

CSAE Working Paper WPS/2016-34

Job Flexibility and Occupational Selection: An Application of Maximum Simulated Likelihood Using Data from Ghana*

Jonathan Lain[†]

November 6, 2016

Abstract

In many African labour markets, the vast majority of self-employed workers are female. It is often hypothesised that this is because self-employment enables workers to balance income-generation with caring for children and other domestic tasks and, since responsibility for these activities is divided unequally in the household, this effect is stronger for women than men. However, testing whether ‘job flexibility’ matters is difficult because variables that proxy for domestic obligations — such as the number of dependents in the household — may be endogenous to occupational choice. In this paper, we build a new estimator using maximum simulated likelihood, which allows us to couch the idea that selection on observables can be used as a guide to selection on unobservables within the multinomial choice problem individuals face when they select their occupations. We test this approach using detailed cross-sectional data from Ghana. Our results show that having extra dependents in the household pushes women towards low-input self-employment substantially more than men.

Keywords: Maximum Simulated Likelihood, Unobservable Selection, Occupational Choice, Self-Employment.

JEL classification: C15, J24, J46.

*I would like to thank my supervisor, Margaret Stevens, for her generous support and guidance on this paper. I am also grateful to Ian Crawford, Beata Javorcik, Simon Quinn, Neil Rankin, Francis Teal, and the members of the Firms and Development Research Group for their important inputs. This research was funded by the Economic and Social Research Council, and the data were collected by the Ghana Statistical Service and the Centre for the Study of African Economies. All errors are my own.

[†]Correspondence: The World Bank Office, Indonesia Stock Exchange Building Tower 2, 12th Floor, Jl. Jend. Sudirman Kav. 52–53, Jakarta 12190, Indonesia. E-mail: jlain@worldbank.org

1 Introduction

The non-pecuniary benefits of certain jobs — such as job security, working conditions, and ‘warm glow’ — may influence occupational selection, alongside the desire to maximise earnings. It is often hypothesised that self-employment enables workers to balance a career with domestic work, including childcare and other household chores, because these types of jobs are more ‘flexible’. These effects may be stronger for women than men if household-level domestic obligations are divided unequally due to social norms, individual preferences, or other factors. The main aim of this paper is to examine whether this logic holds using nationally-representative cross-sectional data from Ghana. Understanding the factors that drive self-employment is especially important in developing countries, where informal jobs are prevalent and where the self-employed typically generate low earnings.

We begin by outlining what it means for a job to be ‘flexible’, and consider what this implies for the relationship between individuals’ domestic obligations — the amount of domestic work they must provide to the household — and occupational choice. The concept of ‘flexibility’ is often invoked in the empirical literature on occupational selection, but is rarely defined explicitly. We illustrate three possible mechanisms, backed up by a simple model of time allocation. Firstly, some jobs may allow for ‘multi-tasking’, where income-generating work and domestic work are undertaken concurrently. Secondly, certain formal sector jobs may have minimum working hours. Thirdly, individuals may face costs for adjusting their hours of work, which depend on their chosen occupation. Such adjustment costs may be important when domestic obligations increase suddenly, for example, when a family member becomes ill. Across all three stories, it emerges that workers with greater domestic obligations select into more flexible occupations, even if these jobs yield lower earnings.

To test whether this prediction is supported by our data, we first examine which types of occupations may be regarded as flexible, using information on work location, hours worked, and time-use. We look not only at broad differences between wage- and self-employment, but also consider whether intra-sector heterogeneity, in terms of the technology deployed by self-employed workers, may be important. Specifically, we disaggregate the self-employed into ‘own account’ workers — self-employed individuals who work alone — and ‘employers’ — those who employ others in their business. We also refer to these jobs as ‘low-input’ and ‘high-input’ self-employment respectively throughout this paper. Our data suggest that low-input self-employment jobs are more flexible than both wage jobs and high-input self-employment.

Examining the job flexibility story also relies on finding a suitable proxy for individuals’ domestic obligations. To do this, we use the household dependency ratio: the ratio of children/elderly individuals

to working-age individuals in the household. We can then assess the importance of job flexibility by testing whether having additional dependents in the household increases the likelihood of working in low-input self-employment. Vitally, we disaggregate the results by sex, because domestic obligations for the household as a whole may be split unevenly between female and male household members.

Occupational selection may be endogenous to household composition, if there are factors that drive decisions about both family structure and employment that cannot be easily observed. For example, certain types of households engaged in low-return activities may choose to have more children to boost their potential earnings power. This is an especially difficult issue to deal with, because occupational selection is a multinomial choice problem, for which workers' decisions cannot be easily nested to form a set of binary choices, nor ranked to create an ordered choice problem. Moreover, it is difficult to find suitable instrumental variables or natural exogenous variation in household structure.

To solve this issue, we construct a new estimator, which builds the logic of Altonji et al. (2005) into a multinomial probit using maximum simulated likelihood. The main idea is to use selection on *observables* as a guide to selection on *unobservables*. Although this method has been widely used on continuous and binary outcome variables, we believe this is the first attempt to apply it to a multinomial choice problem. This is the main contribution of the paper.

Our results suggest that having extra dependents in the household drives women into low-input self-employment more than men. A one standard deviation increase in the dependency ratio implies women are 3.4 percent more likely to enter low-input self-employment, whereas the same change in household structure means men are just 0.8 percent more likely. Moreover, these effects on women's occupational choice survive even if we make the alternative assumption that selection on unobservables is *as strong as* selection on observables. This is not the case for men. These results are even more robust to concerns about endogeneity in the unmarried sub-sample. This fits with our priors that endogeneity may be less of problem for these individuals, as they have less influence over household structure.

This paper proceeds as follows. In Section 2 we review the related literature. In Section 3, we outline three simple ways to relate occupational choice to job flexibility and domestic obligations. In Section 4 we describe our data. In Section 5 we outline our econometric approach, explaining how we allow for the endogeneity of household structure to occupational choice. In Section 6 we report our main results. In Section 7 we examine the robustness and heterogeneity of these results. In Section 8 we conclude.

2 Related Literature

In order to examine the importance of non-wage job attributes for patterns of labour market participation, economists often seek to test whether household composition influences how individuals choose their jobs. However, family structure may be endogenous to occupational selection, making it difficult to estimate the causal effect of household composition. For example, Becker (1985) argues that occupational choice affects family stability and hence the likelihood of having children, insofar as higher earnings make marriage, as a source of income, relatively less desirable. Also, Rosenzweig and Wolpin (1980) suggest that children (or indeed other family members) may be added to increase the earnings of the household at large, especially if individuals work in jobs with low returns. Additionally, occupation may be correlated with a wide range of variables related to family planning, such as knowledge and understanding of contraception, which may not be easily observed and measured.

Natural experiments may provide one avenue for addressing these endogeneity concerns. Angrist and Evans (1998), for example, use an instrumental variable approach to assess the impact of family size on individuals' labour supply, exploiting exogenous variation in child sex and parental preferences for a mixed sibling-sex composition. In similar vein, the birth of twins may also be treated as a shock to fertility, which generates an exogenous change in household demographics (Rosenzweig and Wolpin, 1980; Bronars and Grogger, 1994). However, although these techniques have been widely applied to the labour supply decision, their use in the literature on occupational selection — with multiple employment sectors — is limited.

Others have used panel techniques to try and examine the causal effect of changes in household structure on selection into self-employment (Evans and Leighton, 1989; Fajnzylber et al., 2006). For example, using data from the United States, Wellington (2006) examines the impact of family size on women's likelihood of being self-employed. Estimating both a cross-section model, and a 'longitudinal' model, which controls for time-invariant individual heterogeneity, she finds that the impact of having larger families is somewhat mixed and depends, in particular, on women's level of education.

Structural models may also help disentangle the relationship between household structure and occupational choice. For example, Lombard (2001) builds a structural model, which disaggregates the drivers of occupational choice into differences in earnings and differences in other job attributes between wage- and self-employment. Using data on married women from the United States, she shows that the chances of participating in self-employment rise not only with potential earnings in that sector, but also with demand for a non-standard work week and demand for the flexibility to vary one's work schedule. The latter two attributes are also found to be strongly linked to the demographic composition of the

household, especially the number of children present.

We hope to complement this literature by adopting a new approach to assessing the endogeneity of household composition to occupational choice, which relies on using selection on observables as a guide to selection on unobservables.

Given the context in which we apply this approach, this paper also links to an emerging Ghanaian literature that provides substantial descriptive evidence on the links between fertility, education, and female labour force participation. Taken together, these studies suggest that education has positive effects on women’s involvement in the labour market, for both wage-employment and self-employment (Sackey, 2005). However, the effects of fertility and, by extension, household structure, appear to be far more mixed. Whilst having extra children increases the chances of women participating in income-generating activities in general, it emerges that similar effects prevail for men (Ackah et al., 2009; Baah-Boateng et al., 2013). Moreover, fertility only appears to increase the likelihood of engaging in non-farm self-employment when it is undertaken as a secondary activity alongside other work in agriculture or wage-employment (Heintz and Pickbourn, 2012; Ackah, 2013).

In part, we believe these equivocal results stem from the fact that much of this literature tries to reduce occupational selection down to a series of binomial choices — such as between participating in self-employment or not — rather than treating occupational selection as a multinomial choice problem. Additionally, the existing literature does not account for differences *within* self-employment when considering patterns of labour force participation, despite the fact that informal sectors in developing countries are typically large and heterogeneous. We hope to address these issues in this paper, by explicitly treating occupational selection as a multinomial choice problem, which allows for different occupations *and* different technologies.

3 Job Flexibility

The notion of ‘flexibility’ is often invoked in the empirical literature to explain why certain individuals choose to work in self-employment or other informal labour market activities. This concept, however, is rarely formalised explicitly. We suggest three possible ways of thinking about job flexibility, backed up by a simple model of time allocation, and consider what this implies for the relationship between domestic obligations and occupational choice. In our model, and throughout this paper, ‘domestic obligations’ refer specifically to the minimum amount of time that individuals must devote to doing domestic work. We outline these ideas in this section, and reserve the formal treatment of the model for Appendix A.

Multi-Tasking. Certain types of self-employment activities may be undertaken concurrently with domestic work. For example, self-employed retailers may be able to run their stalls whilst simultaneously watching their children. However, multi-tasking in this way is likely to come at the expense of productivity in income-generating activities. Nonetheless, if individuals must provide more childcare or other domestic work to the household, they will be more willing to forego productivity in market work if doing so allows them to multi-task.

Minimum Hours. Alternatively, flexibility may relate to the minimum number of hours of market work that must be supplied in order to participate in a particular occupation. Formal wage jobs often require a certain number of hours to be committed each week, whereas the choice over how much to work in an informal self-employment job may be far less constrained.

The lower bound on the hours required to work in wage jobs may exclude individuals with substantial domestic obligations. If a certain number of hours are already committed to childcare and other domestic work, then individuals can only participate in more flexible types of jobs.

Adjustment Costs. A final possibility is that it may be easier in some jobs than others to adjust to *shocks* to domestic obligations, such as caring for family members who become sick. In formal wage jobs, it may be difficult to suddenly reduce working hours to deal with these types of unexpected events. Wage workers are typically contracted to work a set number of hours, and deviation from this plan can result in monetary or other types of penalties. Since self-employed workers do not have contracts and do not have to maintain a relationship with an employer, these penalties for sudden changes to time allocation are likely to be less.¹ Therefore, individuals who face greater potential shocks to their domestic obligations, may have a preference for self-employment work, even if these types of jobs are characterised by lower earnings.

The notions of job flexibility outlined above all suggest the same thing: individuals with greater domestic obligations are more likely to select more flexible jobs. In the remaining sections of this paper, we operationalise and test this prediction. Although we do not attempt directly to distinguish between the three stories, we do consider which may be more plausible given our data.

¹Even though self-employed workers do not have formal contracts, deviating from planned working hours may lead to other costs. For example, competitors may be able to capitalise on the gap in the market if a self-employed worker is suddenly absent for a sustained period.

4 Data and Descriptive Statistics

4.1 Sample Composition

Our data are taken from the fifth wave of the Ghana Living Standards Survey (GLSS5+), a nationally representative survey sampling approximately 8,000 households across the ten regions of Ghana.² Information on every individual within each surveyed household was collected, either through direct interview of the given respondent, or from the household head. After merging different sections of the survey, we have data on 37,098 individuals, 20,651 of whom are working age (15–65 years). The geographical distribution of the sample is shown in Table 10 in Appendix B.

The key outcome variable on which our analysis focusses is occupation. We show the breakdown of occupation, by sex, in Table 1. We restrict our analysis to primary occupation, defined as the job that the individual had undertaken most recently and spent most time doing. Importantly, we separate out the self-employed sample into enterprises that operate with and without others' labour. We label those that do not use labour besides their own 'own account' or 'low-input' self-employed, whereas those that use others' labour are labelled 'employer' or 'high-input' self-employed.³ This is a broad dimension on which to split the sample, but we believe it captures a key technology choice taken by entrepreneurs, particularly because the employment of labour is correlated with the use of other factors, such as capital (Woodruff, 2006). By dividing the sample in this way, we are able to investigate heterogeneity *within* the self-employment sector. This is important because, as we show below, factors that may be associated with job flexibility differ between low- and high-input self-employment activities.

Overall, 64 percent of working age men participate in the labour market, compared with 51 percent of women. This difference is somewhat larger than expected, mainly because of the composition of the 'Out of LF/Other' category, which includes not only those who are not working and not searching for work, but also apprentices, domestic workers (that is, 'house help'), and unpaid family workers. For parsimony, we classify those workers who do not receive money income as not employed (Sen, 1975).

The most striking differences in terms of occupational selection arise in low-input non-farm self-employment, where 76 percent of the participants are female. The sample of employers, however, is far more balanced where only 61 percent of the sample are women. This fact, in itself, suggests that the selection forces behind female and male participation in low- and high-input self-employment may

²The collection of GLSS5+ was undertaken by the Ghana Statistical Service (GSS) in conjunction with the World Bank. The survey contains a unique module collecting detailed information on non-farm household enterprises.

³The employers in our sample use labour from inside or outside the household for their businesses.

Table 1: Occupational Choice by Sex

Occupation	Sex					
	Female		Male		Total	
	N	%	N	%	N	%
Wage Employed	658	3.19	1768	8.56	2426	11.75
SE - Agriculture	1181	5.72	2736	13.25	3917	18.97
Non-Farm SE (Own Account)	2640	12.78	813	3.94	3453	16.72
Non-Farm SE (Employer)	807	3.91	526	2.55	1333	6.45
Unemployed	467	2.26	347	1.68	814	3.94
Out of LF/Other	5173	25.05	3535	17.12	8708	42.17
Total	10926	52.91	9725	47.09	20651	100.00
Sample of individuals of working age (15–65)						

differ. Also, women comprise just 27 percent of the wage-employed, further emphasising a disparity in the factors driving occupational choice.

In order to test the predictions from Section 3, we need to find a suitable proxy for domestic obligations. We capture this concept using the ‘dependency ratio’, which is calculated using Equation (1). ‘Dependents’ are defined as anyone aged under 15 or over 65. The dependency ratio is measured at the household-level, but the extra care requirements from having more children or elderly people in the household may not be split equally between women and men. Therefore we divide our results in Sections 6 and 7 by sex.

$$\text{Dependency Ratio} = \frac{\text{No. Dependents}}{\text{Household Size} - \text{No. Dependents}} \quad (1)$$

Information on the dependency ratio is provided in the summary statistics shown in Tables 11, 12, and 13 in Appendix B.

4.2 Job Characteristics

In this sub-section, we present descriptive statistics on the wage and non-wage characteristics of each occupation. This enables us to determine which occupations may be understood as flexible, drawing on the ideas outlined in Section 3.

4.2.1 Earnings

Our measures of earnings are calculated from the amount of money an individual receives from a job, including bonuses, commissions, allowances, or tips. We use this definition of earnings, as opposed to measures of enterprise value-added, to reduce bias when making cross-sectoral comparisons. We convert our measures of earnings from Second Ghana Cedis (GHC) to 2005 United States Dollars (USD).

Total monthly earnings are shown by sex and occupation in Table 2.

Table 2: Monthly Earnings

	Female				Male			
	N	Mean	S.Dev.	Median	N	Mean	S.Dev.	Median
Full Sample	5110	55.57	82.52	28.65	5092	80.26	99.26	47.74
Wage Employed	645	91.73	98.70	55.09	1728	113.52	110.04	87.65
Non-Farm SE (Own Account)	2615	54.11	79.35	28.65	788	84.59	108.73	47.74
Non-Farm SE (Employer)	791	72.80	103.04	38.20	510	106.55	115.98	66.11
Observations	5110				5092			
Outliers trimmed at the 1 st and 99 th percentiles								
Sample of individuals of working age (15–65)								

As expected, for the three sectors shown, median monthly earnings are highest in wage-employment, and lowest in low-input self-employment. Also, for the full sample, and all three occupations shown, median monthly earnings are higher for men than women.

We also adjust our earnings measures for the hours actually worked per month, and report the summary statistics in Table 3. Some economists have argued it is inappropriate to account for working hours in this way because it may not be possible for casual workers to scale up the time they work in a given month (Günther and Launov, 2012). However, given our focus on job flexibility, of which work hours may be an important dimension, we believe this adjustment is important.

Virtually the same patterns emerge in the summary statistics for hourly earnings, as compared to monthly earnings. At the median, men earn more than women across all sectors. The wage-employed enjoy the highest earnings, whilst the low-input self-employed earn the least.

The earnings differences between occupations at the median are echoed across the distribution. This is shown in Figure 9 in Appendix B.

Table 3: Hourly Earnings

	Female				Male			
	N	Mean	S.Dev.	Median	N	Mean	S.Dev.	Median
Full Sample	5090	0.44	0.86	0.18	5110	0.54	0.89	0.27
Wage Employed	641	0.60	0.96	0.34	1726	0.68	0.95	0.41
Non-Farm SE (Own Account)	2598	0.44	0.85	0.18	795	0.61	1.04	0.26
Non-Farm SE (Employer)	797	0.57	1.09	0.23	522	0.67	0.98	0.34
Observations	5090				5110			

Outliers trimmed at the 1st and 99th percentiles

Sample of individuals of working age (15–65)

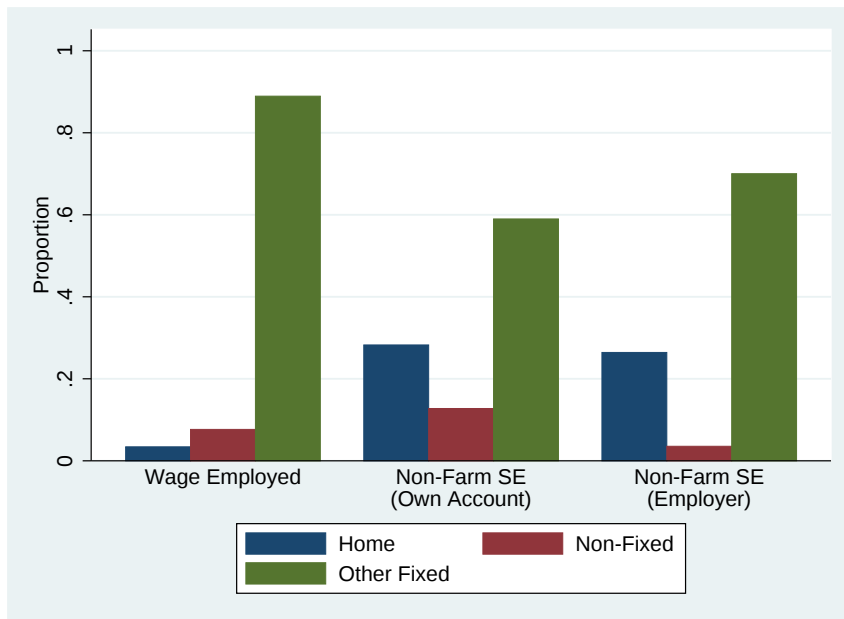
4.2.2 Job Location

In Figure 1 we show the proportion of individuals in wage-employment, low-input self-employment, and high-input self-employment, working in different locations. The blue bars represent individuals working at home or on their own land, the red bars represent individuals working without a fixed location, such as street vendors or those engaged in transport services, whilst the green bars represent individuals working in a fixed location away from home, such as an office, workshop, or stationary stall.

As expected, the wage-employed appear to work mainly in fixed locations away from home. There also appear to be small differences between low- and high-input self-employment activities. In particular, the proportion of individuals working in fixed locations away from home appears to be somewhat larger for the self-employed that utilise labour besides their own.

What can job location tell us about flexibility? Having a fixed working location away from home limits the extent to which it is possible to undertake domestic and market work simultaneously. Additionally, working in a fixed location away from home suggests that individuals are involved in work-based relationships with other employees, their customers, or their competitors. Insofar as these other employees, customers, or competitors operate according to fixed hours this may, in effect, constrain working hours and make it more costly to leave work to cater for shocks to domestic obligations. Thus, in-keeping with the three stories outlined in Section 3 these data suggest that low-input self-employment jobs are the most flexible.

Figure 1: Primary Work Location by Occupation



4.2.3 Work Hours

We also plot the hours worked per month for each occupation to investigate whether any of the occupations display evidence of sharp lower bounds on the amount of time that must be dedicated to market work. This is shown in Figure 2, disaggregating the data for women (Panel A) and men (Panel B).

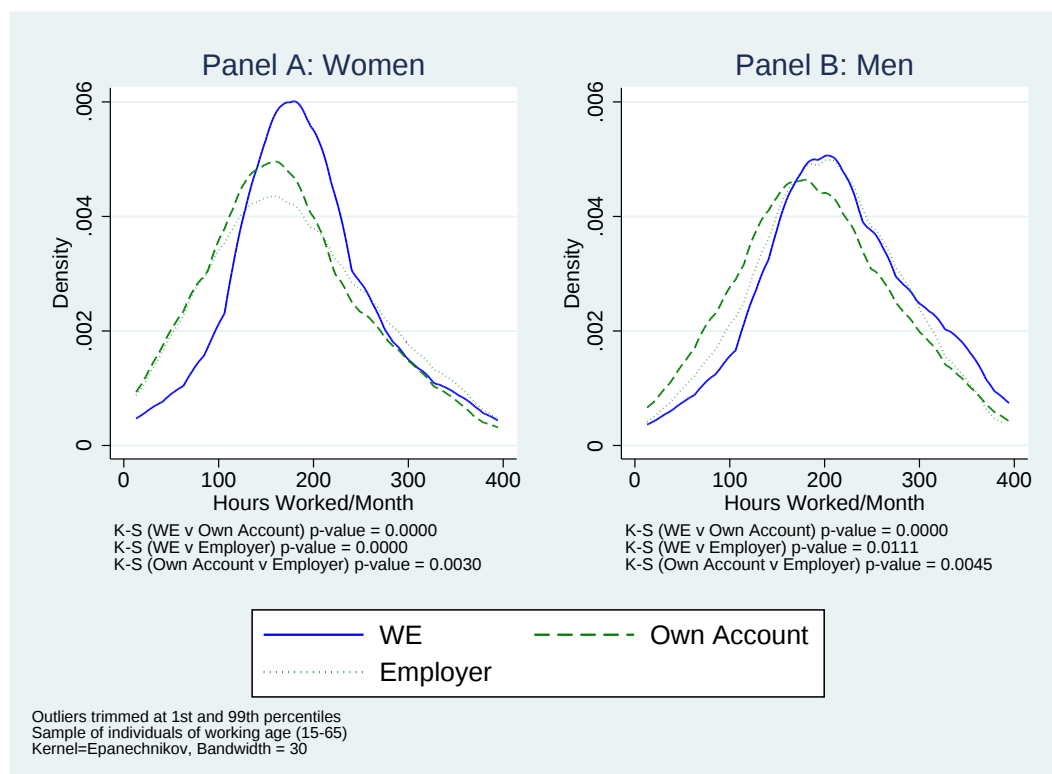
The distributions differ significantly across all three sectors, for both women and men.⁴ The most pronounced differences appear to be at the bottom of the distribution, where there are far fewer wage jobs than self-employment jobs, especially for the sub-sample of women. This is consistent with the notion that formal wage sector jobs are characterised by minimum hours. However, there is significant *intra*-sector heterogeneity, and there do appear to be some wage and high-input self-employment jobs, where individuals can participate but still work very few hours per month.

4.2.4 Time-Use

The previous sub-sections suggest that low-input self-employment jobs are the most flexible, and wage jobs are the least flexible. We now consider whether this is consistent with patterns of time-use.

⁴This is tested formally using the Kolmogorov-Smirnov (K-S) method, with the p-values reported under each graph.

Figure 2: Distributions of Hours Worked per Month by Occupation and Sex



In Figure 3, we show the average time that individuals devote to activities outside of market work each week, in different occupations. We separate out the time devoted to caring for other household members (in blue) from the time spent on other household chores, such as washing, cleaning, and running errands (in red).

Figure 3: Time Spent on Domestic Work by Occupation

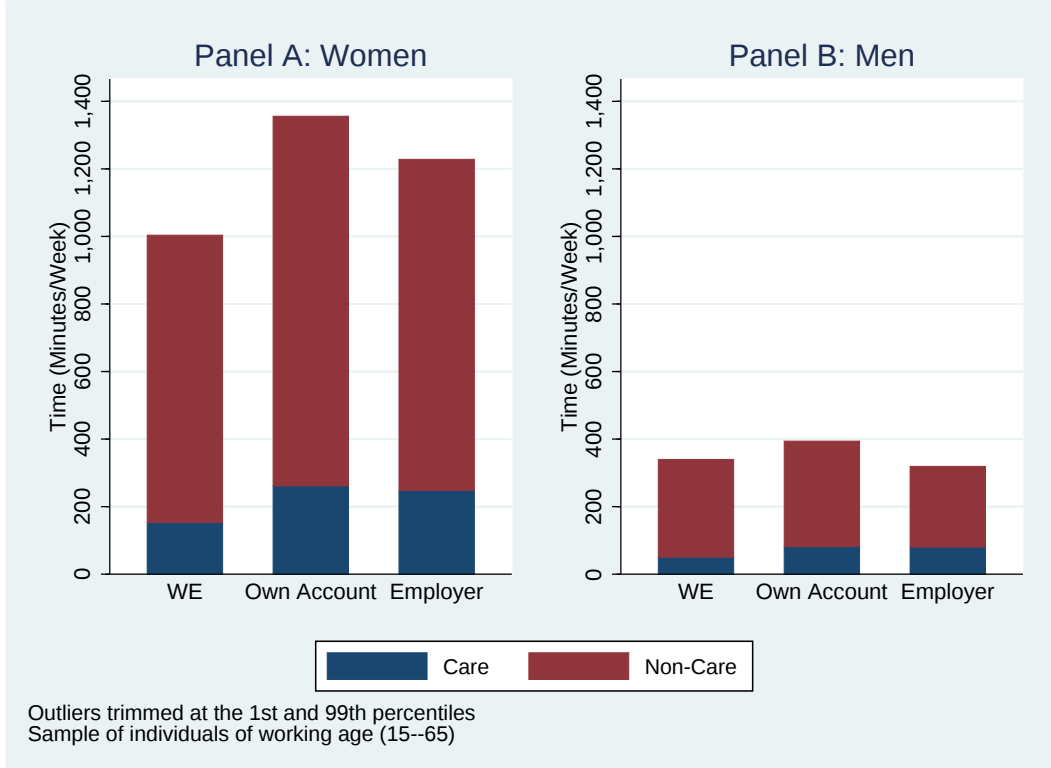


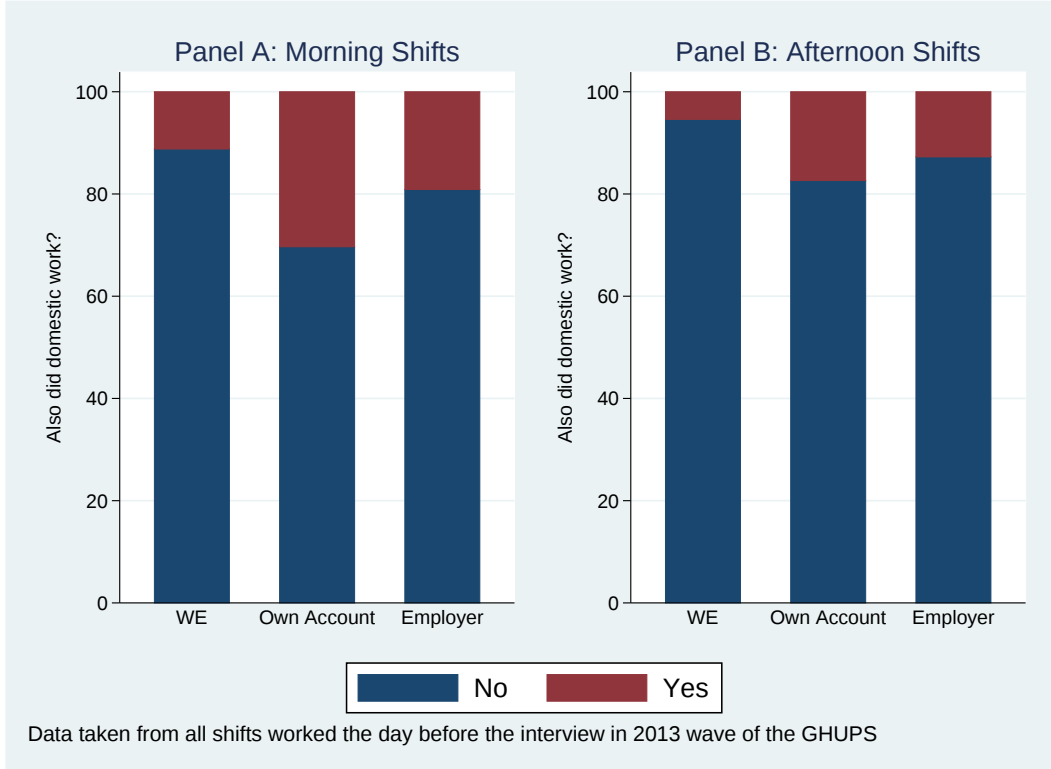
Figure 3 shows that, for both sexes, the wage-employed do the least domestic work, whilst the low-input self-employed individuals do the most. Furthermore, women do substantially more domestic work than men, across all occupations. These data are therefore consistent with two notions: (1) household-level domestic obligations fall disproportionately on women, and (2) low-input self-employment jobs are more flexible in the senses outlined in Section 3.

Whilst the prevalence of multi-tasking is not measured in the GLSS5+ data, the 2013 wave of the Ghana Household Urban Panel Survey — collected by the Centre for the Study of African Economies — contains a small time-use module, where individuals were able to *list* all the activities they undertook during the morning, afternoon, and evening on the previous day.⁵ We show the proportion of morning and afternoon market ‘work shifts’ in which the respondent also reported doing domestic work in Panels

⁵The morning was defined as 08:00-13:00. The afternoon was defined as 13:00-17:00. The evening was defined as 17:00-22:00.

A and B of Figure 4 respectively. We define a ‘work shift’ as a morning or afternoon during which the main activity was working in wage- or self-employment.⁶

Figure 4: Prevalence of Multi-Tasking by Occupation



In both the morning and afternoon shifts, domestic work is undertaken concurrently with market work more in self-employment, especially amongst own account workers. Again this supports the notion that low-input self-employment offers more job flexibility. Moreover, these data suggest that the multi-tasking model outlined in Section 3 may be a tenable way of interpreting our data on occupational choice.

⁶The sample size available from the 2013 wave of the GHUPS is far smaller than the GLSS5+. Additionally, the sample is drawn only from urban areas, focussing in particular on Accra, Kumasi, Cape Coast, and Takoradi. Pooling women and men, there are 864 wage-employed workers, 607 own account workers, and 165 employers. We therefore elect not to split the results in Figure 4 by sex.

5 Econometric Approach

In this section, we describe our main approach for assessing whether extra domestic obligations push workers towards more flexible jobs, in order to test the logic outlined in Section 3. To operationalise this question, we assume that low-input self-employment jobs are the most flexible. This is supported by the data presented in Section 4. We also use the household dependency ratio as a proxy for domestic obligations. However, we recognise that the strength of this proxy may differ according to sex, because the additional domestic work required to care for dependents may not be divided equally between female and male household members. We therefore split up the results for women and men.

Initially, we estimate this model using a multinomial logit, controlling for a wide range of observables relating to the individual's physical and human capital, their ethnicity, sources of unearned income, and their parents' profession. We also control for the size and location of the household.⁷

However, as discussed in Section 2, household structure may be endogenous to individuals' occupational selection. Unobserved characteristics may jointly influence decisions about employment and family planning. Moreover, individuals' occupations may affect family stability and thus fertility, whilst having extra children could also be a rational means of boosting household income for those working in jobs with low returns (Becker, 1985; Rosenzweig and Wolpin, 1980). In order to assess whether our results are sensitive to this potential endogeneity problem, we use selection on observables as a guide for selection on unobservables, building on the logic of Altonji et al. (2005).

To explain this technique, we begin by setting up the occupational choice problem as an Additive Random Utility Model.⁸ Utility for individual (i) in occupation ($j = 1, 2, \dots, J$) is a function of individual characteristics ($\mathbf{x}_i$) as well as some potentially endogenous variable relating to household structure, D_i . For this paper, D_i is the household dependency ratio. We observe alternative k being chosen by individual i if $u_{ik} > u_{il}$ for $\forall l \neq k$. Although $\mathbf{x}_i$ and D_i vary only at the individual level, and not across occupations, the coefficients and error terms are occupation-specific. Utility in the J sectors can thus be written:

$$\begin{aligned} u_{i1} &= \mathbf{x}_i' \boldsymbol{\alpha}_1 + \gamma_1 D_i + \xi_{i1} \\ u_{i2} &= \mathbf{x}_i' \boldsymbol{\alpha}_2 + \gamma_2 D_i + \xi_{i2} \\ &\vdots \\ u_{iJ} &= \mathbf{x}_i' \boldsymbol{\alpha}_J + \gamma_J D_i + \xi_{iJ} \end{aligned} \tag{2}$$

⁷Specifically, we control for location in terms of province and rural versus urban.

⁸This is the same model that underlies a regular multinomial logit or multinomial probit estimator.

We assume that the error terms are distributed multivariate Standard Normal, as in Equation (3).⁹

$$\begin{pmatrix} \xi_{i1} \\ \xi_{i2} \\ \vdots \\ \xi_{iJ} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right] \quad (3)$$

In order to estimate multinomial problems of this type, we must first respecify the model in terms of the *difference* between utility in each occupation compared to some base category. Following the convention in Train (2009), we use the first category as the base category.¹⁰ We operationalise this by setting $u_{i1} = 0$. The *relative* utilities $\forall j = 2, 3, \dots, J$ can then be written:

$$\begin{aligned} y_{ij'}^* &= u_{ij} - u_{i1} \\ &= \mathbf{x}_i'(\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_1) + (\gamma_j - \gamma_1)D_i + (\xi_{ij} - \xi_{i1}) \\ &= \mathbf{x}_i'\boldsymbol{\beta}_{j'} + \psi_{j'}D_i + \epsilon_{ij'} \end{aligned} \quad (4)$$

The new error terms, $\epsilon_{ij'}$, are thus distributed as in Equation (5).

$$\begin{pmatrix} \epsilon_{i2'} \\ \epsilon_{i3'} \\ \vdots \\ \epsilon_{iJ'} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{pmatrix} \right] \quad (5)$$

We cater for endogeneity by assuming a specific structure for the unexplained component of the endogenous household structure variable, D_i . First, we imagine a ‘first-stage’ selection equation, which relates the endogenous variable D_i to the other observable variables in $\mathbf{x}_i$. We do not include any excluded instrumental variables in this equation. Instead, we tackle endogeneity by considering different

⁹Specifying the model with a multivariate Normal error variance-covariance matrix, in itself, adds computational complexity for very little gain compared to a regular multinomial logit model, especially as the off-diagonal terms are all set to zero (Cameron and Trivedi, 2005; Long and Freese, 2006). Indeed, fixing the off-diagonal terms to zero in this way is analogous to imposing the Independence of Irrelevant Alternatives assumption, associated with the multinomial logit. However, we eventually use this functional form assumption to allow for the presence of the endogenous household structure variable, D_i .

¹⁰Although the choice of the base category will influence the parameters in $\boldsymbol{\alpha}_j$ and γ_j , this does not affect the marginal effects.

assumptions about the error term v_i .

$$D_i = \mathbf{x}_i' \boldsymbol{\pi} + v_i \quad (6)$$

Under the assumption that D_i is exogenous, the error terms in Equations (4) and (6), $\epsilon_{ij'}$ and v_i , are uncorrelated. Endogeneity arises when this correlation is non-zero. We follow Rosenbaum and Rubin (1983) and Greene (2003) and formalise this potential endogeneity using a multivariate Normal distribution for the error terms.

$$\begin{pmatrix} \epsilon_{i2'} \\ \epsilon_{i3'} \\ \vdots \\ \epsilon_{iJ'} \\ v_i \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 & \cdots & 1 & \rho\sigma \\ 1 & 2 & \cdots & 1 & \rho\sigma \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 2 & \rho\sigma \\ \rho\sigma & \rho\sigma & \cdots & \rho\sigma & \sigma^2 \end{pmatrix} \right] \quad (7)$$

The correlation between v_i and *all* the $\epsilon_{ij'}$ terms is governed by the parameter ρ .¹¹ Under the assumption that D_i is exogenous, $\rho = 0$. However, by estimating the model with different values of ρ , we can test the sensitivity of our results to relaxing this exogeneity assumption. Therefore, we do not estimate ρ , as one might with a regular Heckman selection model with excluded instruments, but rather change it manually. As such, the only free parameter in the error variance-covariance matrix is the standard deviation of v_i , labelled σ .¹²

In order to write the likelihood function, we first need to derive the distribution of $(\epsilon_{i2'}, \epsilon_{i3'}, \dots, \epsilon_{iJ'})$ conditional on v_i . Using the properties of a Multivariate Normal distribution we can write:

$$\left(\begin{array}{c} \epsilon_{i2'} \\ \epsilon_{i3'} \\ \vdots \\ \epsilon_{iJ'} \end{array} \middle| v_i \right) \sim N \left[\begin{pmatrix} \rho\sigma^{-1}v_i \\ \rho\sigma^{-1}v_i \\ \vdots \\ \rho\sigma^{-1}v_i \end{pmatrix}, \begin{pmatrix} (2-\rho^2) & (1-\rho^2) & \cdots & (1-\rho^2) \\ (1-\rho^2) & (2-\rho^2) & \cdots & (1-\rho^2) \\ \vdots & \vdots & \ddots & \vdots \\ (1-\rho^2) & (1-\rho^2) & \cdots & (2-\rho^2) \end{pmatrix} \right] \quad (8)$$

¹¹The assumption that this correlation is the same for all of the latent variable error terms is strong. However, since ρ is a parameter to be altered by the econometrician, we believe this approach is still informative about the impact of endogeneity. Allowing for different values of ρ for each sector would dramatically increase the number of possible assumptions about endogeneity, and is beyond the scope of this paper.

¹²To simplify computation in our empirical results section, we normalise our variables such that $\sigma = 1$.

Following Greene (2003), we rewrite the latent variable equations such that the error terms are distributed with a 0 vector for the mean.

$$\begin{aligned} y_{ij'}^* &= \mathbf{x}_i' \boldsymbol{\beta}_{j'} + \psi_{j'} D_i + \epsilon_{ij'} \\ &= \mathbf{x}_i' \boldsymbol{\beta}_{j'} + \psi_{j'} D_i + \rho \sigma^{-1} v_i + \zeta_{ij'} \end{aligned} \quad (9)$$

The error variance-covariance matrix may now be written:

$$\left(\begin{array}{c|c} \zeta_{i2'} & \\ \zeta_{i3'} & v_i \\ \vdots & \\ \zeta_{iJ'} & \end{array} \right) \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} (2-\rho^2) & (1-\rho^2) & \cdots & (1-\rho^2) \\ (1-\rho^2) & (2-\rho^2) & \cdots & (1-\rho^2) \\ \vdots & \vdots & \ddots & \vdots \\ (1-\rho^2) & (1-\rho^2) & \cdots & (2-\rho^2) \end{pmatrix} \right] \quad (10)$$

With this error structure in place, we can now begin to write the likelihood function. Labelling the observed occupation for individual i as y_i , and the parameter vector $\boldsymbol{\theta}$, we can write the likelihood function for the individual (Train, 2009).¹³

$$L_i(\boldsymbol{\theta}; y_i \mid \mathbf{x}_i, D_i) = \prod_{j=1}^J \left\{ \mathbb{1}(y_i = j) \times \Pr(y_i = j \mid \mathbf{x}_i, D_i) \times \frac{1}{\sigma} \phi\left(\frac{v_i}{\sigma}\right) \right\} \quad (11)$$

Replacing v_i using the observables in Equation (6), and taking logs, we can write the log likelihood for the individual.

$$l_i(\boldsymbol{\theta}; y_i \mid \mathbf{x}_i, D_i) = \sum_{j=1}^J \left\{ \mathbb{1}(y_i = j) \times \ln [\Pr(y_i = j \mid \mathbf{x}_i, D_i)] + \ln \left[\frac{1}{\sigma} \phi\left(\frac{D_i - \mathbf{x}_i' \boldsymbol{\pi}}{\sigma}\right) \right] \right\} \quad (12)$$

For the sample as a whole, we can write:

¹³We denote the Standard Normal distribution using ϕ and Φ for the PDF and CDF respectively.

$$l(\boldsymbol{\theta}; \mathbf{y} \mid \mathbf{x}, \mathbf{D}) = \sum_{i=1}^N \sum_{j=1}^J \left\{ \mathbb{1}(y_i = j) \times \ln [\Pr(y_i = j \mid \mathbf{x}_i, D_i)] + \ln \left[\frac{1}{\sigma} \phi \left(\frac{D_i - \mathbf{x}_i' \boldsymbol{\pi}}{\sigma} \right) \right] \right\} \quad (13)$$

We can now estimate the parameters in $\boldsymbol{\theta}$ using maximum simulated likelihood, given our distributional assumptions about the error terms in Equation (10).¹⁴ We calculate the simulated probability for each individual using a ‘logit-smoothed’ simulator (Train, 2009; Adams et al., 2015).¹⁵ Importantly, we can do this for different values of ρ , thus testing different assumptions about the endogeneity of D_i .

Greater values of ρ imply more selection of D_i on the basis of unobservables, and hence more endogeneity. However, it is difficult to assess the magnitude of ρ . The question remains over *how large* ρ can become before we are content our results are robust. Put differently, how much selection of D_i on the basis of unobservables are we willing to allow for? One possibility is to assume that the selection (of D_i) on unobservables is *equal* to the selection on observables (Altonji et al., 2005, 2008; Oster, 2013). In reality, we hope that our control variables are sufficiently relevant to explain household demographics *more* than the unobservables, so the ‘equal selection’ condition can be understood as something of an upper bound on how bad endogeneity could become. Thus, if our main results survive under this condition, this supports the notion that they are robust to concerns about endogeneity.

We label the value of ρ , which implies ‘equal selection’, $\tilde{\rho}_{j'}$. Since there are different sets of unobservables for each occupational category, this value must be indexed by j' . This reflects the fact that the factors driving individuals to select wage-employment, relative to the base category, are different from the factors for self-employment. Thus, when we conduct our sensitivity analysis, we assess our results as though the *common* ρ had been increased to the $\tilde{\rho}_{j'}$ for sector j' as well as for the sector where $\tilde{\rho}_{j'}$ is highest. Although, evaluating our results in terms of a common ρ for all occupational categories is somewhat restrictive, it still allows us to use selection on observables as a guide to selection on unobservables in a parsimonious way.

Following Altonji et al. (2005), we can write the equal selection condition as in Equation (14).

$$\tilde{\rho}_{j'} = \frac{\text{Cov}(\mathbf{x}_i' \boldsymbol{\beta}_{j'}, \mathbf{x}_i' \boldsymbol{\pi})}{\text{Var}(\mathbf{x}_i' \boldsymbol{\beta}_{j'})} \quad (14)$$

¹⁴We prefer this method to using our preliminary assumptions in Equation (3) to derive a closed-form likelihood function. We show, however, that this may be possible in Appendix C.

¹⁵This approach helps overcome some of the short-comings of a simple ‘accept-reject’ simulator, whilst maintaining parsimony for coding the estimator. We simulate with 1000 repetitions.

Thus, by checking whether our effects survive when ρ is pushed to $\tilde{\rho}_j$ and beyond, we directly capture the idea that selection on observables serves as a guide for selection on unobservables.

6 Results

6.1 Multinomial Logit Results

We report marginal effects for the family demographic variables included in the multinomial logit selection equations in Table 4.¹⁶

Table 4: Main Marginal Effects on Occupational Selection

	Female			Male		
	Own Account	Employer	WE	Own Account	Employer	WE
Dependency Ratio	0.0451*** (0.0054)	0.0068** (0.0033)	0.0062* (0.0036)	0.0127*** (0.0048)	0.0038 (0.0039)	-0.0222*** (0.0068)
Married? (1=Y, 0=N)	0.0587*** (0.0102)	0.0214*** (0.0068)	-0.0095* (0.0056)	0.0018 (0.0072)	0.0184*** (0.0065)	0.0133 (0.0090)
Household Size	-0.0095*** (0.0020)	0.0032*** (0.0010)	-0.0034** (0.0014)	-0.0072*** (0.0016)	0.0014 (0.0012)	-0.0048** (0.0019)
Education (Years)	0.0012 (0.0010)	0.0013** (0.0006)	0.0098*** (0.0006)	-0.0013** (0.0006)	-0.0007 (0.0005)	0.0133*** (0.0009)
Log of Age	0.1561*** (0.0110)	0.0988*** (0.0071)	0.0340*** (0.0060)	0.0440*** (0.0088)	0.0202*** (0.0077)	0.0941*** (0.0114)
N	10926	10926	10926	9725	9725	9725
Log-Likelihood	-12290.5825	-12290.5825	-12290.5825	-9118.6356	-9118.6356	-9118.6356
Pseudo- R^2	0.2156	0.2156	0.2156	0.3826	0.3826	0.3826

Standard errors in parentheses

Base category is 'Out of the Labour Force'

Marginal Effects for 'Agricultural Self-Employment' and 'Unemployment' not reported

Standard errors clustered at the household level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Higher domestic obligations for the household push women into own account self-employment more than other occupations. Using Table 4 in conjunction with the summary statistics in Appendix B, we can see that a 1 standard deviation increase in the dependency ratio implies women are 3.4 percent more likely to enter own account self-employment. By contrast, the effects of the dependency ratio on selection into wage-employment and high-input self-employment, whilst statistically significant, are

¹⁶We use Hausman and McFadden's (1984) method to test whether our results are sensitive to the IIA assumption. We re-estimate the multinomial logit, omitting each category in turn, and examine whether the coefficients on all the variables change significantly. For both the female and male sub-samples, it is only when we omit the 'Out of the Labour Force' category that our results change substantially, causing us to reject the null of the Hausman-McFadden test. This suggests imposing the IIA, as is implied by the multinomial logit, may not be too restrictive.

small.¹⁷ The results for the other household demographic variables tell a similar story. In particular, marriage appears to make entry into low-input self-employment much more likely, suggesting that it is women with their own families that are more likely to choose this type of work.

The picture is more mixed for men. The impact of the dependency ratio on male selection into own account self-employment is far smaller than for women. Using the descriptive statistics in Appendix B once again, a 1 standard deviation in the dependency ratio increases the probability of entry into low-input self-employment by just 0.8 percent. Moreover, the effects of the other household demographic variables on selection into own account self-employment appear to be weaker, with marriage having no statistically significant effects.

These results therefore suggest that increased household-level domestic obligations drive women into flexible jobs, such as low-input self-employment, more than men.

We also report the marginal effects for education and age to help link these results to the existing literature. Our findings echo previous work on the effect of human capital on occupational choice in Ghana. Firstly, extra education increases the chances of working in wage-employment, for both women and men. However, the marginal effects of education on entry into own account work are either statistically insignificant, or slightly negative. Thus, access to schooling may serve as a barrier to entry into wage work, but not low-input self-employment. There are also positive marginal effects for age for all three of the occupational categories shown, reflecting the fact that non-participation is more likely among the young.

6.2 Assessing Endogeneity

We now test the sensitivity of the results to different assumptions about endogeneity, using the framework outlined in Section 5. Since this technique is computationally intensive we make three modifications to our original selection equations. First, we reduce the number of categories from six to four, by collapsing ‘Self-Employed — Agriculture’, ‘Unemployed’, and ‘Out of the Labour Force’ into one category. Second, we reduce the number of variables included as controls. We retain age, education, land holdings, and some household-level characteristics, but are unable to include the ethnicity and household location variables. Third, we normalise our variables, such that for each variable x_i , $E(x_i) = 0$ and $\text{Var}(x_i) = 1$.

¹⁷The fact that greater household-level domestic obligations appear to increase women’s chances of working in all three types of jobs echoes existing evidence on the household division of responsibilities in Ghana. In particular, it may be that there are certain goods — such as food or clothes for children — which women are expected to buy for the household due to social norms (Warner et al., 1997; Goldstein, 2004). As such, the presence of extra dependents in the household may drive women’s participation in market work.

The sensitivity of the marginal effect on the dependency ratio to different assumptions about endogeneity is shown in Table 5. In the leftmost column, we show the results from a multinomial logit model, having made the changes to the data described above. We then report the analogous results derived from an independent multinomial probit model, calculated with a closed-form likelihood function.¹⁸ In the remaining columns, we report our maximum simulated likelihood results, under different assumptions about ρ . We also show the absolute value of $\tilde{\rho}_{j'}$ for each sector in the final column, which indicates the level of ρ that would imply that the dependency ratio is selected *equally* by observables and the unobservables for that sector.

Table 5: Assessing Endogeneity using Maximum Simulated Likelihood

	MNL	MNP	$\rho = 0$	Maximum Simulated Likelihood MNP					$ \tilde{\rho}_{j'} $
			$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$	$\rho = 0.2$	$\rho = 0.25$		
<i>Female</i>									
Own Account	0.0426 (11.91)	0.0428 (11.79)	0.0416 (11.66)	0.0347 (9.69)	0.0265 (7.36)	0.0194 (5.36)	0.0119 (3.28)	0.0065 (1.80)	0.1783
Employer	0.0051 (2.25)	0.0046 (2.00)	0.0047 (2.39)	0.0027 (1.16)	-0.0005 (-0.20)	-0.0026 (-1.08)	-0.0050 (-2.09)	-0.0085 (-3.48)	0.0605
WE	-0.0006 (-0.29)	-0.0002 (-0.11)	-0.0021 (-0.95)	-0.0048 (-2.29)	-0.0061 (-2.84)	-0.0082 (-3.71)	-0.0115 (-5.13)	-0.0125 (-5.47)	0.2110
<i>Male</i>									
Own Account	0.0133 (4.07)	0.0119 (3.65)	0.0114 (3.56)	0.0074 (2.33)	0.0056 (1.75)	0.0001 (0.02)	-0.0032 (-1.01)	-0.0083 (-2.57)	0.1904
Employer	0.0067 (2.65)	0.0066 (2.55)	0.0077 (2.93)	0.0051 (1.93)	0.0044 (1.66)	-0.0010 (-0.37)	-0.0030 (-1.14)	-0.0044 (-1.66)	0.1096
WE	-0.0192 (-4.08)	-0.0182 (-4.05)	-0.0183 (-4.00)	-0.0250 (-5.45)	-0.0299 (-6.51)	-0.0375 (-8.22)	-0.0458 (-9.86)	-0.0523 (-11.26)	0.1427

t-statistics in parentheses

Base category is all other working age individuals

Standard errors for maximum simulated likelihood results calculated using Stata's `mprobit` command

Absolute values for $\tilde{\rho}_{j'}$ reported

Firstly, transforming the data leaves the main results largely unchanged, with the largest positive marginal effects pushing women into low-input self-employment. This, in itself, demonstrates the robustness of the multinomial logit results from Section 6.1. Making more conservative assumptions about endogeneity reduces the key marginal effects, as anticipated. However, the marginal effect on women's selection into own account work remains positive and significant at the 10 percent level when the common $\rho = 0.25$. For the unobservables associated with own account self-employment, $|\tilde{\rho}_{j'}| = 0.1783$, so the positive effects would survive even if we assumed the dependency ratio was selected *equally* by observables and these unobservables.¹⁹ Indeed, even if the common ρ were set at the highest level of

¹⁸We use Stata's `mprobit` to derive these results. A closed-form for the log-likelihood is found using the result due to Dunnett (1989). As in the multinomial logit, and indeed our maximum simulated likelihood estimator, the `mprobit` command assumes there is no correlation between the error terms for utility in each occupation, as in Equation (3).

¹⁹Technically, $\tilde{\rho}_{j'}$ can be less than 0. However estimating the model with negative values for ρ strengthens the positive effect of the dependency ratio on women's selection into own account self-employment, so does not serve as a test of the

$|\tilde{\rho}_{j'}| - 0.2110$ for the unobservables associated with selecting wage-employment — the key marginal effect would still be positive and significant.

The extent to which we believe that equal selection on observables and unobservables is sufficient to demonstrate that our results are robust to endogeneity depends on the effectiveness of our controls at explaining the dependency ratio. If the control variables are strongly related to the dependency ratio, then the restriction that unobservables have an equal effect may be quite conservative. However, if the control variables do little to explain the dependency ratio, we may expect unobservables to do more selection than observables (Oster, 2013). Although we are somewhat constrained by the computational intensity of this estimator, we have maintained control variables on age, education, land holdings, and other household characteristics (such as household size and spouse education). We believe these controls drive the dependency ratio sufficiently for the ‘equal selection’ condition, encapsulated by $\tilde{\rho}_{j'}$, to provide a useful yardstick against which to judge our results’ robustness to endogeneity.

As such, it appears our main finding — that having a higher dependency ratio for the household pushes women towards low-input self-employment — is somewhat robust to concerns about the endogeneity of household-level domestic obligations.

7 Robustness and Heterogeneity

7.1 Respecifying Domestic Obligations

Thus far, we have proxied for domestic obligations using the household dependency ratio, whilst also reporting results for the household size and marital status. We now try and disentangle *which* household dependents drive our results by respecifying the selection equations, including the number of infants (aged < 2), young children (aged 2–4), other children (aged 5–14), and elders (aged > 65), as shown in Table 6.

Overall, these results echo our initial findings from Table 4. Also, we see that the number of children has the largest marginal effects on women’s entry into low-input self-employment, whilst the number of elders in the household matters less. A 1 standard deviation increase in the number of young children in the household increases the likelihood of female participation in own account self-employment by around 3 percent, whilst the effects on entry into wage-employment and high-input self-employment are minimal.

robustness of our results. As such, we report the absolute values of $\tilde{\rho}_{j'}$ in Table 5 and use these as the benchmark.

Table 6: Marginal Effects for Occupational Selection:
Respecifying Domestic Obligations

	Female			Male		
	Own Account	Employer	WE	Own Account	Employer	WE
Married? (1=Y, 0=N)	0.0582*** (0.0102)	0.0232*** (0.0068)	-0.0098* (0.0056)	0.0015 (0.0073)	0.0168*** (0.0065)	0.0078 (0.0090)
No. Infants (< 2 years) in HH	0.0387*** (0.0088)	-0.0010 (0.0054)	0.0091 (0.0070)	0.0151** (0.0065)	0.0074 (0.0052)	0.0216** (0.0087)
No. Young Children (2–4 years) in HH	0.0462*** (0.0074)	-0.0079* (0.0044)	0.0056 (0.0044)	0.0119** (0.0055)	0.0131*** (0.0039)	-0.0050 (0.0070)
No. Older Children (5–14 years) in HH	0.0304*** (0.0049)	0.0065** (0.0026)	0.0052* (0.0031)	0.0099** (0.0038)	0.0036 (0.0029)	-0.0047 (0.0046)
No. Elders (> 65 years) in HH	-0.0002 (0.0120)	-0.0135* (0.0070)	-0.0005 (0.0079)	0.0116 (0.0107)	0.0027 (0.0086)	-0.0351** (0.0146)
Household Size	-0.0246*** (0.0032)	0.0016 (0.0016)	-0.0055*** (0.0021)	-0.0114*** (0.0025)	-0.0013 (0.0019)	-0.0055** (0.0028)
N	10926	10926	10926	9725	9725	9725
Log-Likelihood	-12243.2330	-12243.2330	-12243.2330	-9053.1369	-9053.1369	-9053.1369
Pseudo- R^2	0.2186	0.2186	0.2186	0.3871	0.3871	0.3871

Standard errors in parentheses

Base category is 'Out of the Labour Force'

Marginal Effects for 'Agricultural Self-Employment' and 'Unemployment' not reported

Standard errors clustered at the household level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In contrast, the marginal effect on the number of elders for female entry into low-input self-employment is small and statistically insignificant.

The effects that extra children have on men's likelihood of participating in own account work are generally weaker than for women. A 1 standard deviation increase in the number of young children in the household increases the likelihood of male participation in own account self-employment by just 0.7 percent.

Respecifying the selection equations using alternative proxies of domestic obligations not only reinforces our main results, but also allows us to unpack how job flexibility may influence occupational selection. It appears to be extra young children, which push women into self-employment rather than the presence of elderly relatives. Insofar as young children require some level of *constant* care, this suggests that the multi-tasking and minimum hours models outlined in Section 3 are more suitable for interpreting the data. In contrast, the number of elders, whose care requirements may be more variable over time, has little effect. Coping with *volatile* domestic obligations, as in the adjustment costs story, appears to matter less.

7.2 Heterogeneity by Marital Status

We may expect that choices about family structure are likely to be decided more by married individuals rather than unmarried individuals in the household. As such, the potential endogeneity of household structure to occupational choice may be less in the unmarried sub-sample.

To test this hypothesis, we begin by re-estimating our main multinomial logit results, splitting the sample by marital status. These results are reported in Tables 7 and 8. The dependency ratio appears to have stronger effects for the unmarried sub-sample, for both women and men.

Table 7: Main Marginal Effects on Female Job Selection by Marital Status

	Unmarried			Married		
	Own Account	Employer	WE	Own Account	Employer	WE
Household Size	-0.0186*** (0.0023)	-0.0001 (0.0015)	-0.0055*** (0.0020)	-0.0049** (0.0025)	0.0044*** (0.0013)	-0.0011 (0.0012)
Dependency Ratio	0.0520*** (0.0058)	0.0068* (0.0038)	0.0022 (0.0058)	0.0363*** (0.0087)	0.0069 (0.0057)	0.0044 (0.0043)
N	5558	5558	5558	5368	5368	5368
Log-Likelihood	-5619.7711	-5619.7711	-5619.7711	-6783.2227	-6783.2227	-6783.2227
Pseudo- R^2	0.2544	0.2544	0.2544	0.1369	0.1369	0.1369

Standard errors in parentheses

Base category is 'Out of the Labour Force'

Marginal Effects for 'Agricultural Self-Employment' and 'Unemployment' not reported

Standard errors clustered at the household level

Ethnicity dummies, unearned income sources, and parents' profession variables omitted to allow computation

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Main Marginal Effects on Male Job Selection by Marital Status

	Unmarried			Married		
	Own Account	Employer	WE	Own Account	Employer	WE
Household Size	-0.0079*** (0.0016)	-0.0030*** (0.0012)	-0.0109*** (0.0023)	-0.0036* (0.0021)	0.0047*** (0.0016)	0.0049** (0.0025)
Dependency Ratio	0.0169*** (0.0052)	0.0074* (0.0038)	-0.0301*** (0.0098)	0.0112 (0.0080)	0.0075 (0.0068)	-0.0222** (0.0095)
N	5467	5467	5467	4258	4258	4258
Log-Likelihood	-4653.6852	-4653.6852	-4653.6852	-4650.1200	-4650.1200	-4650.1200
Pseudo- R^2	0.3276	0.3276	0.3276	0.2236	0.2236	0.2236

Standard errors in parentheses

Base category is 'Out of the Labour Force'

Marginal Effects for 'Agricultural Self-Employment' and 'Unemployment' not reported

Standard errors clustered at the household level

Ethnicity dummies, unearned income sources, and parents' profession variables omitted to allow computation

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To examine whether the potential endogeneity of domestic obligations is less problematic for the unmarried sub-sample, we repeat the sensitivity analysis from Section 6.2 using maximum simulated likelihood for different assumptions about the value of ρ . As before, we collapse the number of categories from six to four, reduce the number of controls, and normalise our variables to aid computation. The

results are shown in Table 9.

Table 9: Assessing Endogeneity for the Unmarried Sub-Sample

	MNL	MNP	$\rho = 0$	Maximum $\rho = 0.05$	Simulated $\rho = 0.1$	Likelihood $\rho = 0.15$	MNP $\rho = 0.2$	$\rho = 0.25$	$ \tilde{\rho}_{j'} $
<i>Female</i>									
Own Account	0.0398 (10.38)	0.0391 (9.91)	0.0418 (10.31)	0.0352 (8.60)	0.0300 (7.32)	0.0236 (5.75)	0.0171 (4.16)	0.0130 (3.14)	0.1419
Employer	0.0038 (1.60)	0.0032 (1.35)	0.0062 (2.00)	0.0044 (1.80)	0.0033 (1.40)	0.0006 (0.26)	-0.0010 (-0.40)	-0.0028 (-1.17)	0.0859
WE	-0.0022 (-0.63)	-0.0022 (-0.67)	-0.0056 (-1.82)	-0.0066 (-2.12)	-0.0107 (-3.43)	-0.0122 (-3.79)	-0.0142 (-4.41)	-0.0194 (-5.86)	0.1898
<i>Male</i>									
Own Account	0.0114 (3.09)	0.0105 (2.76)	0.0136 (3.48)	0.0114 (2.93)	0.0076 (1.96)	0.0066 (1.73)	0.0002 (0.06)	-0.0016 (-0.40)	0.2840
Employer	0.0022 (0.78)	0.0011 (0.39)	0.0020 (0.79)	0.0002 (0.09)	-0.0006 (-0.24)	-0.0017 (-0.68)	-0.0036 (-1.29)	-0.0052 (-1.91)	0.2424
WE	-0.0311 (-4.47)	-0.0319 (-4.99)	-0.0326 (-5.10)	-0.0400 (-6.23)	-0.0454 (-7.02)	-0.0530 (-8.12)	-0.0575 (-8.73)	-0.0679 (-10.12)	0.2150

t-statistics in parentheses

Base category is all other working age individuals

Standard errors for maximum simulated likelihood results calculated using Stata's `mprobit` command

Absolute values for $\tilde{\rho}_{j'}$ reported

The results for the unmarried sub-sample are more robust to alternative assumptions about endogeneity than the results for the full sample. When the common ρ is set at 0.25, the marginal effect of the dependency ratio on women's entry into own account self-employment remains positive and significant at the 1 percent level and is more than double the size seen in the full sample. Also, $|\tilde{\rho}_{j'}|$ is slightly lower for all three occupations for the unmarried sub-sample, focussing on the women. This suggests that the lower bound estimates of the marginal effects, under the assumption that selection of the dependency ratio by observables and unobservables is equal, are somewhat higher for the unmarried than for the full sample.

8 Conclusion

In this paper, we investigate whether non-monetary characteristics of certain jobs cause female and male workers to select occupations differently. In particular, we test the hypothesis that women are drawn into low-input self-employment activities more than men, as these types of jobs are more flexible. We outline three stories for 'job flexibility' in terms of (1) multi-tasking, (2) minimum hours requirements, and (3) adjustment costs.

To examine the importance of job flexibility empirically, we test whether the household dependency ratio — a proxy for household-level domestic obligations — influences individuals’ likelihood of participating in jobs that are characterised by greater flexibility — namely low-input self-employment. These results are disaggregated by sex, because the extra domestic work requirements associated with having more dependents in the household may not be divided equally between women and men.

We find that women from households with greater domestic obligations are more likely to select into low-input self-employment. A 1 standard deviation increase in the dependency ratio increases women’s chances of doing own account self-employment work by 3.4 percent. The analogous effects for men are much weaker. It appears that this effect is largely driven by having extra young children, and that the number of elderly people in the household has little effect on women’s occupational choice. As such, we believe this result is more consistent with the multi-tasking and minimum hours stories, rather than the adjustment costs.

Household structure and occupational choice may be jointly determined by a number of unobservable factors, such as family stability, understanding of family planning, and social norms. We develop a new estimator using maximum simulated likelihood to capture the idea of using selection on observables as a guide to selection on unobservables to tackle this endogeneity concern. Our main finding — that women from households with greater domestic obligations are pushed into low-input self-employment — is robust to more conservative assumptions about the nature of endogeneity. We also find that this potential endogeneity is likely to be less problematic for the sub-sample of unmarried individuals. We argue that this is because these individuals have less control over the structure of their household.

References

- C. Ackah. Non-farm employment and incomes in rural Ghana. *Journal of International Development*, 2013.
- C. Ackah, C. Ahiadeke, and A. Fenny. Determinants of female labour force participation in Ghana. Working paper, Global Development Network, 2009.
- A. Adams, D. Clarke, and S. Quinn. *Microeconometrics and MATLAB: An Introduction*. Oxford University Press, 2015.
- J. G. Altonji, T. E. Elder, and C. R. Taber. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *The Journal of Political Economy*, 2005.
- J. G. Altonji, T. E. Elder, and C. R. Taber. Using selection on observed variables to assess bias from unobservables when evaluating Swan-Ganz catheterization. *American Economic Review*, 2008.
- J. D. Angrist and W. N. Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 1998.
- W. Baah-Boateng, E. Nketiah-Amponsah, and R. Frempong. The effect of fertility and education on female labour force participation in Ghana. *Ghanaian Journal of Economics*, 2013.
- G. S. Becker. Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, 1985.
- S. G. Bronars and J. Grogger. The economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review*, 1994.
- A. C. Cameron and P. K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005.
- C. W. Dunnett. Algorithm AS 251: Multivariate normal probability integrals with production correlation structure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1989.
- D. S. Evans and L. S. Leighton. Some empirical aspects of entrepreneurship. *The American Economic Review*, 1989.
- P. Fajnzylber, W. F. Maloney, and G. Rojas. Microenterprise dynamics in developing countries: How similar are they to those in the industrialized world? Evidence from Mexico. *World Bank Economic Review*, 2006.
- M. Goldstein. Intrahousehold efficiency and individual insurance in Ghana. Development economics discussion paper, London School of Economics, 2004.

- W. H. Greene. *Econometric Analysis: Fifth Edition*. Pearson Education, 2003.
- I. Günther and A. Launov. Informal employment in developing countries: Opportunity or last resort? *Journal of Development Economics*, 2012.
- J. A. Hausman and D. L. McFadden. Specification tests for the multinomial logit model. *Econometrica*, 1984.
- J. Heintz and L. Pickbourn. The determinants of selection into non-agricultural self-employment in Ghana. *The Journal of Applied Economic Research*, 2012.
- K. V. Lombard. Female self-employment and demand for flexible, non-standard work schedules. *Economic Inquiry*, 2001.
- J. S. Long and J. Freese. *Regression Models for Categorical Dependent Variables using Stata*. Stata Press, 2006.
- E. Oster. Unobservable selection and coefficient stability: Theory and evidence. Working paper, National Bureau of Economic Research, 2013.
- P. R. Rosenbaum and D. B. Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1983.
- M. R. Rosenzweig and K. I. Wolpin. Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy*, 1980.
- H. A. Sackey. Female labour force participation in Ghana: The effects of education. Research paper, African Economic Research Consortium, 2005.
- A. Sen. Employment, institutions and technology: Some policy issues. *International Labour Review*, 1975.
- K. Train. *Discrete Choice Methods with Simulation: Second Edition*. Cambridge University Press, 2009.
- M. W. Warner, R. M. Al-Hassan, and J. G. Kydd. Beyond gender roles? Conceptualizing the social and economic lives of rural peoples in sub-Saharan Africa. *Development and Change*, 1997.
- A. J. Wellington. Self-employment: the new solution for balancing family and career? *Labour Economics*, 2006.
- C. Woodruff. Self-employment: Engine of growth or self-help safety net? In P. Paci and P. Serneels, editors, *Employment and Shared Growth: Rethinking the Role of Labor Mobility for Development*. World Bank, 2006.

A Formal Model for Job Flexibility

Starting with a basic model of time allocation, we suggest three possible ways of thinking about job flexibility and consider the implications for occupational choice. Whilst we do not formally test between these different models, it is useful to fix ideas to help interpret the results we present for the Ghana.

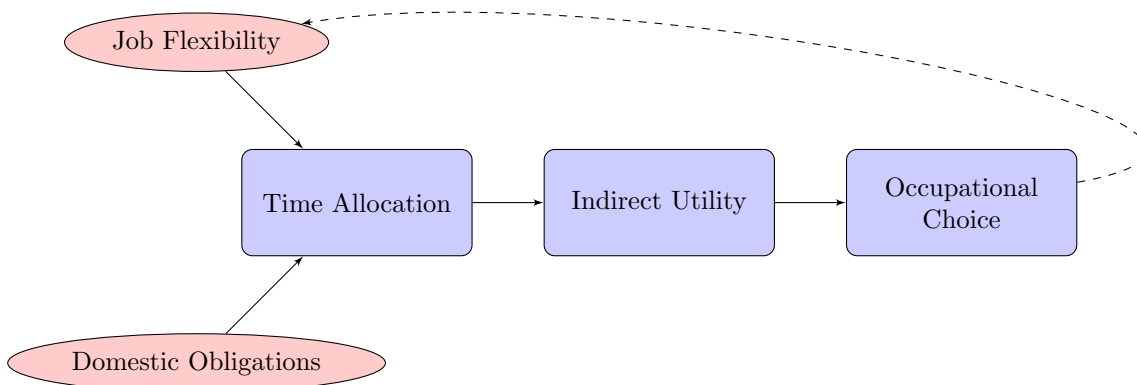
A.1 Basic Model

Our model of occupational choice has two stages:

1. First, we write down a simple model of time allocation. We add individual heterogeneity — in terms of domestic obligations — and job flexibility into the model by altering to the set-up of the constraints. This allows us to calculate indirect utility for each type of individual working in each different type of job.
2. Second, individuals choose the occupation that gives them the most welfare, *given* the results of their time allocation problem.

The overall structure of this framework is shown in Figure 5.

Figure 5: Time Allocation and Occupational Choice



In the time allocation problem, individuals derive utility from market goods — items bought from outside the household — and domestic goods — those goods provided within the household, such as

cooked meals, repairs of existing consumer durables, and care for children. These are labelled x_m and x_d respectively. Market and domestic goods are produced separately through market work (t_m), with return w , and domestic work (t_d), with return ρ . Individuals are only endowed with their time, which adds up to 1. For simplicity, we assume Cobb-Douglas preferences over domestic and market goods.

Taking these components of the model together, the maximisation problem for the individual can be written:

$$\max_{x_m, x_d} U = u(x_m, x_d) = x_m^\beta x_d^{(1-\beta)} \quad (15)$$

$$\text{subject to } t_m + t_d = 1 \quad (16)$$

$$x_m = wt_m \quad (17)$$

$$x_d = \rho t_d \quad (18)$$

$$t_k \geq 0 \quad \forall k = m, d. \quad (19)$$

We capture individual heterogeneity in terms of domestic obligations, by adding a lower bound on the domestic goods that must be provided.²⁰ This may be something of a simplification given that individuals, especially self-employed men, are somewhat willing to substitute between market and domestic work week-by-week. However, the time-horizon that is relevant to occupational choice is likely to be longer than a week. Thus, we believe this captures the imperative to provide domestic goods in a plausible way.

$$x_d \geq \gamma \quad (20)$$

As such, individuals with greater domestic obligations are potentially more constrained in their allocation of time and their choice of goods, which may leave them worse off. We now modify the set of constraints to consider the types of occupations that individuals with different domestic obligations would select.

²⁰Individual domestic obligations are conceptually distinct from household-level domestic obligations, because households' requirements for care and household chores may be divided unequally between certain household members.

A.2 Multi-Tasking

Certain types of self-employment activities may be undertaken concurrently with domestic work. For example, self-employed retailers may still be able to run their stall, whilst watching their children. We anticipate, however, that multi-tasking in this way comes at the expense of productivity in market work activities.

To formalise this, we add a parameter to the basic model, $0 < \pi \leq 1$, which captures these *two* effects of job flexibility. Firstly, we assume that π governs the extent to which domestic work draws down the endowment of time. As such, π measures the proportion of time spent on domestic work, which can simultaneously be devoted to market work. Secondly, we wish to incorporate the idea that more flexible jobs have lower returns. To do this, we write w as a positive linear function of π , such that $w = \hat{w}\pi$, where the parameter $\hat{w}$ measures how strongly job flexibility and market returns are associated.²¹ Since π is actually *lower* for jobs with higher flexibility, we think of π as ‘job rigidity’.

The initial time allocation problem for the individual may now be written:

$$\max_{x_m, x_d} U = u(x_m, x_d) = x_m^\beta x_d^{(1-\beta)} \quad (21)$$

$$\text{subject to } t_m + \pi t_d = 1 \quad (22)$$

$$x_m = \hat{w}\pi t_m \quad (23)$$

$$x_d = \rho t_d \quad (24)$$

$$t_k \geq 0 \quad \forall k = m, d. \quad (25)$$

$$x_d \geq \gamma \quad (26)$$

We begin by deriving analytical expressions for indirect utility in the presence of multi-tasking, then simulate the model for reasonable parameter values to illustrate its main predictions.

First, we write the Lagrangian for the time allocation problem, log linearising the utility function.

$$\mathcal{L} = \beta \ln(x_m) + (1 - \beta) \ln(x_d) - \lambda_1 \left[x_m + \left(\frac{\hat{w}\pi^2}{\rho} \right) x_d - \hat{w}\pi \right] + \lambda_2 [x_d - \gamma] \quad (27)$$

²¹The findings of the model are not sensitive to the assumption of a linear functional form.

The resulting Kuhn-Tucker conditions may then be written:

$$\frac{\partial \mathcal{L}}{\partial x_m} = \frac{\beta}{x_m} - \lambda_1 = 0 \quad (28)$$

$$\frac{\partial \mathcal{L}}{\partial x_d} = \frac{1-\beta}{x_d} - \lambda_1 \left(\frac{\hat{w}\pi^2}{\rho} \right) + \lambda_2 = 0 \quad (29)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = - \left[x_m + \left(\frac{\hat{w}\pi^2}{\rho} \right) x_d - \hat{w}\pi \right] = 0 \quad (30)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = x_d - \gamma = 0 \quad (31)$$

Given the monotonicity of the utility function, the full income constraint necessarily binds, such that $\lambda_1 > 0$. However, the resulting choices for x_m and x_d will depend on whether the domestic obligations constraint binds ($\lambda_2 > 0$) or is slack ($\lambda_2 = 0$). There are thus two scenarios to consider.

Scenario 1: $\lambda_2 = 0$

If $\lambda_2 = 0$ and the domestic obligations constraint is slack, the individual simply chooses x_m and x_d subject to the income constraint — this is a typical maximisation problem. The Marshallian demands that result are:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} \beta \hat{w}\pi \\ (1-\beta) \frac{\rho}{\pi} \end{pmatrix} \quad (32)$$

Scenario 2: $\lambda_2 > 0$

If $\lambda_2 > 0$ and the domestic obligations constraint binds, the consumption of domestic goods is fixed at γ . This, in turn, determines the amount of time left over for doing market work. Thus, the resulting consumptions may be written:

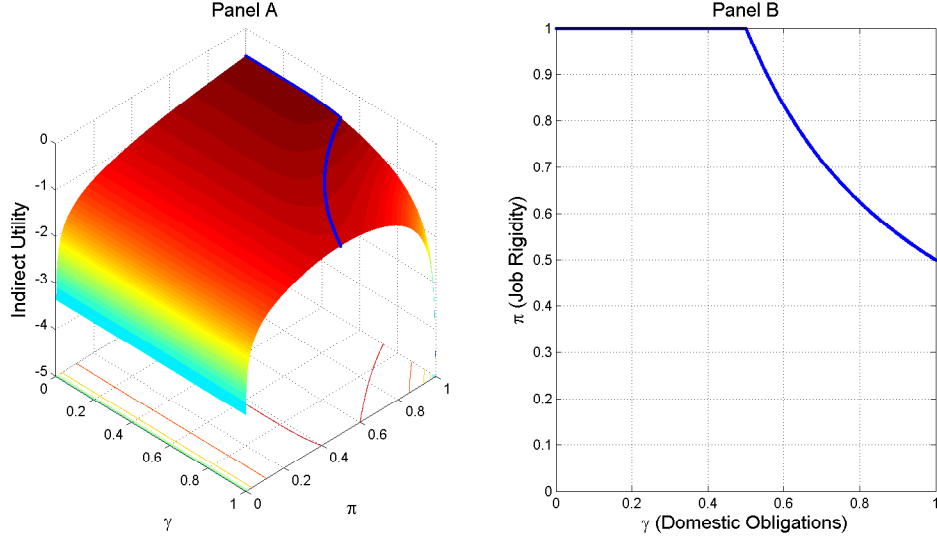
$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} \hat{w}\pi \left(1 - \frac{\pi\gamma}{\rho} \right) \\ \gamma \end{pmatrix} \quad (33)$$

The domestic obligations constraint will only bind if $(1 - \beta)\frac{\rho}{\pi} \leq \gamma$. Thus, we can write the indirect utility function as:

$$V(\beta, \pi, \gamma, \omega(\pi), \rho) = \begin{cases} \beta \ln[\beta \hat{w} \pi] + (1 - \beta) \ln[(1 - \beta)\frac{\rho}{\pi}] & \text{if } (1 - \beta)\frac{\rho}{\pi} > \gamma \\ \beta \ln\left[\hat{w} \pi \left(1 - \frac{\pi \gamma}{\rho}\right)\right] + (1 - \beta) \ln[\gamma] & \text{if } (1 - \beta)\frac{\rho}{\pi} \leq \gamma \end{cases} \quad (34)$$

We can then use this formulation of the indirect utility function to evaluate the welfare of individuals with different levels of domestic obligations in different jobs, allowing us to predict their occupational choice.

Figure 6: Occupational Selection in a Multi-Tasking Model



To demonstrate the main intuition of the model, we plot indirect utility for individuals with different levels of domestic obligations, γ , and job rigidity/flexibility, π . This is shown in Panel A of Figure 6.²² Indirect utility is, on average, decreasing in γ , as anticipated. This is because individuals with greater domestic obligations are more constrained in their time-use, and hence their consumption of goods. For individuals with low domestic obligations, indirect utility is always increasing in π . Intuitively, individuals with fewer domestic obligations are willing to endure extra job rigidity in pursuit of higher returns to market work. However, for individuals with high domestic obligations, indirect utility has an ‘inverted-U’ shape in π . At first, jobs with higher π raise welfare, because of the greater returns to market work. However, a higher level of π makes it more likely that the domestic obligations constraint

²²We set $\rho = \hat{w} = 1$ and $\beta = 0.7$. As shown above, β must be sufficiently high for the domestic obligations constraint to bind, and therefore affect indirect utility in the model.

will bind, which eventually reduces welfare. The higher the level of γ , the lower the level of job rigidity π needed for the domestic obligations constraint to start binding and bringing down indirect utility.

The shape of the indirect utility function, which results from our initial time allocation problem, also tells us the occupations that individuals with different levels of domestic obligations would choose. In particular, an individual with a given level of γ , will select the job type π that yields the highest indirect utility. This is shown by the blue ridge in Panel A of Figure 6, and then recast as the optimal job schedule in two dimensions in Panel B. This clarifies the relationship between γ and π — individuals with sufficiently high levels of domestic obligations will choose less rigid/more flexible jobs even if that means foregoing returns to market work.

A.3 Minimum Hours

Some jobs, particularly in the formal wage sector, may require individuals to work a minimum number of hours. Again, these types of less flexible jobs are also likely to generate the highest hourly earnings.

As before, we build this idea of minimum hours into our basic framework by changing the set of constraints in the model. Job flexibility once again has *two* effects on the time allocation problem, which we capture with a parameter τ . Firstly, τ places a lower bound on the time that must be spent doing market work in order to work in that job. However, we also incorporate the idea that jobs with minimum hours may have higher returns by relating w to τ . We adopt a simple linear functional form such that $w = k\tau$, where the parameter k captures the association between minimum hours and the earnings rate.²³ As such, τ captures *both* of the relevant aspects of job flexibility. Once again, by writing down the time allocation problem and finding out the indirect utility for different values of γ and τ , we can recover the types of jobs that individuals with different domestic obligations would prefer. We also understand τ in terms of ‘job rigidity’, since a *greater* value of τ implies a *less* flexible job.

The maximisation problem may now be written:

²³Once again, the main intuition of the model is robust to different functional form assumptions about this relationship.

$$\max_{x_m, x_d} U = u(x_m, x_d) = x_m^\beta x_d^{(1-\beta)} \quad (35)$$

$$\text{subject to } t_m + t_d = 1 \quad (36)$$

$$x_m = k\tau t_m \text{ if } t_m \geq \tau \quad (37)$$

$$x_d = \rho t_d \quad (38)$$

$$t_k \geq 0 \quad \forall k = m, d. \quad (39)$$

$$x_d \geq \gamma \quad (40)$$

$$(41)$$

To derive indirect utility in the minimum hours model, we once again begin by writing out the Lagrangian.

$$\mathcal{L} = \beta \ln(x_m) + (1 - \beta) \ln(x_d) - \lambda_1 \left[x_m + \left(\frac{k\tau}{\rho} \right) x_d - k\tau \right] + \lambda_2 \left[x_d - \gamma \right] + \lambda_3 \left[\frac{x_m}{k\tau} - \tau \right] \quad (42)$$

The Kuhn-Tucker conditions may then be written:

$$\frac{\partial \mathcal{L}}{\partial x_m} = \frac{\beta}{x_m} - \lambda_1 + \lambda_3 = 0 \quad (43)$$

$$\frac{\partial \mathcal{L}}{\partial x_d} = \frac{1 - \beta}{x_d} - \lambda_1 \left(\frac{k\tau}{\rho} \right) + \lambda_2 = 0 \quad (44)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = - \left[x_m + \left(\frac{k\tau}{\rho} \right) x_d - k\tau \right] = 0 \quad (45)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = x_d - \gamma = 0 \quad (46)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_3} = \frac{x_m}{k\tau} - \tau = 0 \quad (47)$$

Since the utility function is monotonic, we know that the income constraint binds and $\lambda_1 > 0$. We also know that it is not possible for both the domestic obligations constraint and the lower bound on market work to simultaneously bind, so it cannot be that $\lambda_2 > 0$ and $\lambda_3 > 0$. Thus we are left with three potential scenarios to consider.

Scenario 1: $\lambda_2 = \lambda_3 = 0$

If both the domestic obligations constraint and the lower bound on market work are slack, then the individual simply maximises utility subject to the income constraint. This produces regular Marshallian demands of the form:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} \beta k \tau \\ (1 - \beta) \rho \end{pmatrix} \quad (48)$$

This situation will prevail if, both $\beta > \tau$ and $(1 - \beta) \rho > \gamma$, as this implies both extra constraints are too small to affect time allocation and consumption patterns.

Scenario 2: $\lambda_2 > 0; \lambda_3 = 0$

In this scenario, both the domestic obligations constraint and the income constraint bind. This occurs, if $\beta > \tau$ and $(1 - \beta) \rho \leq \gamma$. The resulting outcomes for market and domestic goods are:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} k \tau \left(1 - \frac{\gamma}{\rho} \right) \\ \gamma \end{pmatrix} \quad (49)$$

Scenario 3: $\lambda_2 = 0; \lambda_3 > 0$

Finally, it may be that the income constraint and the lower bound on market work hours bind. This will happen if $\beta \leq \tau$ and $(1 - \beta) \rho > \gamma$. The resulting consumption outcomes are:

$$\begin{pmatrix} x_m^* \\ x_d^* \end{pmatrix} = \begin{pmatrix} k \tau^2 \\ \rho(1 - \tau) \end{pmatrix} \quad (50)$$

Thus, we can now write down the indirect utility function, taking these three potential scenarios into account, as in Equation (51).

$$V(\beta, \tau, \gamma, k, \rho) = \begin{cases} \beta \ln [\beta k \tau] + (1 - \beta) \ln [(1 - \beta) \rho] & \text{if } \tau \text{ and } (1 - \beta) \rho > \gamma \text{ and } \beta > \tau \\ \beta \ln \left[k \tau \left(1 - \frac{\gamma}{\rho} \right) \right] + (1 - \beta) \ln [\gamma] & \text{if } \tau \text{ and } (1 - \beta) \rho \leq \gamma \text{ and } \beta > \tau \\ \beta \ln [k \tau^2] + (1 - \beta) \ln [\rho(1 - \tau)] & \text{if } \tau \text{ and } (1 - \beta) \rho > \gamma \text{ and } \beta \leq \tau \end{cases} \quad (51)$$

Plotting the indirect utility function for possible values of γ and τ allows us to show the link domestic obligations to occupational selection more intuitively.²⁴ This is shown in Panel A of Figure 7. The surface only goes up to the line $\tau = 1 - \frac{\gamma}{\rho}$, because it is impossible for individuals with high levels of γ to choose jobs with a high level of τ — the domestic obligations constraint and the lower bound on minimum hours could not be met simultaneously. Aside from this restriction at $\tau = 1 - \frac{\gamma}{\rho}$, indirect utility has an ‘inverted-U’ shape in job rigidity τ , for *all* levels of domestic obligations γ . At first, individuals are better off with higher τ as this increases their returns to market work, and hence their potential consumption of market goods. However, if τ becomes too high, the returns to market work are not sufficient to compensate individuals for the restricted choice they have over their allocation of time, and hence mix of consumption of market and domestic goods. Thus, if τ becomes too high, extra job rigidity begins to reduce welfare.

Individuals select the occupation that maximises their indirect utility, after having solved the time allocation problem. The optimum level of τ for each individual is shown with the blue ridge in Panel A of Figure 7. This is plotted in two dimensions in Panel B. Individuals with low domestic obligations are unconstrained in their occupational choice, so they pick the level of τ at the top of the ‘inverted-U’. However, for individuals with higher levels of γ , they simply select the occupation with the highest level of τ , in which they can still meet their domestic obligations. This runs along the line $\tau = 1 - \frac{\gamma}{\rho}$.

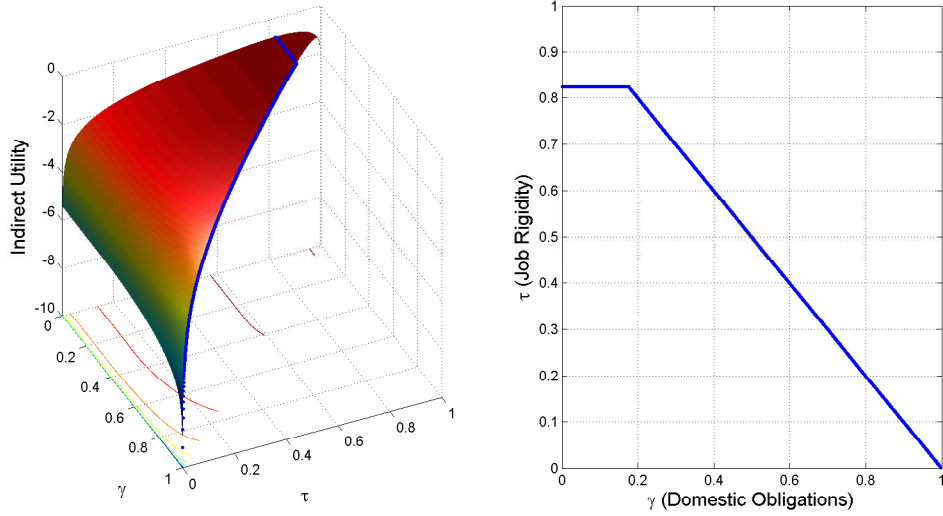
A.4 Adjustment Costs

Domestic obligations may be imposed stochastically if, for example, certain household members suddenly become ill or domestic appliances suddenly fail. It may be easier to adjust to these shocks in some jobs than in others. In wage jobs, the costs of making these adjustments to market work may result in greater costs than in self-employment, because working hours may be explicitly outlined by formal contracts. Again, though, hourly earnings are likely to be greater in jobs where these adjustment costs are higher.

To recast the basic model in terms of adjustment costs, we begin by making the simplifying assump-

²⁴As before, we set $\beta = 0.7$ and $\rho = k = 1$.

Figure 7: Occupational Selection in a Minimum Work Hours Model



tion that individuals derive utility only from market goods, $x_m = wt_m$. However, the actual number of hours they will be able to work in the market is constrained according to some stochastic parameter which, maintaining notation, we label γ .²⁵ The parameters of the *ex ante* distribution of γ are labelled θ . Vitally, the individual chooses the amount of market work they wish to undertake, $\hat{t}_m$, *before* the shock to domestic obligations is realised. In the context of wage-employment, we can think of $\hat{t}_m$ as the ‘contracted’ number of hours. The analogous interpretation for the self-employed may be the number of hours to which the individual has, in some sense, committed prior to working, due to relationships with suppliers, customers, and competitors. As such, the initial problem is now one of choosing the optimal level of $\hat{t}_m$, rather than allocating time between market and domestic work per se.

Once again, job flexibility, which we capture with the parameter c , has *two* related components. Firstly, departure from the pre-committed level of market work incurs some convex adjustment penalty to wages, the size of which is determined by c . However, c also determines the wage rate that would be paid, *absent* any difference between $\hat{t}_m$ and t_m , which we label w_0 . We make the simplifying assumption that w_0 and c are linearly associated through some parameter h , such that $w_0 = hc$. Thus c relates to both the returns to market work and the costs of adjustment when plans need to change.

Making the assumption of log utility, we can write the modified individual maximisation problem.²⁶

²⁵We have implicitly made the normalisation $\rho = 1$.

²⁶We want to set up the model so that individuals are risk averse.

$$\max_{\hat{t}_m} E[U] = E[\ln(x_m)] = E[\ln(wt_m)] \quad (52)$$

$$\text{subject to } t_m \leq 1 - \gamma \quad (53)$$

$$\gamma \sim f(\gamma; \theta) \quad (54)$$

$$w = w_0 \left(1 - \frac{c}{2} (\hat{t}_m - t_m)^2 \right) = hc \left(1 - \frac{c}{2} (\hat{t}_m - t_m)^2 \right) \quad (55)$$

We can write the problem more succinctly by substituting the constraints into the maximand. This reinforces the idea that the initial problem is now one of choosing a ‘contracted’ number of hours, $\hat{t}_m$, rather than time allocation per se. After solving this problem, individuals with different domestic obligations, parameterised by θ , can then choose the level of c which gives them the highest level of expected (indirect) utility.

$$\max_{\hat{t}_m} E[U] = \int_{-\infty}^{\infty} \left[\ln \left(hc \left(1 - \frac{c}{2} (\hat{t}_m - t_m)^2 \right) t_m \right) \right] f(\gamma; \theta) d\gamma \quad (56)$$

In order to further the analytical treatment of this problem, we suggest a functional form for the distribution of γ . We assume that there are just two states of the world. In the ‘bad’ state, which occurs with probability q , the individual faces a lower bound on domestic work of $\tilde{\gamma}$.²⁷ In the ‘good’ state of the world, however, the individual devotes no time to domestic work, such that $\gamma = 0$. We assume that the individual has perfect information about their own values of $\tilde{\gamma}$ and q . Summarising this formulation, the distribution of γ may be written:

$$\gamma(q, \tilde{\gamma}) = \begin{cases} \tilde{\gamma} & \text{w.p. } q \\ 0 & \text{w.p. } (1 - q) \end{cases} \quad (57)$$

As such, individuals’ susceptibility to shocks to domestic obligations is captured by q and $\tilde{\gamma}$.

To solve the model, we note that the problem effectively spans two periods. Whilst the level of $\hat{t}_m$ is chosen *before* the value of γ has been realised, the individual chooses t_m *after*. Therefore, in Period 1, the individual chooses their contracted hours, $\hat{t}_m$, on the basis of predictions about domestic obligations

²⁷We assume $0 \leq q \leq 1$ and $0 \leq \tilde{\gamma} < 1$.

(q and $\tilde{\gamma}$), and the characteristics of their job, c . Then, in Period 2, the individual chooses the level of t_m to maximise consumption $x_m = wt_m$ conditional on $\hat{t}_m$, as well as the revealed value of γ .

We solve the model using backwards induction. First, we find a ‘reaction function’ for the optimal level of t_m given the parameters $\Gamma = \{\hat{t}_m, q, \tilde{\gamma}, h, c\}$. We then substitute $t_m(\Gamma)$ back into the original problem, of the form in Equation (56) to solve for the optimum $\hat{t}_m^*$ and hence the expected utility, for different levels of the job characteristic and domestic obligation parameters.

To find this reaction function, we write the maximisation problem that faces the individual in Period 2. We recall that, by Period 2, the level of γ has been realised as either $\tilde{\gamma}$ or 0, and $\hat{t}_m$ has been chosen.

$$\max_{t_m} x_m = t_m hc \left(1 - \frac{c}{2} (t_m - \hat{t}_m)^2 \right) \quad (58)$$

$$\text{subject to } t_m \leq 1 - \gamma \quad (59)$$

The Lagrangian for the problem is:

$$\mathcal{L} = t_m hc \left(1 - \frac{c}{2} (t_m - \hat{t}_m)^2 \right) - \lambda [t_m - 1 + \gamma] \quad (60)$$

We can then write the resulting Kuhn-Tucker conditions.

$$\frac{\partial \mathcal{L}}{\partial t_m} = hc - \frac{hc^2}{2} \left(3t_m^2 - 4t_m \hat{t}_m + \hat{t}_m^2 \right) - \lambda = 0 \quad (61)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -[t_m - 1 + \gamma] = 0 \quad (62)$$

As such, there are two possible scenarios depending on whether or not the constraint binds. If the constraint does in fact bind, and $\lambda > 0$, the optimal choice of t_m is simply $1 - \gamma$. However, if the constraint does not bind, we use the quadratic formula to find the optimal choice of t_m , which we label t_m^* , as a function of $\hat{t}_m$ and c .

$$t_m(\Gamma) = \begin{cases} 1 - \gamma & \text{if } t_m^* \geq 1 - \gamma \\ t_m^* = \frac{2ct_m + \sqrt{c^2 t_m^2 + 6c}}{3c} & \text{otherwise} \end{cases} \quad (63)$$

We can now substitute the reaction function $t_m(\Gamma)$ back into the Period 1 objective function. This allows us to write an unconstrained maximisation, which the individual solves to choose their preferred contracted hours, $\hat{t}_m$, given the parameters of the model.

$$\begin{aligned} \max_{\hat{t}_m} E[U] = & q \ln \left[hc \left(1 - \frac{c}{2} (\hat{t}_m - t_m(\Gamma|\gamma = \tilde{\gamma}))^2 \right) (t_m(\Gamma|\gamma = \tilde{\gamma})) \right] \\ & + (1 - q) \ln \left[hc \left(1 - \frac{c}{2} (\hat{t}_m - t_m(\Gamma|\gamma = 0))^2 \right) (t_m(\Gamma|\gamma = 0)) \right] \end{aligned} \quad (64)$$

Once we solve for the optimal $\hat{t}_m$, we can recover the expected utility from different jobs (c) for individuals with different domestic obligations, ($\tilde{\gamma}$ and q). Insofar as individuals select the jobs which give highest welfare ex ante, this enables us examine occupational choice. Since solving the model analytically is not straightforward, we use the MATLAB program `fmincon` to maximise Equation (64) with respect to $\hat{t}_m$, and calculate expected utility.

We plot expected utility for individuals with different levels of domestic obligations and different jobs in Panel A of Figure 8. In particular, we focus on heterogeneity in $\tilde{\gamma}$.²⁸ We hold q constant, setting $q = 0.5$ and $h = 1$, and simulating the model for all possible values of $\tilde{\gamma}$.²⁹

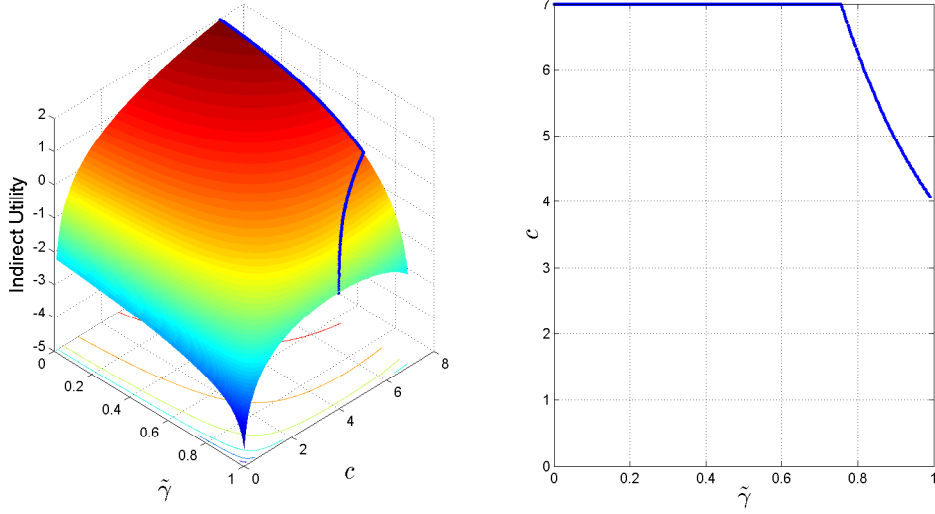
Looking at the shape of the surface in Panel A, we can see that expected utility is everywhere decreasing in $\tilde{\gamma}$, as expected. For low values of $\tilde{\gamma}$, expected utility is increasing in c , as this corresponds to higher returns to returns to market work, and hence higher consumption. When domestic obligations are larger, however, expected utility is at first increasing but then *decreasing* in c . Individuals face more variation in their outcomes when $\tilde{\gamma}$ is higher.³⁰ The extra adjustment penalties that come with an increase in c therefore make the risk averse individuals in the model worse off, in expectation.

²⁸The implications of changing q and holding $\tilde{\gamma}$ constant are somewhat more complex. Overall, it emerges that the optimal level of job rigidity, c , is a non-monotonic function of q .

²⁹In order to maintain positive consumption levels, consistent with log utility, we ensure $c < \frac{2}{(t_m - \hat{t}_m)^2}$. We show the results for $0 < c \leq 7$, since as c becomes too large, the model is insoluble for all values of $\tilde{\gamma}$.

³⁰In particular, both the mean and the variance of γ are increasing in $\tilde{\gamma}$. $E(\gamma) = q\tilde{\gamma}$, therefore $\frac{\partial E(\gamma)}{\partial \tilde{\gamma}} = q \geq 0$. $\text{Var}(\gamma) = q\tilde{\gamma}^2 - q^2\tilde{\gamma}^2$, therefore $\frac{\partial \text{Var}(\gamma)}{\partial \tilde{\gamma}} = 2\tilde{\gamma}q(1 - q) \geq 0$.

Figure 8: Occupational Selection in an Adjustment Costs Model — Changing $\tilde{\gamma}$



The expected utility function translates into a model of occupational choice, insofar as individuals with different levels of domestic obligations choose the level of c that maximises their (expected) welfare. This is shown by the blue ridge in Panel A, and then plotted in two dimensions in Panel B of Figure 8. Individuals facing a higher potential shock to domestic obligations, $\tilde{\gamma}$, have a lower optimal level of job rigidity, c . When $\tilde{\gamma}$ is low, the expected involuntary adjustment of t_m away from $\hat{t}_m$ is also low. Thus, individuals will expose themselves to higher adjustment penalties, in pursuit of higher initial wages, hc . The converse is true when $\tilde{\gamma}$ is high, and individuals' expected involuntary adjustments of t_m away from $\hat{t}_m$ are also high. It is then optimal to choose occupations with lower c , even if this means foregoing higher initial wages hc .

A.5 Theoretical Predictions

In each of the three stories outlined above, there is some range of the parameter values for which individuals with greater domestic obligations optimally choose occupations with more flexibility, even at the expense of reduced hourly earnings in market work. Primarily, we want to test this relationship by examining whether individuals from households with a greater proportion of dependent members choose low-input self-employment activities instead of choosing high-input self-employment activities or wage-employment. Writing down the three stories formally helps to fix ideas and to think about which interpretations might be more important in Ghana.

B Summary Statistics

Table 10: Sample Location

Region	Location					
	Rural		Urban		Total	
	N	%	N	%	N	%
Western	1138	5.51	655	3.17	1793	8.68
Central	860	4.16	517	2.50	1377	6.67
Greater Accra	278	1.35	2502	12.12	2780	13.46
Volta	1151	5.57	442	2.14	1593	7.71
Eastern	1272	6.16	671	3.25	1943	9.41
Ashanti	1760	8.52	1695	8.21	3455	16.73
Brong Ahafo	1112	5.38	678	3.28	1790	8.67
Northern	1818	8.80	549	2.66	2367	11.46
Upper East	1560	7.55	230	1.11	1790	8.67
Upper West	1616	7.83	147	0.71	1763	8.54
Total	12565	60.84	8086	39.16	20651	100.00

Sample of individuals of working age (15–65)

Table 11: Summary Statistics (Females)

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
Age (Years)	10926	33.22	13.56	15.00	22.00	31.00	43.00	65.00
Education (Years)	10926	4.86	4.90	0.00	0.00	5.00	9.00	18.00
Tenure (Months)	10926	64.22	107.97	0.00	0.00	0.00	84.00	540.00
Married? (1=Y, 0=N)	10926	0.49	0.50	0.00	0.00	0.00	1.00	1.00
Household Size	10926	5.93	3.43	1.00	4.00	5.00	7.00	29.00
Dependency Ratio	10926	0.85	0.75	0.00	0.33	0.67	1.00	7.00
No. Infants (< 2 years) in HH	10926	0.28	0.52	0.00	0.00	0.00	1.00	4.00
No. Young Children (2–4 years) in HH	10926	0.44	0.66	0.00	0.00	0.00	1.00	5.00
No. Older Children (5–14 years) in HH	10926	1.62	1.63	0.00	0.00	1.00	2.00	12.00
No. Elders (> 65 years) in HH	10926	0.16	0.41	0.00	0.00	0.00	0.00	5.00
HH Land (Acres)	10926	8.60	33.60	0.00	0.00	0.00	6.00	320.00
HH Owns Land? (1=Y, 0=N)	10926	0.47	0.50	0.00	0.00	0.00	1.00	1.00
Farmer Father? (1=Y, 0=N)	10926	0.58	0.49	0.00	0.00	1.00	1.00	1.00
Professional Father? (1=Y, 0=N)	10926	0.05	0.21	0.00	0.00	0.00	0.00	1.00
Service Sector Father? (1=Y, 0=N)	10926	0.10	0.31	0.00	0.00	0.00	0.00	1.00
Farmer Mother? (1=Y, 0=N)	10926	0.50	0.50	0.00	0.00	0.00	1.00	1.00
Professional Mother? (1=Y, 0=N)	10926	0.01	0.10	0.00	0.00	0.00	0.00	1.00
Service Sector Mother? (1=Y, 0=N)	10926	0.02	0.14	0.00	0.00	0.00	0.00	1.00
HH Unearned Income? (1=Y, 0=N)	10926	0.01	0.09	0.00	0.00	0.00	0.00	1.00
Observations	10926							

Sample of individuals of working age (15–65)

Table 12: Summary Statistics (Males)

	N	Mean	S.Dev.	Min.	25th P.tile	Median	75th P.tile	Max.
Age (Years)	9725	32.58	13.85	15.00	20.00	30.00	43.00	65.00
Education (Years)	9725	6.69	5.08	0.00	0.00	8.00	10.00	18.00
Tenure (Months)	9725	100.21	127.97	0.00	0.00	40.00	180.00	540.00
Married? (1=Y, 0=N)	9725	0.44	0.50	0.00	0.00	0.00	1.00	1.00
Household Size	9725	5.70	3.46	1.00	3.00	5.00	7.00	29.00
Dependency Ratio	9725	0.69	0.66	0.00	0.17	0.50	1.00	7.00
No. Infants (< 2 years) in HH	9725	0.24	0.48	0.00	0.00	0.00	0.00	4.00
No. Young Children (2–4 years) in HH	9725	0.39	0.63	0.00	0.00	0.00	1.00	5.00
No. Older Children (5–14 years) in HH	9725	1.49	1.58	0.00	0.00	1.00	2.00	12.00
No. Elders (> 65 years) in HH	9725	0.12	0.37	0.00	0.00	0.00	0.00	5.00
HH Land (Acres)	9725	8.20	31.43	0.00	0.00	0.00	6.00	320.00
HH Owns Land? (1=Y, 0=N)	9725	0.48	0.50	0.00	0.00	0.00	1.00	1.00
Farmer Father? (1=Y, 0=N)	9725	0.50	0.50	0.00	0.00	1.00	1.00	1.00
Professional Father? (1=Y, 0=N)	9725	0.05	0.22	0.00	0.00	0.00	0.00	1.00
Service Sector Father? (1=Y, 0=N)	9725	0.10	0.30	0.00	0.00	0.00	0.00	1.00
Farmer Mother? (1=Y, 0=N)	9725	0.42	0.49	0.00	0.00	0.00	1.00	1.00
Professional Mother? (1=Y, 0=N)	9725	0.01	0.10	0.00	0.00	0.00	0.00	1.00
Service Sector Mother? (1=Y, 0=N)	9725	0.02	0.14	0.00	0.00	0.00	0.00	1.00
HH Unearned Income? (1=Y, 0=N)	9725	0.02	0.12	0.00	0.00	0.00	0.00	1.00
Observations	9725							

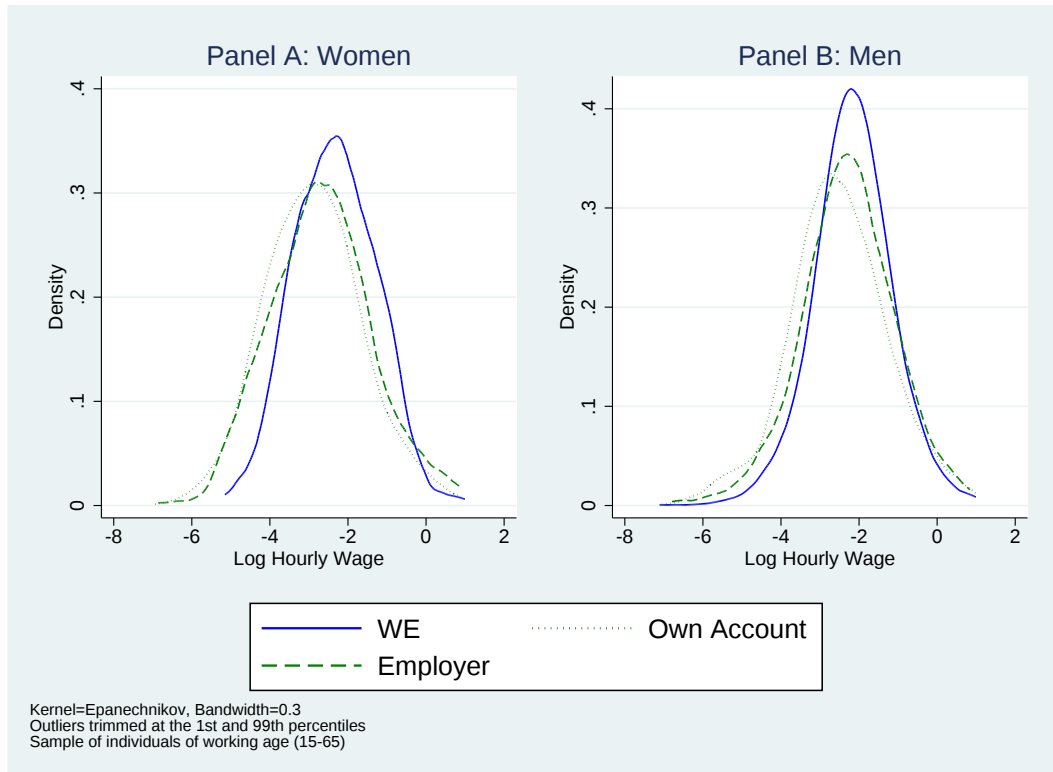
Sample of individuals of working age (15–65)

Table 13: Summary Statistics by Sex and Occupation

	Wage Employed				Non-Farm SE (Own Account)				Non-Farm SE (Employer)			
	Male		Female		Male		Female		Male		Female	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Age (Years)	37.85	36.00	34.36	32.00	37.94	37.00	37.21	36.00	39.15	38.00	40.12	40.00
Education (Years)	10.47	10.00	10.62	11.00	7.29	9.00	5.03	5.00	7.44	9.00	4.58	3.00
Tenure (Months)	112.81	72.00	89.98	43.00	141.52	108.00	105.01	63.50	153.99	120.00	146.53	120.00
Married? (1=Y, 0=N)	0.57	1.00	0.41	0.00	0.61	1.00	0.63	1.00	0.73	1.00	0.67	1.00
Household Size	4.22	4.00	4.52	4.00	4.53	4.00	5.23	5.00	5.19	5.00	6.15	6.00
Dependency Ratio	0.52	0.33	0.64	0.50	0.70	0.50	0.96	1.00	0.78	0.67	0.92	0.75
No. Infants (< 2 years) in HH	0.19	0.00	0.17	0.00	0.25	0.00	0.28	0.00	0.29	0.00	0.28	0.00
No. Young Children (2–4 years) in HH	0.27	0.00	0.27	0.00	0.38	0.00	0.46	0.00	0.49	0.00	0.41	0.00
No. Older Children (5–14 years) in HH	0.97	0.00	1.06	1.00	1.18	1.00	1.52	1.00	1.43	1.00	1.88	2.00
No. Elders (> 65 years) in HH	0.04	0.00	0.08	0.00	0.07	0.00	0.12	0.00	0.07	0.00	0.14	0.00
HH Land (Acres)	5.24	0.00	2.71	0.00	6.10	0.00	8.06	0.00	6.09	0.00	7.05	0.00
HH Owns Land? (1=Y, 0=N)	0.27	0.00	0.19	0.00	0.39	0.00	0.40	0.00	0.42	0.00	0.43	0.00
Farmer Father? (1=Y, 0=N)	0.46	0.00	0.34	0.00	0.59	1.00	0.62	1.00	0.60	1.00	0.65	1.00
Professional Father? (1=Y, 0=N)	0.10	0.00	0.13	0.00	0.06	0.00	0.06	0.00	0.06	0.00	0.06	0.00
Service Sector Father? (1=Y, 0=N)	0.19	0.00	0.23	0.00	0.13	0.00	0.14	0.00	0.16	0.00	0.14	0.00
Farmer Mother? (1=Y, 0=N)	0.38	0.00	0.28	0.00	0.50	0.00	0.53	1.00	0.51	1.00	0.56	1.00
Professional Mother? (1=Y, 0=N)	0.02	0.00	0.04	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Service Sector Mother? (1=Y, 0=N)	0.05	0.00	0.06	0.00	0.04	0.00	0.03	0.00	0.04	0.00	0.03	0.00
HH Unearned Income? (1=Y, 0=N)	0.03	0.00	0.02	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.01	0.00
Observations	1768		658		813		2640		526		807	

Sample of individuals of working age (15–65)

Figure 9: Log of Hourly Earnings by Sex



C Closed-Form Log-Likelihood Function

In order to estimate the log-likelihood function without simulation, we need to be able to calculate the probability for each outcome, $\Pr(y_i = j \mid \mathbf{x}_i, D_i)$. For a general error variance-covariance matrix, this would require evaluating a complicated multidimensional integral. However, since the distribution of $(\zeta_{i2'}, \zeta_{i3'}, \dots, \zeta_{iJ'}) \mid v_i$ is *exchangeable* (or characterised by compound symmetry), we can use the result due to Dunnett (1989) to reduce this multidimensional integral to one dimension. We can thus write the probability that individual i chooses k as:³¹

$$\Pr(y_i = k \mid \mathbf{x}_i, D_i) = \frac{1}{\sqrt{\pi}} \int_0^\infty \left\{ \prod_{j=1}^{J-1} \Phi\left(\frac{-[z\sqrt{2}\sqrt{1-\rho^2}] - \lambda_{ij}}{\rho}\right) + \prod_{j=1}^{J-1} \Phi\left(\frac{[z\sqrt{2}\sqrt{1-\rho^2}] - \lambda_{ij}}{\rho}\right) \right\} e^{-z^2} dz \quad (65)$$

Empirically, we substitute for λ_{ij} using:

$$\lambda_{ij} = \mathbf{x}_i^\top \boldsymbol{\beta}_j + \psi_j D_i + \rho \sigma^{-1} (D_i - \mathbf{x}_i' \boldsymbol{\pi}) \quad (66)$$

Substituting Equation (65) into Equation (13) results in a closed-form for the log likelihood.

³¹Following the help file for Stata's `mprobit` command, the Dunnett (1989) result can be approximated using Gaussian quadrature.