answer follows directly from Figure 1, where it is easily seen that the individual will participate in home production only if the indifference curve at point D slopes less than the home production function. Writing U_X and U_L for the marginal utilities of the aggregate good and leisure, the participation condition takes the form:

$$\frac{U_L(V + W\overline{N}, T - \overline{N})}{U_X(V + W\overline{N}, T - \overline{N})} \begin{cases} \geq ZF_H(0) & \Rightarrow \text{ no participation } (H = 0) \\ < ZF_H(0) & \Rightarrow \text{ participation } (H > 0) \end{cases} . \qquad (1)$$

Condition (1) implies that the individual will participate in home production if the utility forgone by one hour less leisure falls short of the utility gained through the output from an extra hour in home production. This is the case if, at D, the value of the marginal productivity in home production $(ZF_H(0))$ exceeds the individual's shadow wage, $W^* = U_L/U_X$. Furthermore, for a participant in home production utility can be gained through increasing hours in home production up to point Q in Figure 1. At Q the value of the marginal

[2] The validity of this restriction is critical to the analysis. Deardorff and Stafford (1976) and Oswald and Walker (1993) demonstrate that there may be strong theoretical reasons for the existence of restrictions on workers' choice of working hours. Biddle (1988) and Moffitt (1982) provide empirical evidence.

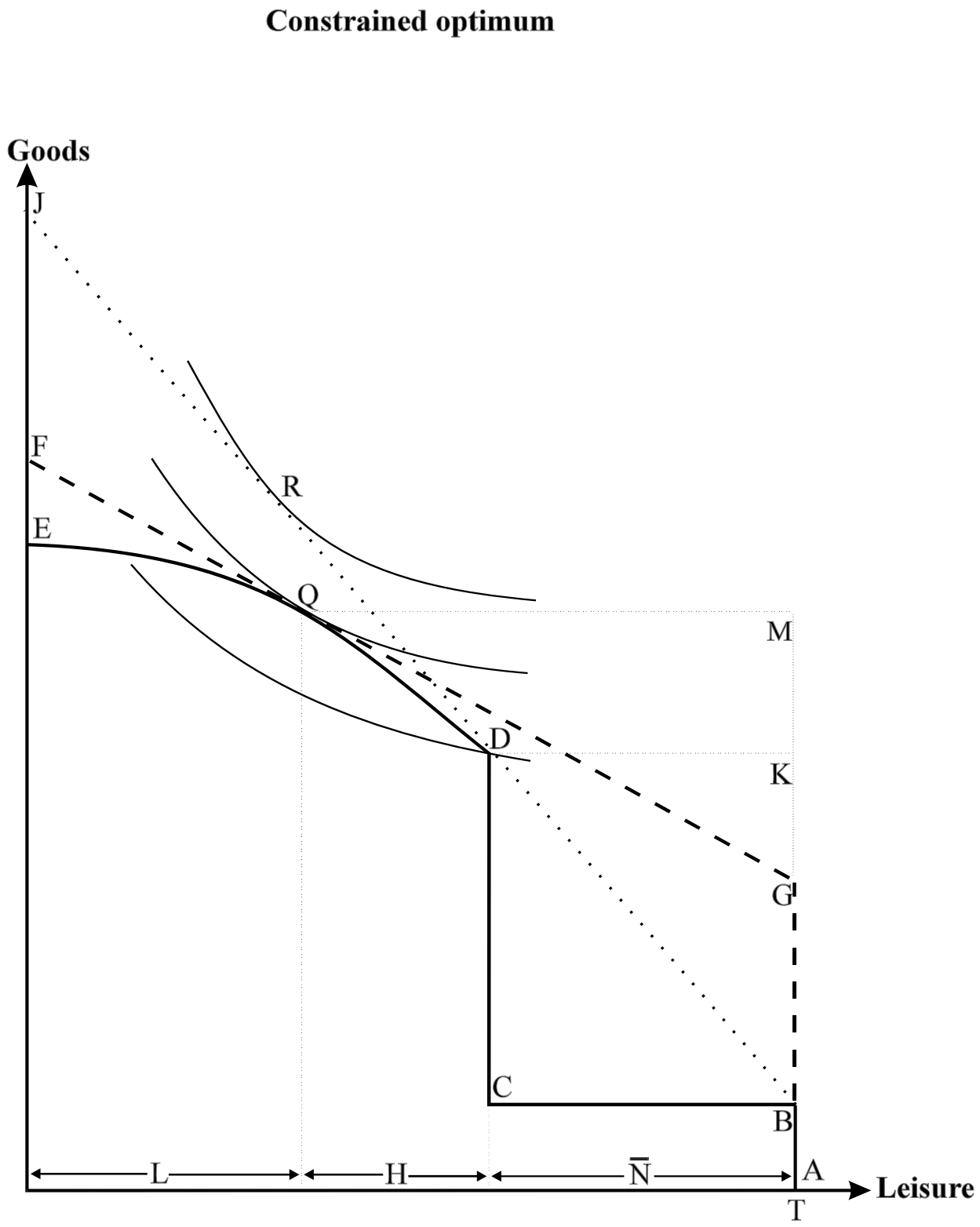

Figure 1 Consumer choice

productivity equals the shadow wage. Hence, for a participant in home production we have the first order condition:

$$\frac{U_L\left(V + W\overline{N} + ZF(H), T - \overline{N} - H\right)}{U_X\left(V + W\overline{N} + ZF(H), T - \overline{N} - H\right)} = ZF_H(H). \qquad (2)$$

This yields the conditional supply function for hours worked in home production:

$$H = f\left(\hat{W}, \hat{V}\right) - \overline{N}, \qquad (3)$$

where $\hat{W} = ZF_H(H)$ is the value of the marginal productivity in home production, which is equivalent to the concept of a "virtual wage rate" used in the following. In Figure 1, the virtual wage rate corresponds to the slope of the linearized budget constraint GQF supporting the optimum at Q. $\hat{V}$ in Eq. (3) is "virtual non-earned income", which equals the distance AG in Figure 1. More precisely, $\hat{V} = V + \left(\overline{W} - \hat{W}\right)H + \left(W - \hat{W}\right)\overline{N}$, where $\overline{W} = ZF(H)/H$ is the average "wage rate" in home production.[3] Killingsworth (1983, pp 88-91) discusses the use of virtual concepts, which have the great advantage that the conditional supply function can be written on the simple structural form given by Eq. (3). In Eq. (3) $f\left(\hat{W}, \hat{V}\right)$ is the (total) labour supply function, conditional on holding a full-time formal sector job. Since $\hat{W}$ and $\hat{V}$ generally are functions of H, Eq. (3) gives the supply of labour to home production only on implicit form. This complication is handled in Section 3 by imposing additional restrictions on the home production function.

The restriction on formal sector working hours plays an important role in our model. Figure 1 illustrates, however, only one of the cases that may occur, namely the case of a worker who is "under-employed" in his or her formal sector job. Since the virtual wage rate at the optimum (Q) in Figure 1 falls short of the formal sector wage rate, the worker would have preferred to work more than $\overline{N}$ hours in the formal sector, and to reduce working hours in home production. Due to the restriction on $\overline{N}$ this is not possible. Hence, the individual can be characterised as under-employed in the formal sector.

---

[3] From figure 1, $\hat{V}$ =AG=AB+BK+KM-MG= $V + W\overline{N} + ZF(H) - \hat{W}\left(\overline{N} + H\right) =$
$V + \left(ZF(H) - \hat{W}H\right) + \left(W - \hat{W}\right)\overline{N} = V + \left(ZF(H)/H - \hat{W}\right)H + \left(W - \hat{W}\right)\overline{N}$
$= V + \left(\overline{W} - \hat{W}\right)H + \left(W - \hat{W}\right)\overline{N}$

There may also be cases of "over-employment" in the formal sector. This may occur if the function $F(H)$ is strongly concave. The individual may then want to work only a few hours in home production. In order to obtain a desired volume of goods, however, work in home production has to be combined with a formal sector job. The utility of such an individual could be increased by reducing somewhat the number of hours worked in the formal sector - and expanding hours worked in home production. Due to the restriction on $\overline{N}$ such a reallocation of hours is not possible. Hence, the individual can be characterised as over-employed in the formal sector. In Section 4 we examine empirically whether individuals in our sample are over-employed or under-employed in the formal sector job.

2.2 Multi-person households and multiple forms of informal work

The model set out above can easily be extended to the case of a multi-person household. Such a strategy is not followed in the present paper, since the available data precludes exploiting the finer structure of a multi-person model. Consequently, we follow the strategy of including incomes earned by other household members than the one in focus, in the exogenous non-earned income variable. In a similar vein, hours devoted to home production by other household members than the one we focus on are taken as exogenously given. Shiskho and Rostker (1976), Conway and Kimmel (1998) and a number of other papers on multiple-job-holding tacitly follow a similar strategy.

Our model can easily be extended to multiple goods.[4] In such a version of the model, home production of some goods may exceed consumption. For such goods there will be a surplus to sell on the market, and the individual is in effect part-time self-employed. Home production and self-employment are, however, distinguished only by whether $X_H^i < X^i$ or

[4] In order to extend the model to multiple goods, interpret X, X_M, X_H and H as n-element vectors, replace the unit price of X by an n-element price vector, $P' = (p^1 ..... p^n)$, and let the production function for each good take the form $X_{H^i} = Z^i F^i(H^i)$, with the same properties as the production function already introduced. Under the simplifying assumption that all goods are consumed in optimum, this leads to conditions $U_{X^i}/p^i = U_{X^j}/p^j$ characterising optimal allocation of household resources to the consumption of different goods. The individual will now participate in home production of good i if, at point D in the multidimensional space, we have $U_L/U_{X^i} < Z^i F_H^i$, and not participate if $U_L/U_{X^i} \geq Z^i F_H^i$. From the hours' conditions and the conditions for optimal allocation of household resources to goods consumption we also obtain, for all goods $i \neq j$ that are produced in optimum, the relationship $p^i Z^i F_{H^i}^i = p^j Z^j F_{H^j}^j = \hat{W}$. This relationship characterises optimal allocation of time to the production of different goods at home, and makes clear that for each individual a common virtual wage rate $(\hat{W})$ applies for all goods he or she produces at home.

$X_H^i > X^i$, where H stands for home production and i indexes goods. The individual may also engage in part-time employment in a second (informal) job. Under the assumptions stated in footnote 4 there is, however, no need to distinguish between home production, part-time self-employment and part-time employment as a worker in the informal sector. In other words, the real value of wage income earned in an informal sector job may be aggregated with output from home production and self-employment to form an aggregate good. Hence, we will in the following use the more general term informal production rather than home production, and proceed as if the individual participates in informal production of a single aggregate good. This facilitates substantially the econometric analysis, to which we now turn.[5]

3. Econometric model

3.1 Participation in informal production

We parametrise the r.h.s. of the participation condition by assuming that the volume of informal production is proportional to hours allocated to such production $\left(X_H = ZH\right)$, implying that the virtual and average wage rates in the informal sector co-inside. This may seem restrictive, but it is in line with the assumption of an exogenous wage rate made in the vast majority of labour supply and multiple-job-holding studies.[6]

For the l.h.s of the participation condition we consider two alternative parametrisations, based on two different specifications (A and B) of the utility function:

$$U\left(X, H+\overline{N}; R, \varepsilon_A\right) = \left(\frac{\omega_{A\hat{V}}\left(H+\overline{N}\right)-\omega_{A\hat{W}}}{\left(\omega_{A\hat{V}}\right)^2}\right)\exp\left\{\frac{\omega_{A\hat{V}}\left(\omega_{A0}+\omega_{A\hat{V}}X+\omega'_{AR}R+\varepsilon_A\right)-\omega_{A\hat{W}}}{\omega_{A\hat{V}}\left(H+\overline{N}\right)-\omega_{A\hat{W}}}\right\}, \quad (4a)$$

[5] The fact that informal production may be aggregated to a single good does not preclude us from studying separately individuals' participation in home production, self-employment and part-time wage earning, or from studying multiple-job-holding with even more narrowly defined goods. In the present paper, however, we consider informal sector participation only at the aggregate level.

[6] Moffitt (1984) is one of the few papers with an endogenous hourly wage rate, determined simultaneously with hours worked. I am not aware of multiple-job-holding studies that deviate from the assumption of exogenous earnings per hour. On the other hand, with access to data on each individual's input of time in informal production, and the volume of goods produced informally, it would be possible to estimate an informal sector production function. The estimated production function could in turn be used for estimating each individual's average and virtual wage rate in informal production. The data available in the present study are, however, insufficient to pursue such a strategy. Hence, in the empirical part of this paper there is actually no choice but to assume proportional home production functions.

$$U(X, L; R, \varepsilon_B) = \left(V + W\overline{N} + Z\left(H - \varepsilon_B - \mu - \omega'_{BR}R\right)\right)^{\alpha}\left(T - \overline{N} - \left(H - \varepsilon_B - \mu - \omega'_{BR}R\right)\right)^{\beta}, \qquad (4b)$$

where the ω_{Ai}'s $(i = 0, \hat{W}, \hat{V})$, μ, α and β are single parameters, and ω'_{AR} and ω'_{BR} are row vectors of parameters. The stochastic error terms, ε_A and ε_B, can be interpreted as taste shifters that are known to the individual worker, but unknown to the researcher. R is a column vector of taste shifters known not only to the individual worker, but also to the researcher. For further discussion of Specification A in (4a) and Specification B in (4b), see Pencavel (1986) and Killingsworth (1983).

Our parametrisations lead to two alternative participation conditions:

$$\omega_{A0} + \omega_{A\hat{V}}\hat{V}^k + \omega_{A\hat{W}}\hat{W}^k + \omega'_{AR}R^k - \overline{N}^k + \varepsilon_A^k \begin{cases} \leq 0 & \Rightarrow \text{ Non-participation} \\ > 0 & \Rightarrow \text{ Participation} \end{cases} \quad (k=1, , K), \qquad (5a)$$

$$\omega_{B0} + \omega_{B\hat{V}}\frac{\hat{V}^k}{\hat{W}^k} + \omega'_{BR}R^k - \overline{N}^k + \varepsilon_B^k \begin{cases} \leq 0 & \Rightarrow \text{ Non-participation} \\ > 0 & \Rightarrow \text{ Participation} \end{cases} \quad (k=1, , K), \qquad (5b)$$

where k indexes individuals, $\omega_{B0} = \mu + \alpha T/(\alpha + \beta)$, and $\omega_{B\hat{V}} = -\beta/(\alpha + \beta)$, cf. Appendix A.

For a non-participant in informal production, the l.h.s. of Conditions (5a) and (5b) are negative, and can be interpreted as the latent supply (H_*^k) of hours to informal production. For a participant in informal production, the l.h.s. of Conditions (5a) and (5b) are positive, and give the actual (=latent) hours supplied to informal production. Hence, we can write down the two variants of the conditional latent informal labour supply functions:

$$H_*^k = \omega_{A0} + \omega_{A\hat{V}}\hat{V}^k + \omega_{A\hat{W}}\hat{W}^k + \omega'_{AR}R^k - \overline{N}^k + \varepsilon_A^k \qquad (k=1, , K), \qquad (6a)$$

$$H_*^k = \omega_{B0} + \omega_{B\hat{V}}\frac{\hat{V}^k}{\hat{W}^k} + \omega'_{BR}R^k - \overline{N}^k + \varepsilon_B^k \qquad (k=1, , K). \qquad (6b)$$

Eqs. (6a) and (6b) correspond to Eq. (3), but notice that our assumptions imply that the dependent variable in Eqs. (6a) and (6b) appears only on the l.h.s., while it appeared on both sides of Eq. (3).

Next, define the indicator, $\tilde{H}_*^k$, equal to 1 if an individual participates in informal production, equal to 0 if he or she does not participate. This yields the participation model:

$$\tilde{H}_*^k = \begin{cases} 0 \text{ if } H_*^k \leq 0 \\ 1 \text{ if } H_*^k > 0 \end{cases} \qquad (k=1,\ldots..,\ K), \qquad (7)$$

with H_*^k given by Eq. (6a) or (6b). Assuming that the error-terms $\left(\varepsilon_A^k, \varepsilon_B^k\right)$ are logistically distributed, relationship (7) taken together with either (6a) or (6b) constitute two variants of a logit model.

3.2 Modelling the virtual wage rate

With data on the virtual wage rate, and on human capital variables, one could use the sub-sample of those actually working in the informal sector for estimating a virtual wage equation. The estimated virtual wage equation could in turn be used to predict the virtual wage rates for all individuals in the sample. Such a procedure was pursued in the seminal paper of Shishko and Rostker (1976), and has become standard in multiple-job-holding studies based on data from developed economies. In the few multiple-job-holding studies conducted on data from developing countries, however, researchers have so far not had access to virtual wage data. Hence, they have typically relied on reduced-form approaches, cf. Glick (1999), Jolliffe (2004), and Theisen (2005). This methodological difference reflects the much greater difficulties in measuring the virtual wage rate in developing countries than in developed countries. These measurement problems are inter alia due to the fact that individuals in developing countries often hold several informal jobs simultaneously, switch back and forth between different informal activities over time, and are heavily engaged in

home production, for which the volume is hard to measure and the value is difficult to impute. Such complications led Deaton (1997, p29-30) to a sceptical view on income measurement in developing countries.[7] Similar complications arise in the measurement of informal sector working hours. Since the virtual wage rate usually is calculated as the ratio between the income earned and hours worked in the informal sector, the consequence of these measurement problems is that virtual wage rate data often are unavailable in developing countries. In fact, these measurement problems are also the main reason why we in the present paper carry out a participation study rather than estimating an informal labour supply function.

As already mentioned, a reduced form approach provides one possible solution to the problem of a non-observable virtual wage rate. Theisen (2005) demonstrates that a reduced form approach may provide consistent estimates, but only under restrictions that in practice are unlikely to be completely fulfilled. In addition, a reduced form approach is less informative than a structural form approach. Hence, in the present paper we will take a step forwards by pursuing a structural form approach. It takes as the starting point that wages in any sector of the economy can be modelled by sector specific earnings functions of the type introduced by Mincer (1974):

$$\mathrm{Ln}\hat{W}_j^k = \eta_j + \sigma_j' S^k + e_j^k \qquad (j = -2, -1, 0, 1, 2;\ k=1,\ldots.., K), \qquad (8)$$

where $\mathrm{Ln}\hat{W}_j^k$ is the natural logarithm of the hourly wage rate, η_j is a constant term, σ_j' is a row vector of parameters, S^k is a column vector of explanatory variables, e_j^k is a stochastic error term, and j indexes sector of employment (with j = - 2 for home production, j = - 1 for the (monetary) informal sector, j = 0 for the private formal sector, j = 1 for the parastatal sector, and j = 2 for the public sector).

[7] Deaton (1997, p30) summarises these problems by stating that "The practical and conceptual difficulties of collecting good income data are severe enough to raise doubts about the value of trying; the costs are large and the data may not always be of great value once collected. Apart from some early experiments, the Indian NSS has not attempted to collect income data in their consumer expenditure surveys." NSS here stands for National Sample Survey.

When data on the virtual wage rate are unavailable, it is not possible to estimate the informal sector earnings function directly. The informal sector is, however, a sub-sector of the private sphere of the economy, which encompasses three sub-sectors: (I) the private formal sector, (II) the (monetary) informal sector, and (III) the (non-monetary) home production sector. If workers flow freely between the formal and informal parts of the private sector, it can be argued that competition between workers about jobs, and competition between employers about workers will work in the direction of an equilibrium where an individual with a given set of characteristics earns an (untaxed) "wage rate" in the informal (private) sector that is equal to the after tax wage rate he or she could earn in the formal private sector. On the assumption that this wage-equalization process works sufficiently smoothly, earnings functions in the formal and the informal parts of the private sector will tend to be similar. Hence, we assume: $(\eta_{-2}, \sigma'_{-2}) = (\eta_{-1}, \sigma'_{-1}) = (\eta_0, \sigma'_0)$. Gindling (1991) found empirical evidence that the last of these equalities may hold as an approximation, but Funkhauser (1996) obtained diverging results.[8] Notice also that the assumption of wage-equalization can be seen as an extension of the result obtained in Section 2 that a common virtual wage rate applies in all sub-sectors of informal production. In terms of Figure 1 our assumption means that workers in the private part of the formal sector face a budget constraint ABCDJ, which within our model implies that underemployment is the reason for private sector workers' participation in informal production. For workers in the government or parastatal sectors, however, the budget constraint may look very different. We return to the implications of this in Section 5.

One important complication remains: Since the parameters (η_0, σ'_0) of Eq. (8) must be estimated from the sub-sample of workers holding a job in the formal private sector, application of OLS is likely to lead to sample selection bias. As demonstrated by Lee (1983), however, this problem can be solved through a two-step procedure.[9] The first step is to model

[8] Funkhauser obtained results implying that the returns to schooling in the informal sector of Central American countries on average fell 15 – 20 per cent short of returns in the formal sector. Since schooling is by far the most important variable in the earnings function, this suggests that using the earnings function for the formal part of the private sector to predict the virtual wage rate may lead to an overestimation of the virtual wage rate. In Section 7 we will therefore report results from a sensitivity analysis carried out in order to assess the consequences of possible distortions in the predicted virtual wage rates.

[9] For applications of the technique, see for instance Trost and Lee (1984), and Lassibille and Tan (2004).

the allocation of workers to sectors of formal employment. In the present case, with three sectors of formal sector employment (sector 0: private, sector 1: parastatal, sector 2: public), we use a multinomial logit model, which amounts to estimating the parameter vectors α'_j in:

$$P_j = \frac{\exp(\alpha'_j Z)}{1+\sum_{j=1,2}\exp(\alpha'_j Z)} \qquad (j=1, 2), \tag{9}$$

where P_j denotes the probability that a worker is allocated to sector j, Z is a vector of covariates, and where the coefficient vector for the private sector, α_0, is normalised to zero, cf. Green (2000) for details.

The next step is to use the estimation results for the multinomial logit model to generate the following selection term for the private sector earnings function:

$$\hat{\lambda} = \varphi[J(\hat{\alpha}'_j Z_j)]/F(\hat{\alpha}'_j Z_j), \tag{10}$$

where φ is the standard normal density function, F is the logistic marginal distribution, and J is equal to $\Phi^{-1}F$, where Φ is the cumulative distribution function of the standard normal distribution.

In the second stage we include the selection term defined in Eq. (10) in the earnings function. Hence, we obtain:

$$\mathrm{Ln}\hat{W}^k = \eta + \sigma' S^k + \tau\ \lambda^k + \varepsilon^k \qquad (k=1,\ldots..,K), \tag{11}$$

where τ^k is an unknown parameter to be estimated, ε^k is an error-term, and where we for simplicity have omitted sector indexes. The relationship between the error terms of Eqs. (8) and (11) are given by $E(e_0^k | \text{Sectoral choice}) = \tau\ \lambda^k + \varepsilon^k$. Application of ordinary least squares to Eq. (11) will give consistent estimates of all parameters in the earnings function. Moreover, this approach allows us to test whether sample selection is present in our case or not.

3.3 The complete econometric model, its predictions, and test strategies

Our econometric approach can now be summarised as follows: First, we employ a multinomial logit model for the allocation of workers to their formal sector of employment (private, parastatal, or public sector). Second, we estimate an earnings function from the sub-sample of workers in the private formal sector, as well with a selection term included, as with no selection term, and test whether sample selection is a problem or not, cf. Section 5. Third, we use the estimated private sector earnings function to predict the virtual wage rate for all individuals. The predicted virtual wage rates are in turn used for calculating virtual non-earned income. Fourth, we estimate the two variants of the logit model given by relationship (7) combined with either (6a) or (6b), cf. Section 7

From the presumption that leisure is a non-inferior good, it follows that the virtual income parameter and the real virtual income parameter $\left(\omega_{A\hat{V}} \text{ and } \omega_{B\hat{V}}\right)$ will take a negative sign in both specifications. The parameter $\omega_{A\hat{W}}$, which captures the uncompensated wage effect in Specification A, cannot be signed from the assumptions made above. In the more restrictive Specification B, however, it follows from $\omega_{B\hat{V}} < 0$ that the uncompensated wage effect will be positive and of the same magnitude as the negative income effect. Furthermore, in both specifications the coefficient affiliated with formal sector working time is expected to be equal to -1.

4. Data collection, descriptive analysis and assessment of modelling assumptions

4.1 Data collection

Data were collected in three-steps. First, five towns were chosen as places to conduct interviews.[10] Second, 45 formal sector organisations of different size, and from different sectors, were selected. Finally, workers to be interviewed were selected within each

[10] The five towns were Dar es Salaam, Dodoma, Iringa, Morogoro and Mzumbe. These towns represent the variety in Tanzanian nature, industrialisation and population. Dar es Salaam is by far the largest of the five towns, and continues to be the commercial centre of the country. The new capital, Dodoma, is the centre of the Dodoma region, and is dominated by Government organisations. Only a few sizeable private companies and state-owned (parastatal) companies are located in Dodoma. Iringa and Morogoro are regional administrative and economic centres. The small ward Mzumbe is located close to Morogoro.

organisation. At the last stage managers were asked to assist in selecting workers randomly within their organisation.[11] This resulted in a response rate close to 100 %, and a sample of 261 workers. Because of non-response to questions concerning participation in informal production, the analyses in the following are, however, based on a sample of size 247.

The sample contains only workers holding a formal sector job. Delimiting the sample to formal sector workers seems natural, since our focus is on whether formal sector workers are multiple-job-holders. Restricting the sample in the manner just described also enables us to address the phenomenon of multiple-job-holding with a sample of modest size.[12]

4.2 A descriptive examination of participation in informal production

Formal sector workers were asked about their participation in home production for own consumption as well as in income generating informal production. Data on participation in home production cover six goods that are most important in the diets of Tanzanian households: maize, rice, fruit, vegetables, eggs and milk. Data on participation in income generating informal production cover the production of agricultural commodities for sale, catering, working as craftsmen, etc. Some ways of earning monetary incomes are in effect informal part-time self-employment (agriculture), which is distinguished from home production mainly by the scale of the activity and by the fact that output is sold in the market. Other informal ways of earning monetary incomes (private teaching) come very close to holding a second job as a part-time wage earner.[13]

---

[11] In order to avoid strategic selection by the manager, workers were selected "on-the-spot", after enumerators had arrived at the workplace, and in collaboration with enumerators. In addition to the main questionnaire used to collect data from workers, a separate questionnaire was used to interview managers. A comparison of answers given by workers and their manager revealed that workers to some extent gave answers deviating from the manager's answers to similar questions, even on sensitive issues. This indicates that managers' role in selecting workers hardly lead to serious problems like strategic selection or manipulation of workers to give the "right" answers.

[12] In order to obtain a sufficient number of formal sector workers in a random sample drawn from the entire Tanzanian population, the sample would have to be much larger than permitted by the resources available for the present study. Shisko and Rostker (1976) and Conway and Kimmel (1998) also restricted their samples to individuals holding at least one job, even though they had access to samples representative of the entire population.

[13] The complete list of informal income generating activities includes husbandry, poultry, piggery, making pancaces, roasting meat, selling ice cream, running a bar, running a shop, working as a disc jockey, tailoring, sewing, weaving, doing laundry, shoeshining, carpentry, masonry, car repair, working as a draughtsman, transportation by car or lorrey, milling, teaching, and being a landlord with a house for rent.

The sample of size 247 consists of 90 persons (36%) who did not participate in informal production and 157 persons (64 %) who participated in informal production. Of the 157 participants in informal production 145 took part in home production, 59 took part in income generating informal production, and of the 59 participants in income generating informal production as much as 47 participated also in home production.

Participation rates differed between subgroups of the sample, with a rate of 66 % for men but 58 % for women. Among women with children below school age the participation rate was 49 %, while it was 73 % for women without children below school age. Among workers in the public sector 72 % participated in informal production, compared to 58 % in the parastatal sector and 63 % in the private sector. Finally, while only 43 % of formal sector workers living in Dar es Salaam participated in informal production, 72 % of those living in the other towns included in the sample took part in informal production.

A closer examination of answers to open ended questions revealed that formal sector workers of almost all kinds participated in home production. In income generating informal production, however, there seems to be a more distinct pattern of specialisation: Skilled blue collar male workers who held formal sector jobs as mechanics, carpenters, masons etc. usually carried out the same type of professional work in the informal sector. Workers holding administrative positions in the formal sector were more often engaged in informal trade, or in agricultural production on a large scale, like running a farm. Unskilled female workers earned extra by doing laundry, selling ice cream, selling pancakes, etc. In other words, unskilled female workers seem to carry out similar work in the informal sector as they do in the household. Taken together, these observations indicate that formal sector workers who earned monetary incomes informally usually participated in informal activities where they were likely to have a comparative advantage.

4.3 Assessment of two modelling assumptions

In Section 2 we distinguished between workers who were under-employed and over-employed in their formal sector job. In order to throw some light over the relative importance of these categories, respondents were asked if they would like to work overtime in their

formal sector job. Of the 197 workers who responded to this question (with a clear yes or no, rather than with don't know), 176 (90 %) answered the question in the affirmative, indicating that the vast majority in our sample participated in informal production because they were under-employed in their formal sector jobs. This in turn suggests that it is warranted to disregard the possibility that a significant number of individuals in our sample took part in informal production because they found less displeasure in work in the informal sector compared to work in the formal sector. This provides empirical support for our assumption that work in the formal and the informal can be regarded as perfect substitutes. Without this assumption, a model of the type developed by Conway and Kimmel (1998), would have been required.

In Section 3 we introduced the assumption of proportional informal production functions. In the case of no uncertainty it follows from proportional informal sector production functions that each individual in optimum either will participate in informal production of a single good, or not participate at all. Once stochastic elements are included in the model, however, participation in informal production of more than one good may occur even with proportional production functions. Weather conditions, for instance, introduce a random element in the relationship between inputs and output in home production of food. The individual may then, as a hedge against uncertainty, chose to participate in the production of several goods in addition to the good where he or she has the maximum productivity. Uncertainty concerning future market prices for goods may work in a similar way. On the reasonable assumption that an individual's productivity differs across goods it is, however, likely that the individual will participate in production of only a small number of goods. This prediction gains empirical support from our data, which reveal that among those who participated in home production, 88 persons (61 %) produced only one good, 32 (22 %) produced two, 19 (13 %) three, and 10 (7 %) produced four goods. No individual produced more than four of six goods examined.

4.4 Descriptive statistics for variables in the participation equations

Table 1 provides definitions and descriptive statistics for all variables used in the estimation of participation functions. Notice that there in the sample is limited variation in

formal sector working hours. The impact coefficient of this variable may therefore be hard to identify with our data. We return to this in Section 6.

Table 1. Definitions and descriptive statistics of variables in the participation functions

Variable	Definition	Mean	St. dev.
Informal work	Binary variable equal to 1 if respondent participates in some form of informal production, otherwise 0	0.6356	0.4822
Virtual income	Total household expenditures minus monetary incomes earned by the respondent through informal work minus the income the respondent would have earned by working the hours actually worked in the formal sector at the virtual wage rate. Measured on a monthly basis in 1000 Tanzanian Shillings, 1991	10.1308	17.8486
Virtual wage rate	The wage rate that the respondent is predicted to earn if working in the private sector. Tanz. Shillings, 1991	9.2104	5.2056
Real virtual income	Virtual income divided by the virtual wage rate	1.6866	2.4522
Age	Age of respondent, measured in years	35.1336	8.0770
Size	Number of persons in the respondent's household	3.7652	2.4581
Female	1 if respondent is female, 0 if male	0.2713	0.4455
Mother	1 if respondent is female and has children below 7 years of age, 0 if not	0.1660	0.3728
School-aged	Number of children in respondent's household between 7 and 15 years of age	1.0850	1.3453
Dar es Salaam	1 if respondent lives in Dar es Salaam, 0 if not	0.3077	0.4625
Hours	Hours worked per day in the formal sector	8.0121	0.4640
Public	1 if respondent holds a full-time job in the public sector, 0 if not	0.3482	0.4774
Parastatal	1 if respondent holds a full-time job in the parastatal sector, 0 if not	0.4170	0.4941

5. Predicting the virtual wage rate

From Section 3 it follows that the first step in predicting the virtual wage rate amounts to estimating the parameters of the earnings function specified in Eq. (11), using the sub-sample of 58 private sector workers. Table 2 provides definitions and descriptive statistics of the variables that enter the earnings function, including the selection variable which has been calculated from estimated parameters of the multinomial logit model, for which estimation results are shown in Table B2 of Appendix B. Two variants of the earnings function have been estimated: one with the selection term included, and one without the selection term.

Estimation results are contained in Table 3. The estimated function fits the data very well. The coefficient of the selection term is insignificant at standard levels of statistical significance, indicating there is no sample selection bias in the earnings function. Hence, when predicting the virtual wage rate we use the estimation results for the earnings function without a selection term. From Table 3 it is seen that except for the variable Experience squared, which is highly correlated with Experience, all independent variables in that earnings function have a statistically significant impact on the dependent variable. Moreover, all the variables with statistically significant coefficients have an economically significant impact on the predicted wage rate, and they carry the signs usually found in earnings function studies. Hence, the estimated earnings function seems to provide us with a reasonably good basis for predicting the virtual wage rate.

Based on the assumption $(\eta_{-2}, \sigma'_{-2}) = (\eta_{-1}, \sigma'_{-1}) = (\eta_0, \sigma'_0)$ our proxy for the virtual wage rate can be determined using the parameter estimates in Table 3. For the 58 individuals in the sample who hold a job in the formal private sector, the predicted virtual wage rate deviates of course only randomly from their actual formal sector wage rate. The picture is very different, however, for the 189 individuals in sample who hold a formal sector job in one of the two other sectors. For 80 % of those who hold a job in the public sector, the predicted virtual wage rate exceeded their formal sector wage rate. For parastatal sector workers the picture is more mixed: those at very low and very high formal sector wage rates earned higher wages in their formal sector job than our predictions show that they could earn in the informal sector.

These results are in line with the "compressed wage structure" observed by Knight and Sabot (1980) in the Government and Parastatal sectors of Tanzania, relative to the private sector.

Table 2 Definitions and descriptive statistics of variables in the earnings function. Sub-sample of formal private sector workers, n = 58.

Variable	Definition	Mean	St dev.
$\text{Ln}\hat{W}_P^k$	The natural logarithm of salary	1.8542	0.5570
Schooling	Years of formal schooling	8.3621	3.1992
Experience	Age minus years of formal schooling minus 7	18.4483	8.8460
Expsq	Experience squared divided by 100	4.1724	4.0736
Female	1 if respondent is female, otherwise zero	0.2241	0.4207
Dar es Salaam	1 if respondent lives in Dar es Salaam, otherwise zero	0.3276	0.4734
Lambda	λ calculated by means of Eq. (10)	0.3557	0.2209

Table 3 Estimation results for the earnings function [*]

	Without selectivity correction		With selectivity correction	
Variable	Coefficient	Standard error	Coefficient	Standard error
Constant	0.1441		- 0.0570	
Schooling	0.1311***	(0.0161)	0.1364***	(0.0321)
Experience	0.0434*	(0.0222)	0.0439*	(0.0226)
Expsq	- 0.0512	(0.0475)	- 0.0511	(0.0480)
Female	- 0.2441*	(0.1274)	- 0.2337*	(0.1391)
Dar es Salaam	0.2517**	(0.1054)	0.2467**	(0.1094)
Lambda	-	-	0.0874	(0.4489)
	R^2 =0.6373	F(5,52)=18.28	R^2 =0.6376	F(6,51)=14.95

[*] Standard errors in parentheses. Significance at 10 % level indicated by *, at 5% level by ** and at 1 % level by ***.

6. Empirical results

6.1 Main results

We have estimated three variants of the participation function based on specification A (A1, A2, and A3), and two variants based on specification B (B1 and B2). Estimation results are shown in Table 4. All the estimated relationships exhibit an acceptable fit (Log-likelihood, Chi square or McFadden's pseudo-R^2), and acceptable predictive properties (share of cases correctly predicted).

In each of the five estimated equations, the coefficients affiliated with five explanatory variables are significantly different from zero. The impact coefficients of Virtual income (in A1, A2 and A3) and Real virtual income (in B1 and B2) carry the predicted negative sign, and they are all different from zero at standard levels of statistical significance. We return below to a broader discussion of the relationship between income and participation in informal production. Next, notice that living in the country's largest city reduces the probability of participating in informal sector activities, as implied by the strongly negative and statistically significant coefficient of the variable Dar es Salaam. The estimated coefficients affiliated with variables Female and Mother are statistically significant, with a positive sign of the Female-coefficient, but a negative sign for the Mother-coefficient. The negative sign of the Mother-coefficient is in good accordance with the findings of inter alia Glick (1999). The positive sign of the Female-coefficient is more surprising, but must be interpreted in view of the fact that the share of women who hold a formal sector job is rather small. A somewhat more comprehensive discussion of this is found in Theisen (2005). The positive, and statistically significant, coefficient of the Age-variable implies that participation in informal production is an increasing function of Age. This is not in accordance with results from multiple-job-holding studies based on data from industrialised countries, and is therefore discussed in more detail below. Another surprising result is the positive, but statistically insignificant, Hours-coefficient, to which we now turn.

Table 4 Estimation results for structural-form logit equations*.

	Specification A			Specification B	
	A1	A2	A3	B1	B2
Constant	- 2.2494	- 1.4221	- 1.4199	- 1.2677	- 0.9014
	(2.9831)	(0.8117)	(0.8110)	(3.0162)	(0.8283)
Virtual income	- 0.0290***	- 0.0292***	- 0.0290***	-	-
	(0.0062)	(0.0108)	(0.0106)		
Virtual wage rate	- 0.0026	- 0.0032	-	-	-
	(0.0372)	(0.0372)			
Real virtual income	-	-	-	- 0.2117***	- 0.2125***
				(0.0710)	(0.0706)
Hours	0.0995	-	-	0.0405	-
	(0.3453)			(0.3478)	
Size	0.0675	0.0692	0.0692	0.0641	0.0649
	(0.0652)	(0.0650)	(0.0649)	(0.0650)	(0.0647)
Female	1.1540**	1.1614**	1.1711**	1.3787**	1.3831**
	(0.5591)	(0.5571)	(0.5455)	(0.5621)	(0.5602)
Mother	- 1.4726**	- 1.4919**	- 1.4948**	- 1.6059**	- 1.6147***
	(0.6168)	(0.6122)	(0.6110)	(0.6248)	(0.6199)
School-aged	0.1162	0.1114	0.1120	0.1143	0.1125
	(0.1298)	(0.1289)	(0.1287)	(0.1310)	(0.1301)
Age	0.0629**	0.0629**	0.0621***	0.0501**	0.0500**
	(0.0252)	(0.0252)	(0.0234)	(0.0233)	(0.0233)
Dar es Salaam	- 1.1987***	- 1.2211***	- 1.2256***	- 1.3196***	- 1.3294***
	(0.3470)	(0.3384)	(0.3343)	(0.3445)	(0.3342)
Public	0.3617	0.3500	0.3450	0.3381	0.3332
	(0.4215)	(0.4199)	(0.4159)	(0.4201)	(0.4182)
Parastatal	0.1055	0.0795	0.0685	0.0567	0.0455
	(0.4145)	(0.4049)	(0.3842)	(0.3974)	(0.3857)
Log-likelihood	- 137.0322	- 137.0739	- 137.0776	- 136.3118	- 136.3185
Log-likelihood (Slopes=0)	- 162.0055	- 162.0055	- 162.0055	- 162.0055	- 162.0055
Chi-Sq.	49.94661	49.86304	49.85567	51.38741	51.37389
Pseudo R^2	0.1541	0.1539	0.1539	0.1586	0.1586
Correct predictions	0.7166	0.7206	0.7166	0.7247	0.7247
Sample size	247	247	247	247	247

* Standard errors in parentheses. One asterix, *, indicates that the coefficient is statistically different from zero at the 10 % level, ** at the 5 % level, and *** at the 1% level.

6.2 Formal sector working time

From our theoretical and econometric model, the Hours-coefficient is expected to take the value minus 1. In assessing the positive, but statistically insignificant, coefficient of the Hours-variable in columns A1 and B1 of Table 4 it is, however, important to take into

account that there in our data is very little variation in formal sector working time. In other words, the Hours-variable comes close to being a second constant. The impact coefficient of such a variable is hard to identify. Consider therefore the results from re-estimating the participation equation with the Hours-variable excluded. The estimation results for this case are shown in columns A2 and B2 of Table 4. A pair-wise comparison of coefficients in A1 vs. A2, and B1 vs. B2, reveals that the impact coefficients of other variables change very little when the Hours-variable is left out, while the magnitude of the constant term is reduced substantially. In addition, the drop in fit when the Hours-variable is excluded is negligible. Taken together these observations have led us to focus on the participation equation where the Hours' variable is left out. It is important to notice, though, that the arguments for excluding the formal sector working time variable from the participation equation are due only to characteristics of the data, reflecting characteristics of the Tanzanian formal sector labour market. From a theoretical point of view, there is no doubt that the Hours'-variable belongs in the participation equation. In order to identify the impact of formal sector working hours on participation in informal production, however, a data set with much more variation in the Hours-variable is needed.

7. Discussion of the main results and of the preferred specification

7.1 Income and participation

As already noted, the coefficient affiliated with the income variable carries the expected negative sign, and it is different from zero at conventional levels of statistical significance. In other words, the estimation results for both specifications consistently imply that leisure is a normal good. In order to see that not only the sign, but also the magnitude of the income effects show a consistent pattern, let us calculate the probability of participation in each of the two specifications. When the calculations are carried out for a man working in the private sector outside Dar es Salaam, of average age, average family size, with all children above 7 years of age, and with average virtual income and average virtual wage rate, we find a participation probability of 0.673 in Specification A2, and 0.704 in Specification B2. Moreover, if a person with such characteristics doubles his income, the probability of

participation drops by 6.8 percentage points in Specification A2, and by 5.1 percentage points in Specification B2. These calculations show that when it comes to the magnitude of the effects of variations in income, the two specifications give roughly the same picture.

Consider next the results for Specification, B2. In order to get a better grasp of the relationship between participation in informal production and real virtual income, we have in Figure 2 plotted the participation rate as a function of real virtual income. From the graph it is seen that a man living outside Dar es Salaam (indicated by MEN in the figure) is predicted to have a participation probability of almost 0.8 if his (real virtual) income is very low, but somewhat less that 0.4 if his income is at the high end of the income distribution. The income variations covered by Figure 2 are, however, very large. Let us therefore also consider an increase in income from average real virtual income (approximately 1.5 in Figure 2) to twice the average. Such an increase leads to a drop in the probability of participation by only a few percentage points. Hence, we conclude that real virtual income has a only a moderate impact on participation.

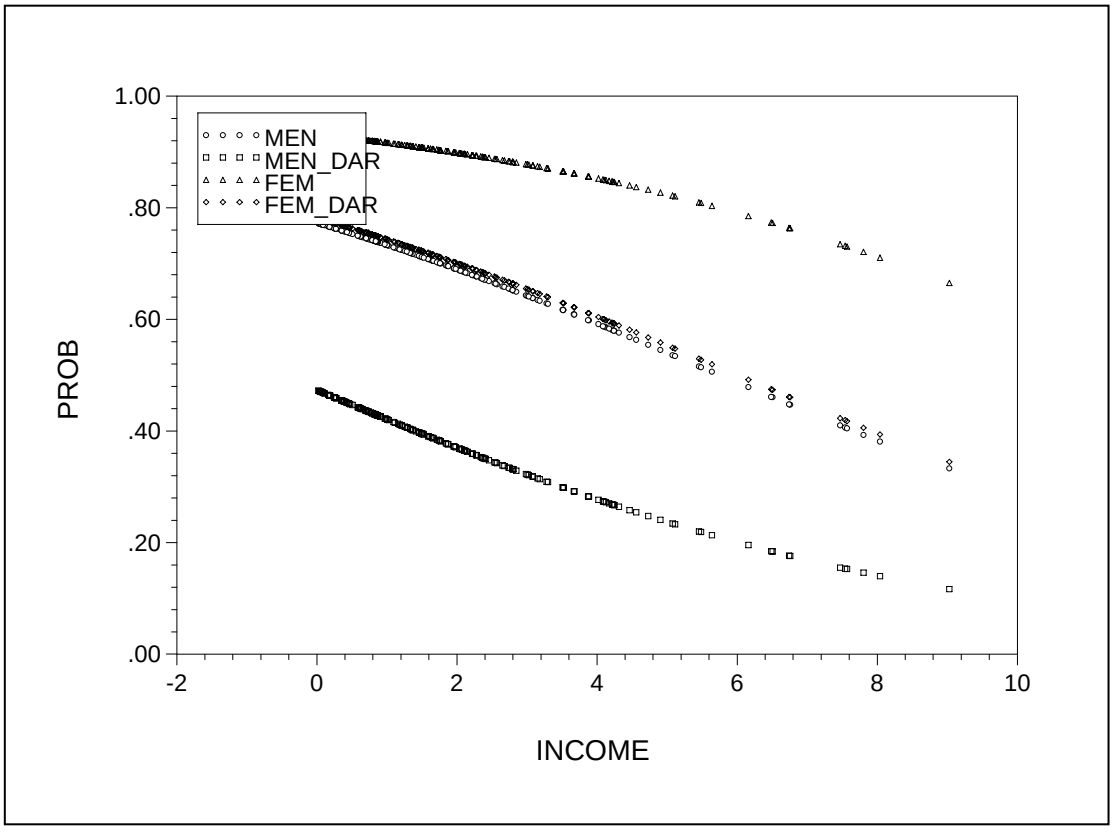

Figure 2 Probability of participation in the informal sector as a function of real virtual income

We have followed the normal practice and included a number of socio-economic variables as taste shifters in the participation function. This practice has little basis in neo-classical theory of consumer choice, and may even lead to inferior estimates of the parameters of primary interest. In particular, the correlation matrix provided in Appendix C indicates that there possibly might be a problem with the variables Female, Age, and Dar es Salaam, which are more strongly correlated with the economic variables than are the other taste shifters. In order to assess whether this represents a serious problem, the participation equations have been re-estimated with different taste shifters left out of the equation, one or two at a time. Table 5 shows how this affects the coefficients of the economic variables. The conclusion is clear: leaving out taste shifters has in most cases little impact on the estimated coefficients of variables Virtual income and Real virtual income. There are a few exceptions, however. Leaving out Dar es Salaam in Specification A2 gives an estimate of the Virtual income coefficient outside a 95 % confidence interval of the estimate in the basic equation. Similarly, leaving out variables Female or Age in Specification B2 results in an estimate of the Virtual income coefficient that lies outside the 95 % confidence interval of the estimate in the basic equation. On the other hand, the magnitude of the deviation between the estimated coefficient

Table 5 Estimated income and wage coefficients when different (groups of) of taste shifters are left out of the participation equation. Standard errors in parenthesis.

	Specification A2		Specification B2
	Virtual income	Virtual wage rate	Real virtual income
Basic equation	- 0.0292 (0.0108)	- 0.0032 (0.0372)	- 0.2125 (0.0706)
Size	- 0.0287 (0.0108)	- 0.0035 (0.0372)	- 0.2102 (0.0700)
Female	- 0.0283 (0.0106)	- 0.0214 (0.0359)	- 0.1756 (0.0663)
Mother	- 0.0310 (0.0108)	- 0.0097 (0.0367)	- 0.2062 (0.0679)
School-aged	- 0.0292 (0.0108)	- 0.0050 (0.0371)	- 0.2112 (0.0703)
Age	- 0.0269 (0.0107)	- 0.0031 (0.0338)	- 0.2376 (0.0714)
Dar es Salaam	- 0.0346 (0.0107)	- 0.0246 (0.0360)	- 0.2095 (0.0695)
Sector dummies	- 0.0285 (0.0105)	- 0.0030 (0.0352)	- 0.2100 (0.0695)

in the basic equation and the re-estimated equations is in all cases less than 20 % (when we don't pay attention to the statistically insignificant coefficients affiliated with the virtual wage rate). Hence, we conclude that the estimation results for the impact parameters of economic variables are quite robust towards the specification of taste shifters.

All results referred to so far are based on the assumption that the earnings function for the private formal sector can be used for predicting the virtual wage rate, which in turn is an essential element in the calculation of virtual income and real virtual income. From Table 3 we see that the virtual wage rate depends heavily on years of schooling. As mentioned in Section 3.2, however, some researchers have found that the virtual wage rate in some developing countries falls short of the wage rate in the formal private sector by 10 – 20 per cent. We have therefore carried out a sensitivity analysis in order to assess whether our results would be substantially changed if the virtual wage rate is less strongly dependent on schooling. Specifically, we have in the sensitivity analysis assumed that the schooling parameter in the virtual wage rate function is 25 per cent smaller that the number given in Table 3, while all other parameters are unchanged. In A2 this lead to a negligible change in the coefficient affiliated with virtual income (the coefficient changed from – 0.0292 to – 0.0294). In B2, however, the change in the virtual wage equation lead to a change in the coefficient affiliated with real virtual income from – 0.2125 to – 0.1777. For B2, the new estimate is statistically significantly different from the original estimate. Notice, however, that despite the substantial change in the coefficient of the real virtual income variable in specification B2, the relationship between real virtual income and the probability of participating in informal production was only moderately affected. Hence, we conclude that the inference drawn from our estimation results seem to be quite robust towards measurement errors in the virtual wage rate.

In order to further assess the robustness of our estimates, let us also compare the results for Specification A2 with results obtained by Theisen (2005) in estimating a reduced-form participation equation with a similar specification, and from the same data. The income parameter, $\omega_{A\hat{V}}$, in the structural form model is identified also in the reduced form model, cf. Theisen (2005). In the reduced form estimation, a probit model was used, and the income

coefficient was estimated to - 0.0168, with a standard deviation of 0.0062. However, as explained by Maddala (1983, p 23) the probit coefficient has, due to differences in the variance of the logistic and normal distributions, to be multiplied by $\pi/\sqrt{3}$ in order to be made comparable to the logit coefficient. Multiplying the estimate of - 0.0168 by $\pi/\sqrt{3}$ gives a coefficient of – 0.030, which is practically identical to our estimate of – 0.029 in Specification A2 of Table 4. This provides additional evidence that our estimate of the income coefficient is quite robust.

7.2 Age, experience and participation

The coefficient affiliated with the Age-variable carries a positive sign, and is statistically significantly different from zero, implying that individuals are more likely to participate in informal production the higher their age. This is in accordance with the results obtained by Pradhan and van Soest (1995). It is also in accordance with interpreting age as a measure of human capital acquired through experience. In plain words, it may take some time and experience to develop the skills required to produce goods at home, or to establish oneself as a part-time self-employed or informal sector worker. This explanation does not, however, correspond well with the weak correlation between education and participation in informal production in our sample.

An alternative, and more likely, explanation for the positive relationship between age and informal sector participation may be that Tanzanians have strong incentives to engage in informal production as they approach retirement, because pensions are far from sufficient to cover the costs of living after retirement. This interpretation is also supported by the observation that 43 % of the individuals in the sample envisaged income from informal production as an important source of living after retirement, while only 31 % envisaged pension incomes among the important sources of income after retirement. This in turn implies that there may be a relationship between the level of retirement benefits and the size of the informal sector. Indeed, the fact that the increase in retirement benefits for a long period of time have fallen short of inflation may provide part of the explanation for the expansion of the informal sector in Tanzania.

7.3 The choice between Specification A and Specification B

In Specification A, Virtual income and the Virtual wage rate enter as separate variables. In Specification B, by contrast, only the ratio between these two variables enters. This is the only difference between the two specifications. In order to choose between them, consider first Eqs. A2 and A3 in Table 4. A comparison of these equations reveals that very little is lost in fit etc. by leaving out the Virtual-wage-rate variable, for which the coefficient in Eq. A2 is statistically insignificant. Next, let us compare Eqs. A3 and B2. These equations contain the same number of independent variables, but Eq. B2 performs slightly better than Eq. A3 both when it comes to fit, and when it comes to the share of correct predictions. Taken together, these observations indicate that dividing the virtual income variable with the virtual wage rate, as in Eq. B2, gives a "refined" variable – real virtual income – which explains participation slightly better than the "unrefined" virtual income variable in A3. This also suggests that our predictions of the virtual wage rate contain relevant, albeit certainly imprecise, information on the virtual wage rate variable. In conclusion, this leads us to prefer Specification B (Equation B2 in Table 4) over Specification A. This conclusion is strengthened by the fact that Specification B is more parsimonious than Specification A. On the other hand, the sensitivity analysis carried out in Section 7.1 revealed that Specification B is more vulnerable to measurement errors in the virtual wage rate. Consequently, when there is reason to believe that the virtual wage rate is distorted, there may be a case for preferring Specification A. As demonstrated in Section 7.1, however, it is only when the problems of measurement are severe that the estimation results are strongly affected.

8. Conclusions

We have estimated structural-form functions explaining multiple-job-holding in the form of participating in informal production, in addition to holding a job in the formal sector. The estimated participation functions complement the work of Gindling (1991), Pradhan and van Soest (1998) and Funkhauser (1996). While these papers focus on individuals' choice of primary sector of work, the focus in the present paper has been on multiple-job-holding in the form of informal sector participation, conditional on holding a job in the formal sector. The

present paper also complements the reduced form multiple-job-holding studies of Glick (1999), Joliffe (2004), and Theisen (2005). To our knowledge, structural form participation functions of multiple-job-holding in developing countries have not previously been estimated.

Strong assumptions were needed to come around measurement problems, so that structural form participation functions could be estimated. The way these measurement problems were solved in the present paper seems to work reasonably well, and may be of interest in itself. In addition, we have indicated how the methods could be improved through joint estimation of participation functions and informal sector production functions. This would require a new and more informative data set, though.

The empirical results imply that leisure is a normal good. Consequently, formal-sector-workers who experience declining real wages in their main job can, at least to some extent, be expected to compensate by increasing their participation in informal production. This is in accordance with the view of Fields (1975) that the informal sector plays the role of a buffer sector. To some extent this may reduce the need for Government interventions in order to alleviate poverty problems in periods of economic recession.

The estimation results show that living in Dar es Salaam and being the mother of a small child affect participation negatively, while female gender and high age have an opposite effect. Comparing these results with those obtained by for instance Shisko and Rostker (1977) and Conway and Kimmel (1998) provide some interesting insights: First, income, gender and presence of small children have a negative effect on labour supplied to a second job both in Tanzania and in the US. On the other hand, while age affects Tanzanian formal-sector workers participation in informal production positively, Shisko and Rostker (1977) and Conway and Kimmel (1998) found a strong negative effect of age on the probability of US workers holding a second job. We argued that these differences most likely can be attributed to differences in pension systems. Specifically, in the case of Tanzania participation in informal production may function as a “substitute” for a pension system. In developed countries like the US, by contrast, the negative impact of age on participation suggests that workers when they approach retirement first withdraw from their second job, then from their

main job. These differences in the transition from a full-time formal sector job to retirement should be paid more attention to in future research.

References

Biddle, J. (1988): "Intertemporal Substitution and Constraints on Labour Supply," The Review of Economics and Statistics, 70, 347-351.

Conway, K. S. and J. Kimmel (1998): "Male labour supply estimates and the decision to moonlight," Labour Economics, 5, 135-166.

Deaton, A. (1997): The analysis of household surveys. A Microeconometric Approach to Development Policy, The Johns Hopkins University Press, Baltimore, Maryland, USA.

Deardorff, A. V. and F. P. Stafford (1976): "Compensating and Cooperating Factors," Econometrica, 44, 671-684.

Feige, E. L. (1990): "Defining and estimating underground and informal economies: The new institutional economics approach," World Development, 18, 989-1002

Fields, G. S. (1975): "Rural urban migration, urban unemployment and underemployment, and job-search activity in LDCs," Journal of Development Economics, 2, 165-187.

Funkhauser, E. (1996): “The Urban Informal Sector in Central America: Household Survey Evidence,” World Development, Vol. 24, 1737-1751.

Gindling, T. H. (1991): "Labour market segmentation and the determination of wages in the public, private-formal, and informal sectors in San Jose, Costa Rica," Economic Development and Cultural Change, 39, 585-606.

Glick, P. (1999): “Simultaneous Determination of Home Work and Market Work of Women in Urban West Africa,” Oxford Bulletin of Economics and Statistics, Vol. 61, 57-84.

Green, W. H. (2000): Econometric Analysis, Fourth Edition, Prentice Hall, Upper Saddle River, N. J.

Gronau, R. (1977): "Leisure, Home Production, and Work - the Theory of the Allocation of time Revisited," Journal of Political Economy, 85, 1099 - 1123.

Jolliffe, D. (2004): “The Impact of Education in Rural Ghana: Examining Household Labor Allocation and Returns On and Off the Farm,” Journal of Development Economics, Vol. 73, 287-314.

Killingsworth, M. R. (1983): Labour Supply, Cambridge University Press, Cambridge.

Knight, J. B. and R. H. Sabot (1990): Education, Productivity, and Inequality: The East

African Natural Experiment, Oxford University Press, Oxford.

Lassibille, G. and J-P. Tan (2004): “The Returns to Education in Rwanda,” Journal of African Economies, Vol. 14, 92-116.

Lee, L-F. (1983): “Generalized Economic Models with Selectivity,” Econometrica, Vol. 51, 507-512.

Leibbrand, M., C. Woolard and I Woolard (2000): “The Contribution of Income Components to Income Inequality in the Rural Former Homelands of South Africa: a Decomposable Gini Analysis,” Journal of African Economics, Vol. 9, 79-99.

Lindauer, D. L. and R. H. Sabot: “The public/private wage differential in a poor urban economy,” Journal of Development Economics, 12, 137-152.

Maddala, G. S. (1983): Limited-dependent and qualitative variables in econometrics, Cambridge University Press, Cambridge, New York.

Mincer, J. (1974): Schooling, experience and earnings, Columbia University Press, NY.

Moffitt, R. (1982): "The Tobit Model, Hours of Work and Institutional Constraints," Review of Economics and Statistics, 64, 510-515.

Moffitt, R. (1984): ”The Estimation of a Joint Wage-Hours Labor Supply Model,” Journal of Labour Economics, Vol. 2, 550-566.

Oswald, A. and I. Walker (1993): "Labour Supply, Contract Theory and Unions," IFS working paper No. W93/21, London

Pencavel, J. (1986) ”Labour Supply of Men: A Survey,” in Ashenfelter, O. and D. Layard, eds., Handbook of Labor Economics, Vol. 1, North-Holland, Amsterdam, 3-102.

Pradhan, M. and A. van Soest (1995): "Formal and informal sector employment in urban areas of Bolivia," Labour Economics, 2, 275-297.

Schneider, F. and D. H. Enste (2002): The Shadow Economy: An International Survey, Cambridge University Press, Cambridge, UK.

Shishko, R. and B. Rostker (1976): "The Economics of Multiple Job Holding," American Economic Review, 66, 298-308.

Theisen, T. (2005): “Tanzanian Formal Sector Workers’ Participation in Informal Production,” Applied Economics, 37, 2469-2485.

Trost, R. P. and L-F. Lee (1984): “Technical Training and Earnings: a Polychotomous Choice Model with Selectivity,” The Review of Economics and Statistics, Vol. 66, 151-156.

Appendix A

1. The participation condition based on specification A

Following Pencavel (1986, p 53) we specify the utility function as Eq. (4a) in the main text. The correspondence between the parameters and variables in Eq. (4a) in the main text and those used by Pencavel is as follows: $\omega_{A0} = \alpha_0$, $\omega_{A\hat{W}} = \alpha_1$, $\omega_{A\hat{V}} = \alpha_2$, $\omega'_{AR} = \alpha_3$, $\left(H + \overline{N}\right) = h$, and $\varepsilon_A = \varepsilon$. Writing expQ for the exponential expression in Eq. (4a) in the main text, and using that $\left(H + \overline{N}\right) = T - L$ we obtain the partial derivates w.r.t. X and L:

$$U_X = \left(\frac{\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}}}{\left(\omega_{A\hat{V}}\right)^2} \right) [\exp Q] \left\{ \frac{\left(\omega_{A\hat{V}}\right)^2}{\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}}} \right\} = \exp Q \qquad \text{(A1)}$$

$$U_L = \frac{\left(-\omega_{A\hat{V}}\right)\left(\omega_{A\hat{V}}\right)^2}{\left(\omega_{A\hat{V}}\right)^4} \exp Q +$$

$$\frac{\left(\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}}\right)}{\left(\omega_{A\hat{V}}\right)^2} [\exp Q] \frac{\left(-\left(\omega_{A\hat{V}}\left(\omega_{A0} + \omega_{A\hat{V}}X + \omega'_{AR}R + \varepsilon_A\right) - \omega_{A\hat{W}}\right)\left(-\omega_{A\hat{V}}\right)\right)}{\left(\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}}\right)^2}$$

$$U_L = -\frac{1}{\omega_{A\hat{V}}} \exp Q \left[1 - \frac{\left(\omega_{A\hat{V}}\left(\omega_{A0} + \omega_{A\hat{V}}X + \omega'_{AR}R + \varepsilon_A\right) - \omega_{A\hat{W}}\right)}{\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}}} \right]$$

$$U_L = -\frac{1}{\omega_{A\hat{V}}} \exp Q \left\{ \frac{\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}} - \omega_{A\hat{V}}\left(\omega_{A0} + \omega_{A\hat{V}}X + \omega'_{AR}R + \varepsilon_A\right) + \omega_{A\hat{V}}}{\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}}} \right\}$$

$$U_L = -\exp Q \left\{ \frac{\left(H + \overline{N}\right) - \left(\omega_{A0} + \omega_{A\hat{V}}X + \omega'_{AR}R + \varepsilon_A\right)}{\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}}} \right\} \qquad \text{(A2)}$$

From Eqs. (A1) and (A2) we obtain the marginal rate of substitution:

$$\frac{U_L}{U_X} = -\left\{ \frac{\left(H + \overline{N}\right) - \left(\omega_{A0} + \omega_{A\hat{V}}X + \omega'_{AR}R + \varepsilon_A\right)}{\omega_{A\hat{V}}\left(H + \overline{N}\right) - \omega_{A\hat{W}}} \right\} \qquad \text{(A3)}$$

Substituting from Eq. (A3) into Condition (1) in the main text, and using that $\hat{W} = ZF_H(H)$ we obtain:

$$-\left\{\frac{(H+\overline{N})-(\omega_{A0}+\omega_{A\hat{V}}X+\omega'_{AR}R+\varepsilon_A)}{\omega_{A\hat{V}}(H+\overline{N})-\omega_{A\hat{W}}}\right\}\begin{cases}\geq \hat{W} & \Rightarrow \text{ no participation } (H=0)\\ < \hat{W} & \Rightarrow \text{ participation } (H>0)\end{cases} . \qquad \text{(A4)}$$

Since $X = V + W\overline{N} + ZH = V + W\overline{N} + \overline{W}H + \hat{W}H - \hat{W}H + \hat{W}\overline{N} - \hat{W}\overline{N}$, which gives $X = V + (\overline{W} - \hat{W})H + (W - \hat{W})\overline{N} + \hat{W}H + \hat{W}\overline{N} = \hat{V} + \hat{W}H + \hat{W}\overline{N}$, substitution into Condition (A4) yields:

$$-\frac{(H+\overline{N})-(\omega_{A0}+\omega_{A\hat{V}}(\hat{V}+\hat{W}H+\hat{W}\overline{N}+)+\omega'_{AR}R+\varepsilon_A)}{\omega_{A\hat{V}}(H+\overline{N})-\omega_{A\hat{W}}}\begin{cases}\geq \hat{W} & \Rightarrow \text{ no participation } (H=0)\\ < \hat{W} & \Rightarrow \text{ participation } (H>0)\end{cases}$$

$$\frac{\omega_{A0}+\omega_{A\hat{V}}\hat{V}+\omega_{A\hat{V}}\hat{W}H+\omega_{A\hat{V}}\hat{W}\overline{N}++\omega'_{AR}R+\varepsilon_A-H-\overline{N}}{\omega_{A\hat{V}}(H+\overline{N})-\omega_{A\hat{W}}}\begin{cases}\geq \hat{W} & \Rightarrow \text{ no participation } (H=0)\\ < \hat{W} & \Rightarrow \text{ participation } (H>0)\end{cases}$$

Since $\left[\omega_{A\hat{V}}(H+\overline{N})-\omega_{A\hat{W}}\right]$ in the denominator of this condition represents the substitution effect of an increase in the virtual wage rate, which is negative, the condition can be written:

$$\omega_{A0}+\omega_{A\hat{V}}\hat{V}+\omega_{A\hat{V}}\hat{W}H+\omega_{A\hat{V}}\hat{W}\overline{N}++\omega'_{AR}R+\varepsilon_A-H-\overline{N}\begin{cases}\leq \hat{W}(\omega_{A\hat{V}}(H+\overline{N})-\omega_{A\hat{W}}) & \Rightarrow \\ & \text{no participation } (H=0)\\ > \hat{W}(\omega_{A\hat{V}}(H+\overline{N})-\omega_{A\hat{W}}) & \Rightarrow \\ & \text{participation } (H>0)\end{cases}$$

$$\omega_{A0}+\omega_{A\hat{V}}\hat{V}+\omega_{A\hat{W}}\hat{W}+\omega'_{AR}R-\overline{N}+\varepsilon_A\begin{cases}\leq 0 & \Rightarrow \text{ no participation } (H=0)\\ > H & \Rightarrow \text{ participation } (H>0)\end{cases},$$

which since we always have $H \geq 0$ can be written:

$$\omega_{A0} + \omega_{A\hat{V}}\hat{V} + \omega_{A\hat{W}}\hat{W} + \omega'_{AR}R - \overline{N} + \varepsilon_A \begin{cases} \leq 0 \quad \Rightarrow \quad \text{no participation (H = 0)} \\ > 0 \quad \Rightarrow \quad \text{participation (H > 0)} \end{cases} . \tag{A5}$$

2. The participation function based on specification B

The starting point is the Cobb Douglas utility function $U(X, L; R, \varepsilon_B) = X^{\alpha}L^{\beta}$, where the specification of X and L is shown in Eq. (4b) in the main text. From this we obtain $U_X = \alpha X^{\alpha-1}L^{\beta} = \alpha X^{\alpha}L^{\beta}/X$ and $U_L = \beta X^{\alpha}L^{\beta-1} = \beta X^{\alpha}L^{\beta}/L$, which in turn yield the marginal rate of substitution:

$$\frac{U_L}{U_X} = \frac{\beta X^{\alpha}L^{\beta}/L}{\alpha X^{\alpha}L^{\beta}/X} = \frac{\beta}{\alpha}\frac{X}{L} = \frac{\beta}{\alpha}\frac{V + W\overline{N} + Z(H - \varepsilon_B - \mu - \omega'_{BR}R)}{T - \overline{N} - (H - \varepsilon_B - \mu - \omega'_{BR}R)}$$

$$\frac{U_L}{U_X} = \frac{\beta X^{\alpha}L^{\beta}/L}{\alpha X^{\alpha}L^{\beta}/X} = \frac{\beta}{\alpha}\frac{X}{L} = \frac{\beta}{\alpha}\frac{V + W\overline{N} + ZH + (-\varepsilon_B - \mu - \omega'_{BR}R)}{T - \overline{N} - H - (-\varepsilon_B - \mu - \omega'_{BR}R)} \tag{A6}$$

As before, using that $Z = \hat{W} = \overline{W}$, and that $X = V + W\overline{N} + ZH = \hat{V} + \hat{W}H + \hat{W}\overline{N}$, substitution into Eq. (A6) yields:

$$\frac{U_L}{U_X} = \frac{\beta}{\alpha}\frac{\hat{V} + \hat{W}\overline{N} + \hat{W}H + \hat{W}(-\varepsilon_B - \mu - \omega'_{BR}R)}{T - \overline{N} - H - (-\varepsilon_B - \mu - \omega'_{BR}R)} \tag{A7}$$

Substituting (A7) and $\hat{W} = Z$ into Eq. (1) in the main text now yields:

$$\frac{\beta}{\alpha}\frac{\hat{V} + \hat{W}\overline{N} + \hat{W}H + \hat{W}(-\varepsilon_B - \mu - \omega'_{BR}R)}{T - \overline{N} - H - (-\varepsilon_B - \mu - \omega'_{BR}R)} \begin{cases} \geq \hat{W} \Rightarrow \text{no participation (H = 0)} \\ < \hat{W} \Rightarrow \text{participation (H > 0)} \end{cases}$$

Multiplying on both sides by $(T - \overline{N} - H - (-\varepsilon_B - \mu - \omega'_{BR}R))$ and dividing through with $\hat{W}$ yields:

$$\frac{\beta}{\alpha}\frac{\hat{V}}{\hat{W}}+\frac{\beta}{\alpha}\overline{N}+\frac{\beta}{\alpha}H+\frac{\beta}{\alpha}\left(-\varepsilon_B-\mu-\omega'_{BR}R\right)\begin{cases}\geq T-\overline{N}-H-\left(-\varepsilon_B-\mu-\omega'_{BR}R\right)\Rightarrow \\ \qquad \text{no participation } (H=0) \\ <T-\overline{N}-H-\left(-\varepsilon_B-\mu-\omega'_{BR}R\right)\Rightarrow \\ \qquad \text{participation } (H>0)\end{cases}$$

$$\frac{\beta}{\alpha}\frac{\hat{V}}{\hat{W}}+\left(\frac{\beta}{\alpha}+1\right)\overline{N}+\left(\frac{\beta}{\alpha}+1\right)H+\left(\frac{\beta}{\alpha}+1\right)\left(-\varepsilon_B-\mu-\omega'_{BR}R\right)-T\begin{cases}\geq 0\Rightarrow & \text{no participation } (H=0)\\ <0\Rightarrow & \text{participation } (H>0)\end{cases}$$

Multiplying on both sides by $(\alpha/\alpha+\beta)$ yields:

$$\frac{\alpha}{\alpha+\beta}\frac{\beta}{\alpha}\frac{\hat{V}}{\hat{W}}+\overline{N}+H+\left(-\varepsilon_B-\mu-\omega'_{BR}R\right)-\frac{\alpha}{\alpha+\beta}T\begin{cases}\geq 0\Rightarrow & \text{no participation } (H=0)\\ <0\Rightarrow & \text{participation } (H>0)\end{cases}$$

Multiplying on both sides by – 1 and rearranging terms yields:

$$\underbrace{\left(\mu+\frac{\alpha}{\alpha+\beta}T\right)}_{\omega_{B0}}+\underbrace{\left(\frac{\beta}{\alpha+\beta}\right)}_{\omega_{B\hat{V}}}\frac{\hat{V}}{\hat{W}}+\omega'_{BR}R-\overline{N}+\varepsilon_B\begin{cases}\leq H\Rightarrow & \text{no participation } (H=0)\\ >H\Rightarrow & \text{participation } (H>0)\end{cases},$$

which, since H = 0 in the case of non-participation and H > 0 in the case of participation, can be written:

$$\omega_{B0}+\omega_{B\hat{V}}\frac{\hat{V}}{\hat{W}}+\omega'_{BR}R-\overline{N}+\varepsilon_B\begin{cases}\leq 0\Rightarrow & \text{no participation } (H=0)\\ >0\Rightarrow & \text{participation } (H>0)\end{cases}. \qquad \text{(A8)}$$

Appendix B

The covariates used for estimating the multinomial logit model for the allocation of workers to sectors of formal employment are show in Table B1, along with descriptive statistics. Notice that workers in the private sector have less education and are slightly younger than in the other two sectors, and that there are large differences in the share of workers living in Dar Es Salaam.

Table B1 Sample means by sector for correlates in the multinomial logit model for sector allocation of workers. Sample standard deviations of continous variables in parentheses.

Covariates	Sub-sectors			All sectors
	Private sector N = 58	Parastatal sector N = 103	Public sector N = 86	N = 247
Secondary*	0.1724	0.2427	0.3372	0.2591
University**	0.1034	0.3786	0.2326	0.2632
Age	33.8103 (8.6277)	35.5631 (7.7582)	35.5116 (8.0684)	35.1336 (8.0770)
Female	0.2241	0.2621	0.3140	0.2713
Small children	0.6552	0.6796	0.6512	0.6640
Married***	0.7241	0.6990	0.7442	0.7206
Dar Es Salaam	0.3276	0.4563	0.1163	0.3077

* Secondary is a dummy variable equal to 1 if the respondent's highest level of education is lower or upper secondary school, 0 if not.

** University is a dummy variable equal to 1 if the respondent's highest level of education is a university degree, 0 if not.

*** Married is a dummy variable equal to 1 if the respondent is married, 0 if not.

The dependent variable in the multinomial logit model is an integer variable equal to 0 if the respondent is employed in the privat sector, equal to 1 if employed in the parastatal sector,

and equal to 2 if employed in the public sector. Estimation results are shown in Table B2. A likelihood ratio test refutes the null hypothesis that the covariates have no impact on the allocation of workers to formal sectors of employment. Notice in particular that the coefficients affiliated with the educational variables are positive and statistically significant at conventional levels of significance, indicating a strong tendency for educated individuals to be allocated to the parastatal or the public sector rather than to the private sector.

Table B2 Estimation results for multinomial logit model*

	Parastatal sector		Public sector	
	Coefficient	Standard dev.	Coefficient	Standard dev.
Constant	- 1.3104		- 0.9639	
Secondary	0.9152**	0.4390	1.1216**	0.4425
University	2.0313***	0.5030	1.2649**	0.5532
Age	0.0258	0.0237	0.0299	0.0256
Female	0.5386	0.4303	0.7922*	0.4289
Small children	0.1831	0.4359	- 0.2234	0.4049
Married	- 0.2640	0.5040	0.0542	0.4696
Dar Es Salaam	0.6840*	0.3598	- 1.3290***	0.4766
Log-likelihood	- 234.6481			
Log-likelihood (Slopes = 0)	- 264.8623			
Chi.Sq.	60.4284			
Pseudo R^2	0.1586			
Percentage of correct predictions	53.0364			

* Standard errors in parentheses. Significance at 10 % level indicated by *, at 5% level by ** and at 1 % level by ***.

Appendix C. Correlation matrix for variables in the participation equations

	Dep. Variable	Virtual income	Virtual wage	Real virt income	Hours	Size	Female	Mother	School-aged	Age	Dar es Salaam	Public	Para-statal
Dependent var.	1.0000												
Virtual income	- 0.1977	1.0000											
Virtual wage	0.0732	-0.2165	1.0000										
Real virt. Income	- 0.2358	0.7619	-0.4250	1.0000									
Hours	0.1107	-0.1094	-0.1642	-0.1035	1.0000								
Size	0.1230	0.0292	0.0460	-0.0087	-0.0631	1.0000							
Female	- 0.0679	0.1686	-0.3719	0.3307	-0.0160	0.0361	1.0000						
Mother	- 0.1370	0.1562	-0.2285	0.2494	-0.1057	0.0338	0.7312	1.0000					
Schoolaged	0.1043	0.0131	0.1536	-0.0649	-0.1840	0.1892	0.0088	0.0852	1.0000				
Age	0.2182	-0.0446	0.4521	-0.2358	-0.0926	0.2276	-0.3434	-0.2139	0.3753	1.0000			
Dar es Salaam	- 0.2790	0.0722	0.1251	0.0213	-0.2827	-0.0399	0.0471	0.0091	0.0689	-0.0143	1.0000		
Public	0.1296	-0.0317	-0.0894	0.0131	0.1277	0.0769	0.0702	0.1079	-0.0273	0.0343	-0.3031	1.0000	
Parastatal	- 0.0933	0.1070	0.2809	0.0420	-0.2882	-0.0563	-0.0174	-0.0463	0.0932	0.0451	0.2723	-0.6181	1.0000