

Who are the Self-employed? A New Empirical Approach

Sarah Brown^a, Lisa Farrell^b and Mark N. Harris^c

^aDepartment of Economics, University of Leicester, U.K.

^bDepartment of Economics, University of Melbourne, Australia.

^cDepartment of Econometrics and Business Statistics,
Monash University, Australia.

September 8, 2004

Abstract

Modelling the incidence of self-employment has proved problematic. Whilst the individual supply side characteristics of the self-employed are well documented, we argue that the literature has largely misspecified demand side aspects. We present results from an econometric framework that allows us to separately, and simultaneously, model the supply and demand side characteristics that determine employment types. We show that whilst individual characteristics are important determinants of self-employment, there are also important factors specific to the type of employment that influence whether an individual is self-employed. Our results suggest that individuals may be captive to a particular employment type because of the sector in which they work, the number of hours they prefer to work and their ethnicity.

Word count: 5173

Keywords: Self-employment; Discrete Choice Models; Logit; OLS

JEL classification: J23 J33 C25 C10

Acknowledgements: We acknowledge financial support from the University of Melbourne and the University of Leicester. We are grateful to the Data Archive at the University of Essex for supplying the Family Expenditure Surveys, 1997 to 2000. We wish to thank Tim Fry for invaluable advice and participants at the Royal Economic Society Conference 2003, European Meeting of the Econometric Society 2003 and the Econometrics Society Australasian Meeting 2003 for helpful comments.

1 Introduction

Recent years have seen the self-employed emerge as an important group of workers in many developed countries, accounting for 16% of the workforce in Australia, 10% in Canada, 9% in the UK and 9% in the US in 1999 (Le 1999). The increasing proportion of self-employed individuals in the labour force has led to an interest amongst economists and policy makers in the characteristics of the self-employed due to its potential role in alleviating poverty and unemployment. Self-employment may be regarded as a way out of unemployment and/or the creation of small businesses may serve to create further employment. Yet modelling the incidence of self-employment has proved problematic. Whilst the individual supply side characteristics of the self-employed are well documented, we argue that the literature to date has largely misspecified the demand side aspects. In this paper we present results from an econometric framework that allows us to separately, and simultaneously, model supply and demand side characteristics.

This paper focuses on the factors that determine the type of individual who is self-employed. The literature concerning the profile of the self-employed has, in the main, looked at the characteristics of individuals who are self-employed compared to fixed wage workers. We extend this approach by additionally considering workers employed on performance related pay (PRP) contracts. Such contracts allow some proportion of remuneration to be conditional upon a measure of performance and, hence, represent an interesting intermediate case between fixed wage employment and self-employment.

We investigate the possibility that individuals may not be strictly free to

choose their preferred type of employment, but may be captive to particular types of employment due to a number of demand side factors which characterise heterogeneity of employment types. We apply a new estimator, the Parameterised Dogit Ordered Generalized Extreme Value (DOGEV) model, that allows for ordering in the observed employment types and for captivity within an observed employment type. To be specific, this estimator allows us to distinguish between the differential effects of factors which lead to individual heterogeneity (i.e. supply side influences) and those which lead to employment type heterogeneity (i.e. demand side influences). Thus we are able to determine the impact of supply side factors on the probability of an individual being self-employed whilst also controlling for demand side effects.

2 Background

A number of approaches have been developed to explain the determinants of self-employment, emphasising to varying degrees sociological, psychological and economic factors. The standard economic argument is that individuals decide whether or not to enter self-employment on the basis of the relative utilities on offer. Such an approach encapsulates unemployment push and pull factors, with displaced workers being pushed or pulled into self-employment by supply side considerations (Taylor 1996).

Recent research has focused on the attributes of the self-employed in order to determine who are the self-employed concentrating on characteristics such as gender, ethnicity and family background (Le 1999). In general, studies specify a reduced form Probit or Logit equation of self-employment whereby the vector of explanatory variables contains a combination of personal and

labour market characteristics.

Turning initially to personal characteristics, marital status has been incorporated into many empirical studies. As argued by Le (1999), marriage is assumed to represent stability and, as such, may provide a suitable background for risky self-employment. Blanchflower and Oswald (1998) and Bernhardt (1994) find that having a working spouse enhances the probability of self-employment. Similarly, Schiller and Crewson (1997) find that a husband's employment increases the probability that his wife will be self-employed.

An individual's age may affect his/her propensity to become self-employed (see, Calvo and Wellisz 1980, Kidd 1993). For instance, age may act as a proxy to capture the effects of an individual's awareness, knowledge and experience in the labour market thereby reflecting general human capital. Alternatively, as an individual becomes older, he/she may have accumulated the financial resources required for self-employment. Detailed analyses of the importance of capital constraints for the probability of becoming self-employed have been conducted by Blanchflower and Oswald (1998), Dunn and Holtz-Eakin (2000) and Johansson (2000). Generally, studies have reported a non-linear relationship between self-employment and age (Rees and Shah 1986).

The role of ethnicity in determining the propensity to become self-employed has also attracted a great deal of attention. The issue of whether discrimination bars employment in certain sectors has been the subject of much debate (Hout and Rosen 2000). Rees and Shah (1986) find that non-white individuals in the UK have a lower propensity to become self-employed whilst Brock and Evans (1986) find the reverse in the US. More recent evidence reported

by Clark and Drinkwater (2000) indicates higher rates of self-employment amongst ethnic minorities in England and Wales.

The role of educational qualifications has been incorporated into many empirical studies. Educational attainment may act as a proxy for ability. Studies reporting a positive relationship between educational attainment and the probability of self-employment, include Rees and Shah (1986), Borjas (1986), Borjas and Bronars (1989) and Evans and Leighton (1989). Alternatively, higher levels of educational attainment may play a signalling role in the labour market with high educational qualifications serving to secure employment in the non self-employed sector. Evidence supporting an inverse relationship between higher levels of educational qualifications and the propensity to become self-employed include Evans (1989), de Wit and Winden (1989) and Kidd (1993).

Finally, Blanchflower and Oswald (1998) argue that the psychological factors associated with entrepreneurial drive influence the propensity to become self-employed. Evans and Leighton (1989) focus on one such psychological factor – willingness to be in charge of one’s own destiny – and find that this is positively correlated with self-employment. A related point concerns an individual’s preference for hours of work. In general, fixed wage employment entails accepting the working week dictated by the employer (although in some cases an employee may be able to exercise some discretion, albeit within the parameters specified by the employer, for example flexi-time schemes). In contrast, a self-employee can determine the number of hours of work, presumably to satisfy their entrepreneurial drive, thereby exerting more control over their destiny.

To summarise, the standard approach adopted to profile the self-employed entails specifying a Probit or Logit model to predict the probability of self-employment. We extend the existing literature by allowing for three types of employment namely; fixed wage employment, PRP and self-employment. The inclusion of the intermediate category allows for the fact that, in reality, the choice between self-employment and non self-employment, is not as stark as that depicted in the existing literature. Moreover, the existing literature allows for supply side and demand side influences, but these two influences have been treated identically empirically. We propose an econometric model that distinguishes between these two influences and therefore allows demand and supply side effects to impact differentially on the probability of self-employment.

3 The Econometric Model

3.1 The DOGEV Model

A new discrete model, the DOGEV model, proposed by Fry and Harris (2004) forms the basis of our econometric analysis. It is based upon separate, independent, generalisations of the Logit model for multiple outcomes. One of the generalisations embodies the Ordered Generalized Extreme Value (OGEV) probabilities (Small 1987), which will allow for the presence or otherwise of ordering in the dependent variable and correlations of alternatives in close proximity. This is important as it may be the case that employment types are ranked. We explore the possibility that employment types are ranked according to the associated earnings uncertainty, from self-employment at

one extreme to fixed wage employment at the other. Moreover, it is likely that neighbouring alternatives will indeed be correlated.¹

The second extension, which the DOGEV model embodies, is the Dogit model of Gaudry and Dagenais (1979). This expands on Logit probabilities by the introduction of additional outcome-specific parameters, θ_j . In our application, $j = 1, 2, 3$, where 1 denotes fixed wage employment, 2 denotes PRP and 3 denotes self-employment. The interpretation of θ_j varies across application, but in general they can be regarded as heterogeneity of the outcome itself (i.e. employment type), as opposed to observed individual heterogeneity. In addition, they may also represent unobserved individual heterogeneity which is common to individuals within an outcome. For example, if the majority of individuals within a particular employment type belong to a particular religious denomination and religious denomination is a missing variable in the data set, then θ_j for this particular outcome will embody the effect of this variable.

It is intuitive to consider the choice-set generation framework of Manski (1977). Specifically, in the Dogit model an individual is assumed to be either captive to one of the J outcomes (here an employment type) or chooses freely from the full set (fixed wage employment, PRP and self-employment). Therefore, the available choice set faced by individual i , $B_i = B \forall i$, comprises $J + 1$ sets, J single outcome captivity sets and one set comprising all J

¹We focus here on testing for an ordering of employment types according to the associated degree of earnings uncertainty. The raw data superficially supports the hypothesis that employment types are ordered according to earnings uncertainty, with the standard deviation of earnings around the mean being £200.72 for fixed wage employment, £371.43 for PRP employment and £828.75 for self-employment. Our framework, however, is sufficiently flexible to explore ordering in other directions.

outcomes from which free choice is (subsequently) exercised by the individual. The choice set generation process itself can be represented as a random utility maximisation model with utilities given by

$$U_{ik}^{(1)} = W_{ik} + \eta_{ik}, \quad i = 1, \dots, n; \quad k = 1, \dots, J + 1. \quad (1)$$

Under the assumptions that: η_{ik} are independently and identically distributed as a Type 1 Extreme Value variate; $W_{ik} = \log(\theta_k)$; and the normalisation that $W_{iJ+1} = 0$, the probability of individual i choosing a single outcome (captive) set is given by

$$P_{ij} = \frac{\theta_j}{1 + \sum_{k=1}^J \theta_k}, \quad (2)$$

and the probability that individual i chooses from the full set is

$$P_{iJ+1} = \frac{1}{1 + \sum_{k=1}^J \theta_k}. \quad (3)$$

For the outcome selection process, the probability that an individual selects the specified outcome j from a single outcome choice set is one. The probability that an individual selects the specified outcome j from the full choice set is in the second stage derived from the standard random utility maximisation model, RUM (Fry, Brooks, Comley, and Zhang 1993), of

$$U_{ij}^{(2)} = V_{ij} + \varepsilon_{ij} \quad (4)$$

where $U_{ij}^{(2)}$ is the utility that individual i gains from alternative j in this second stage, and V_{ij} and ε_{ij} are, respectively, the non-stochastic and stochastic components of this utility. For simplicity, V_{ij} is typically specified as

$$V_{ij} = \mathbf{x}'_i \boldsymbol{\beta}_j, \quad (5)$$

and under the assumption that the ε_{ij} independently follow a Type 1 Extreme Value distribution, the resulting probabilities have the standard Logit form (Maddala 1983). So, utilising the Manski framework, the Dogit model can be parameterised as

$$P_{ij}^{Dogit} = \frac{\theta_j}{1 + \sum_{k=1}^J \theta_k} + \frac{1}{1 + \sum_{k=1}^J \theta_k} (P_{ij}^L) \quad (6)$$

where P_{ij}^L are the standard Logit probabilities for multiple outcomes.

Fry and Harris (2004) suggest utilising this basic set-up, but to replace the Logit probabilities with those of the OGEV model, which are given by

$$P_{ij}^{OGEV} = \frac{\exp(\rho^{-1}V_{ij})}{\sum_{r=1}^{J+1} (\exp(\rho^{-1}V_{i,r-1}) + \exp(\rho^{-1}V_{ir}))^\rho} \quad (7)$$

$$\times \left[(\exp(\rho^{-1}V_{i,j-1}) + \exp(\rho^{-1}V_{ij}))^{\rho-1} + (\exp(\rho^{-1}V_{ij}) + \exp(\rho^{-1}V_{i,j+1}))^{\rho-1} \right],$$

where the parameter ρ allows for any ordering in the data.

Akin to a moving average process, the OGEV (and hence DOGEV) model(s) account for ordering in the data by allowing a correlation between alternatives in close proximity, which decreases the further are the two alternatives apart. Although ρ is not strictly a correlation coefficient - which has no closed form expression in these models (Small 1987) - it is inversely related to this such that when $\rho = 1$, the OGEV (DOGEV) model collapses to the Logit (Dogit) one.²

Therefore the full functional form for the DOGEV choice probabilities are given by

²See (Small 1987) for full details of the correlation structure.

$$P_{ij}^{DOGEV} = \frac{\theta_j}{1 + \sum_{k=1}^J \theta_k} + \frac{1}{1 + \sum_{k=1}^J \theta_k} \times P_{ij}^{OGEV} \quad (8)$$

This specification is convenient in that simple parameter restriction tests of $\rho = 1$ and $\theta = 0$, are tests against the nested sub-models of Dogit and OGEV (and jointly, Logit), respectively. Such parameter restrictions can be easily tested for using usual testing paradigms.

Using the indicator function d_{ij} where

$$d_{ij} = \begin{cases} 1 & \text{if individual } i \text{ chooses alternative } j \\ 0 & \text{otherwise} \end{cases}$$

the parameters of the model are now estimated using the maximum likelihood criterion, where the log-likelihood function is

$$L(\phi) = \sum_{j=1}^J \sum_{i=1}^N d_{ij} \ln P_{ij}^{DOGEV}.$$

with $\phi' = [(\text{vec} \beta_j)', \theta', \rho]$ and P_{ij}^{DOGEV} being given by equation (8).

3.2 The Parameterised DOGEV Model

In a model of employment type, a relatively standard set of observed individual characteristics are likely to directly affect the (second-stage) utilities of the individual via the index functions described in equations (5) and (4). Labour market rigidities, and in particular demand side constraints, may lead to captivity of individuals in particular types of employment. However, an important question is: are these captivity effects constant across individuals? In other words, is there heterogeneity of the various employment types, that will vary in its effect across individuals? For example, consider hours worked - certain individuals may prefer a particular length of working

week. However, such divergences in hours worked are likely to be specific to particular employment types and predominantly controlled by employers - for example fixed wage contracts are likely to be associated with a standard working week.

Therefore, we extend the basic DOGEV model to allow the captivity parameters θ to vary by observed factors z which characterise employment type heterogeneity (or demand side effects) such that

$$\theta_j = \exp(\mathbf{z}'\boldsymbol{\gamma}_j), \quad (9)$$

where the use of the exponential transformation ensures non-negativity of the θ parameters, required for the probabilities of equation (8) to be properly defined (Gaudry and Dagenais 1979). Such a generalisation appears appropriate in a model explaining observed types of employment, as it is possible to identify factors which are specific to employment types that are likely to impact on the probability of self-employment. We will refer to this specification as the Parameterised DOGEV Model. The instruments we use to parameterise the DOGEV model and thereby control for employment type heterogeneity are industry, hours of work and ethnic origin.

Certain industries are more suited to particular types of employment. Where large economies of scale and/or large sunk costs exist, for example, we would not expect to find a high prevalence of self-employed workers. Thus, individuals employed in different industries may be captive, to varying extents, to particular employment types.

Hours of work are usually contractually specified and hours inflexibility is a well documented labour market rigidity. Within fairly narrow bands,

employees cannot, generally, freely choose the number of hours they work (although some flexibility is generated by the availability of overtime and flexi-time, but this too is employer driven). Self-employment is generally associated with working longer hours than other types of employment. Thus employment types may be heterogeneous in terms of the number of hours worked.

Finally, as suggested in the literature, we argue that employer based discrimination may push ethnic minorities into self-employment and thus employment types may be heterogeneous in terms of ethnic groups. However, it is important to note that, in this respect, this is a demand side factor which will impact on the incidence of self-employment among ethnic minority groups.

Including demand side variables as standard regressors does not take into consideration their true impact in terms of tying workers to particular types of employment. This could lead to misspecification and potentially erroneous inference. To ascertain the contribution of modelling employment type heterogeneity to our understanding of who are the self-employed, we compare the Parameterised DOGEV model with the standard multinomial Logit model using the data described below.

4 The Data

Our data are drawn from the Family Expenditure Survey (FES) for the UK, which is a nationally representative survey that is conducted on an annual basis. Approximately 10,000 households are selected each year and the average response rate is around 70%. The main aim of the survey is to

Table 1: Distribution of the Sample by Employment Contract

Self-employed		PRP		Fixed Wage		Total	
Number	%	Number	%	Number	%	Number	%
1287	9.4%	3,623	26.5%	8,765	64.1%	13,675	100%

provide a reliable source of information on household expenditure, income and other aspects of household finances.

We use pooled data from the 1997/98, 1998/99 and 1999/00 surveys. Our sub-sample comprises of working adults aged between 18 and 65 who are either self-employed, employed under a fixed wage contract or on a contract characterised by PRP.³

Sample statistics are set out in Tables 1 - 3. It is apparent from Table 1 that fixed wage contracts are the dominant form of employment type (64.1%) and self-employment (9.41%) is the least common.⁴ Table 2 shows there are interesting differences in the incidence of employment type across individual characteristics. For example, men are relatively more likely to be self-employed. The incidence of self-employment is higher amongst married relative to single, separated, divorced and widowed respondents. With respect to ethnicity, PRP contracts are more heavily concentrated among whites whilst the incidence of self-employment is relatively high among non whites.

One might also hypothesise that both the number and the age of children

³The FES contains information on performance related bonus schemes (e.g. Christmas bonuses; productivity bonuses; profit related bonuses; loyalty bonuses; dividends; incentive schemes and performance/sales bonuses), which allows us to identify those individuals who receive PRP.

⁴The 9.4% self-employment figure ties in closely with that stated earlier of 9% for the UK suggesting that we have a representative working sample of the working population.

may affect a parent's willingness to become self-employed. We therefore consider the number of pre-school and school-age children in the household and find that whilst the former are distributed relatively evenly across the three types of employment, the latter are highest amongst self-employed workers.

The self-employed are, on average, much older - a finding that is consistent with the hypothesis that older, displaced workers might turn to self-employment given their relatively low probability of re-employment. It might also reflect the ability of older workers to acquire the necessary start-up capital for self-employment, on account of their longer accumulation of wealth.

Individuals in self-employment have relatively few formal qualifications. PRP contracts, on the other hand, are concentrated amongst those with formal school qualifications and higher, whilst individuals holding fixed wage contracts are relatively evenly spread across all levels of schooling.

In relation to housing tenure, fixed wage employees are most likely to be found living in local authority housing, whilst the self-employed are more likely to own their homes outright and PRP workers are most likely to live in homes that are mortgaged. This may reflect the fact that housing equity is often used as collateral to secure loans necessary to start up a small business.

Finally, we can see that the self-employed have a much longer working week than those employed under fixed wage or PRP contracts. This is consistent with the known labour market rigidities in terms of hours of work since fixed wage and PRP employees are generally not free to determine the number of hours worked. Hours of work are usually a contractually specified aspect of the employment contract. Fixed wage and PRP workers have, on

Table 2: Employment Type by Individual Characteristics

	Self-employed	PRP	Fixed Wage
Males	11.77	28.19	60.04
Females	4.67	23.07	72.26
Age*	43.62	37.89	38.61
Family Status			
Single	4.15	26.18	69.67
Married/Cohabit	10.28	26.79	62.39
Separated/Widowed/Divorced	9.44	24.50	66.06
Pre-school Children*	0.22	0.23	0.18
School-Age Children*	0.75	0.47	0.52
Ethnicity			
White	9.28	26.78	63.94
Non-white	12.33	20.03	67.64
Education Level			
Less than GCSE	14.06	22.36	63.58
GCSE	9.29	26.94	63.77
Further Education	7.38	28.48	64.14
Higher Education	7.78	27.28	64.94
Housing Tenure			
Rented- Local Authority	6.11	22.17	71.72
Rented - Private	9.82	22.16	68.02
Mortgaged	9.02	28.44	62.54
Owned	13.89	22.20	68.92
Household Income (Weekly)*	627.19	705.33	601.85
Hours of work*	50.77	39.67	39.92

Note: Numbers are expressed as a percentage of the total number of individuals within the three types of employment for each individual characteristic, except for those denoted by * which represent the average for each types of employment

Table 3: Employment Type by Industry

	Fixed Wage	PRP	Self-employed	Total
Sample sizes				
Retail	1917	1043	451	3411
Manufacturing	2990	1553	522	5065
Services/public	3858	1027	314	5199
Total	8765	3623	1287	13675
Percentages				
Retail	56	31	13	100
Manufacturing	59	31	10	100
Services/public	74	20	6	100

average, a standard working week of approximately 40 hours, whilst the self-employed are observed to have, on average, a much longer working week of around 50 hours.

Table 3 shows the relationship between employment and industry. Employment in a given industry may be associated with employment type heterogeneity. Fixed wage contracts are the dominant type of employment across all three (broad) industrial categories. However there are interesting differences regarding the frequency of PRP and self-employment across the industries. Employment types other than fixed wage employment are very rare within the service industry and the self-employed are more common in retail and manufacturing, both of which also have a high frequency of PRP.

5 The Results

For comparison purposes, we report results from both the Parameterised DOGEV model and the multinomial Logit model. The results are set out in Tables 4 and 5 below. In terms of the explanatory variables, a comparison of

the multinomial Logit estimates with the Parameterised DOGEV estimates shows a fairly consistent story, although the significance of some variables varies across the specifications. The high degree of significance of the captivity parameters in the Parameterised DOGEV model is particularly noticeable indicating that they have an important role to play in explaining employment type heterogeneity.⁵

Prior to discussing the results of the Parameterised DOGEV model in detail, we will present a comparison of the predictive capacity of the two models. Table 6 shows the sample proportions across types of employment and the within-sample predicted proportions from the multinomial Logit and the Parameterised DOGEV models. We can see that unsurprisingly both specifications over predict the dominant outcome, fixed wage employment. In terms of overall predictive power, the Parameterised DOGEV model is slightly better, predicting 66% of observations correctly; compared to 65% for the multinomial Logit specification. However the overall predictive capacity of the two models is misleading given the dominance of fixed wage contracts in the data - a naïve model predicting fixed wage employment for the entire sample would correctly predict 64% of the observations. A better criterion for judging the performance of the models is their ability to predict outcomes outside of the dominant category and the Parameterised DOGEV does better than the multinomial Logit in this respect, especially with re-

⁵A likelihood ratio test of the Parameterised DOGEV model against the DOGEV specification yielded a Chi-squared test statistic of 1820 with 12 degrees of freedom. Hence, the parameterisation of the captivity terms does make a significant contribution to the likelihood function thereby supporting the parameterised specification. For brevity, we do not present the results of the DOGEV model - the full results are available from the authors on request.

Table 4: Multinomial Logit Estimates
Reference Category: Fixed Wage Contracts

Variable	PRP Contract			Self-Employed		
	Coef			Coef		
Constant	-1.981	(0.32)	**	-10.310	(0.57)	**
Male	0.353	(0.05)	**	0.453	(0.09)	**
Age	0.915	(0.15)	**	1.388	(0.27)	**
Age squared	-0.122	(0.02)	**	-0.124	(0.03)	**
Married77	0.088	(0.06)		0.236	(0.13)	*
Separated/widowed/divorced	0.133	(0.10)		0.217	(0.17)	
Pre school children	0.071	(0.04)		0.181	(0.07)	**
School age children	-0.147	(0.03)	**	0.156	(0.04)	**
Education						
GCSE	0.097	(0.07)		-0.061	(0.09)	
Further Education	0.238	(0.07)	**	-0.093	(0.11)	
Higher Education	0.131	(0.07)	*	-0.182	(0.11)	
Housing Tenure						
Rented private	-0.021	(0.10)		0.752	(0.18)	**
Mortgaged	0.284	(0.07)	**	0.575	(0.14)	**
Owned	0.124	(0.09)		0.956	(0.16)	**
Year of Survey						
1998/99	-0.263	(0.05)	**	-0.323	(0.08)	**
1997/98	-0.213	(0.05)	**	-0.218	(0.08)	**
Household Income	0.286	(0.04)	**	0.019	(0.07)	
Region						
Wales	-0.146	(0.11)		-0.043	(0.17)	
Scotland	-0.070	(0.13)		0.197	(0.20)	
North England	-0.021	(0.08)		-0.041	(0.13)	
Midlands	-0.168	(0.08)	**	-0.135	(0.13)	
South England	0.048	(0.05)		-0.250	(0.09)	**
Captivity Parameters						
Manufacturing	-0.085	(0.05)		-0.248	(0.08)	**
Services/public	-0.753	(0.05)	**	-0.953	(0.09)	**
Hours of work	-0.191	(0.04)	**	0.995	(0.03)	**
Non-white	-0.266	(0.11)	**	0.284	(0.16)	*
Number of Observations	13675					
Log Likelihood	-10500					

Standard errors in parentheses. ** and * significant at 5 and 10% level, respectively (two-sided).

Table 5: Parameterised DOGEV Estimates

Variable	PRP Contract			Self-Employed		
	Coef			Coef		
Constant	-3.360	(1.05)	**	-8.588	(1.83)	**
Male	0.675	(0.19)	**	0.406	(0.28)	
Age	1.276	(0.51)	**	2.841	(0.79)	**
Age squared	-0.184	(0.07)	**	-0.264	(0.09)	**
Married	0.153	(0.16)		1.024	(0.34)	**
Separated/widowed/divorced	0.092	(0.26)		0.930	(0.42)	**
Pre school children	-0.123	(0.07)		0.272	(0.17)	
School age children	0.437	(0.18)	*	0.237	(0.10)	**
Education						
GCSE	-0.323	(0.18)	*	-0.063	(0.22)	
Further Education	-0.025	(0.17)		-0.128	(0.28)	
Higher Education	-0.125	(0.17)		0.080	(0.28)	
Housing Tenure						
Rented private	-0.216	(0.27)		1.533	(0.47)	**
Mortgaged	0.506	(0.21)	**	1.781	(0.39)	**
Owned	0.469	(0.26)	*	2.502	(0.48)	**
Year of Survey						
1998/99	-0.413	(0.15)	**	-0.574	(0.20)	**
1997/98	-0.397	(0.15)	**	-0.363	(0.20)	*
Household Income	0.437	(0.18)	**	-4.719	(1.04)	**
Region						
Wales	-0.668	(0.34)	*	-0.881	(0.49)	
Scotland	-0.074	(0.30)		0.658	(0.54)	
North England	-0.151	(0.19)		-0.158	(0.27)	
Midlands	-0.433	(0.24)	*	-0.432	(0.31)	
South England	-0.007	(0.13)		-1.009	(0.31)	
Captivity Parameters						
Fixed Wage						
Constant	11.860	(0.75)	**			
Manufacturing	-0.040	(0.16)				
Services	0.387	(0.20)	**			
Hours of work	-2.720	(0.19)	**			
Non-white	0.354	(0.29)				
PRP						
Constant	14.120	(0.89)	**			
Manufacturing	-0.096	(0.17)				
Services	-0.473	(0.20)	**			
Hours of work	-3.456	(0.23)	**			
Non-white	-0.012	(0.30)				
Self-Employed						
Constant	-2.777	(0.36)	**			
Manufacturing	-0.187	(0.15)				
Services	-0.713	(0.17)	**			
Hours of work	0.348	(0.06)	**			
Non-white	0.706	(0.26)	**			
ρ	0.562	(0.27)	**			
Number of Observations	13675					
Log Likelihood	-10359					

Standard errors in parentheses. ** and * significant at 5 and 10% level, respectively (two-sided).

Table 6: Sample proportions and predicted outcomes

	Fixed wage	PRP	Self-employed	Correct
Sample	0.641	0.265	0.094	
Multinomial Logit	0.945	0.016	0.039	0.65
Parameterised DOGEV	0.938	0.011	0.051	0.66

gard to self-employment. The Parameterised DOGEV model does better at predicting outcomes within the self-employed category (although marginally worse in the PRP one).⁶ Thus, explicitly accounting for employment type heterogeneity enhances our ability to predict employment types.

To further consider the within-sample predictive accuracy of the Parameterised DOGEV model, Table 7 presents a simulated hit and miss table. The predictions underlying hit and miss tables are given by assigning individuals the outcome associated with their highest predicted probability across the alternatives and then comparing this to the observed outcome. As is generally the case with discrete choice models (Duncan and Harris 2002) both of these models tended to over predict the most common alternative (as noted above). This is because construction of traditional hit and miss tables implicitly ignores the stochastic elements of the underlying economic model. Therefore we present simulated hit and miss tables where we explicitly take into account both the stochastic and non-stochastic elements of the underlying utility function (over 1,000 random draws). For comparison purposes, the case of a random assignment of individuals' types of employment according to observed sample proportions is also presented. The Parameterised DOGEV model predicts better than simple random assignment. Thus, we find

⁶Neither models predict the PRP category particularly well. One suspects that this is due to a paucity of variables pertaining to the workplace.

Table 7: Hit and Miss Tables
RANDOM ASSIGNMENT

		Predicted			
Actual	Fixed wage	PRP	Self-employed	Total	
Fixed wage	5618	2322	825	8765	
PRP	2322	960	341	3623	
Self-employed	825	341	121	1287	
Total	8765	3623	1287	13675	

PARAMETERISED DOGEV SIMULATED

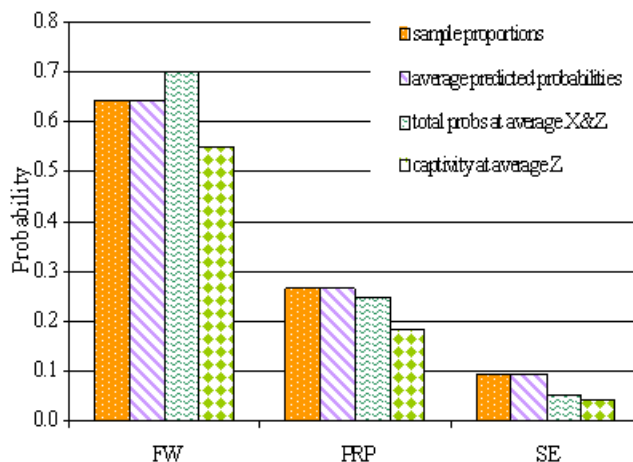
		Predicted			
Actual	Fixed wage	PRP	Self-employed	Total	
Fixed wage	5837	2296	632	8765	
PRP	2301	1069	253	3623	
Self-employed	624	263	400	1287	
Total	8762	3628	1285	13675	

further evidence suggesting that accounting for employment type captivity is particularly important in predicting employment outcomes.

Figure 1 illustrates the predictive power of the Parameterised DOGEV model by graphing the actual sample proportions by type of employment, the average predicted probability across the individuals for each outcome, the total probabilities predicted at the sample means of the explanatory variables and the amount of the latter accounted for by the captivity within each employment type.⁷ Under these two tests of predictive power we can see the model clearly performs well. The average probability of being in each employment type accurately matches the sample proportions. Whilst calculating the predicted probabilities at the sample means of the variables does slightly over predict the number of fixed wage employees and slightly under

⁷This differs from Table 6 which is based on observed individual characteristics as opposed to sample means.

Figure 1: Sample Proportions; Predicted Probabilities - average and at sample means (by total and captive)



predict the number of PRP and self-employed workers, the orders of magnitude are small. Moreover, the graph illustrates the substantial contribution made to the total probabilities arising from the demand side (or captivity) factors.

Another short fall of the multinomial Logit specification is that it cannot account for the possibility of ordering across employment types. The Parameterised DOGEV model, on the other hand, has the flexibility to allow us to test for the presence of ordering in the dependent variable. A simple test of the null hypothesis of no ordering is essentially a test of $\rho = 1$. We find that ρ is statistically significantly different from 0 and 1 (at the 95% level). Thus, we reject the null hypothesis which suggests that employment types display ordering. As suggested earlier, this may reflect the earnings uncertainty associated with each outcome with self-employment at one extreme and fixed wage employment at the other. Ignoring the ordering in our de-

pendent variable would result in misspecification and potentially erroneous inferences.

Given the complexity of the implicit marginal effects of the Parameterised DOGEV model and, for reasons of clarity, we present the implied probabilities of self-employment for different realisations of our explanatory variables in Table 8.⁸ Specifically, we consider the impact of both individual heterogeneity and employment type heterogeneity by considering each explanatory variable in turn whilst setting all other variables to their sample means.

Table 8 has been constructed so that the contribution to the total probabilities generated by the individual and employment type heterogeneity can be distinguished. Note that as we change x (individual heterogeneity or supply side variables) total probabilities change via equations (5) and (8), although the captivity probabilities are unaltered. On the other hand, when we change z (employment type heterogeneity or demand side variables), captive probabilities are directly affected via equations (9) and (8). In addition, the non-captive ones are also affected via a scale factor implied by these equations. That is, changes in x change the total probabilities through the impact on P_{ij}^{OGEV} and hence affect only the second term of equation (8). Whilst changes in z influence θ_j and so affect both terms in (8): the first term directly and the second term via a scale factor.

In Table 8, we present the total probabilities of self-employment for each change as well as the proportions of the total probability accounted for by employment type heterogeneity and individual heterogeneity. In terms of

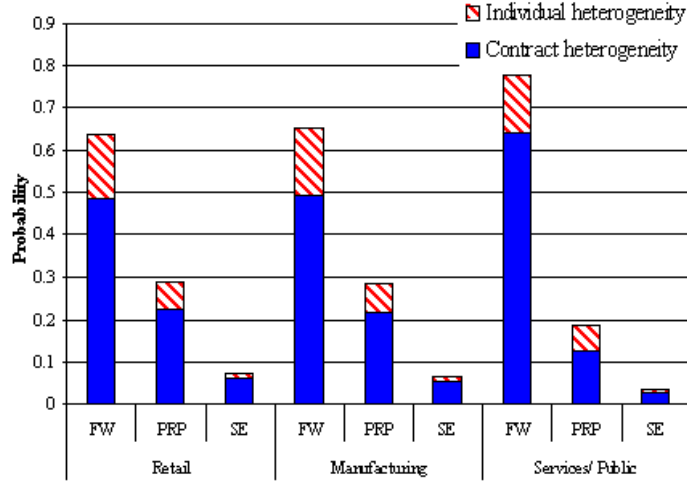
⁸Given that the focus of our paper is on self-employment, the corresponding probabilities for fixed wage employment and PRP have been omitted for reasons of brevity but are available from the authors on request.

gender, employment type heterogeneity appears to be more important for self-employed men than for self-employed women, which may reflect the fact that self-employed women fit self-employment around family commitments - consequently individual heterogeneity may be particularly important here. Regarding age, individual heterogeneity is not very important when young, but becomes increasingly important for individuals aged 40 and above. This ties in with the existing literature reporting that age is positively related to self-employment. Employment type heterogeneity appears to dominant the probability of self-employment for all categories of marital status with very little variation in terms of it's contribution to the total probabilities across each category. In terms of number of children the effect of employment type heterogeneity declines as number of children increases.

There is some variation in the effect of education at the relatively higher levels of education, the contribution of employment type heterogeneity is highest for further education and lowest for higher education. As noted in Section II, the role of education in determining self-employment has led to some mixed results. Our findings suggest that it is particularly important to distinguish between types of education beyond 'A' level.

In terms of housing tenure, the effect of individual heterogeneity is particularly pronounced for the self-employed - a finding which is consistent with the hypothesis that capital/asset accumulation is associated with self-employment, given that housing equity is a common source of loans security. The contribution of employment type heterogeneity to the total probability of self-employment appears to increase with income. In addition, it is apparent that there is significant variation across regions in terms of the im-

Figure 2: Predicted Probabilities: Industry



portance of the two types of heterogeneity with demand side influences being particularly pronounced in the South.

Given that introduction of the captivity parameters represents the novel part of the model, we will consider all types of employment when discussing the results relating to these parameters. Hence, Table 8 is accompanied by three graphs showing the contribution to the total probabilities generated by individual and employment type heterogeneity for each of the three demand side effects for each of the three types of employment. We consider in turn industry, hours worked and ethnicity.

We find that fixed wage workers are most likely to be found in the service/public sector and the increased probability is largely due to the captivity parameters. PRP and self-employment in contrast are least likely to be found in this sector (see Figure 2).

Across the range of hours worked, we find the largest degree of variation

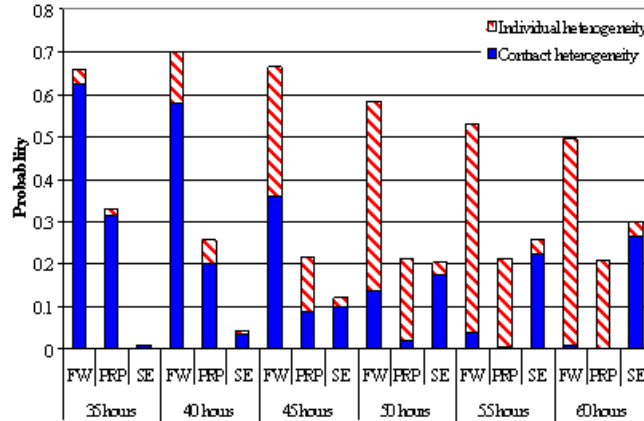
Table 8: Self-employment Predicted Probabilities

	Contract Heterogeneity		Individual Heterogeneity		Total Probability
	Probability	%	Probability	%	
Gender					
Male	0.043	(80.45)	0.010	(19.55)	0.053
Female	0.043	(45.47)	0.051	(54.53)	0.094
Age					
Age 20	0.043	(80.45)	0.001	(19.55)	0.044
Age 30	0.043	(89.91)	0.005	(10.09)	0.048
Age 40	0.043	(75.29)	0.014	(24.71)	0.057
Age 50	0.043	(61.78)	0.026	(38.22)	0.069
Age 60	0.043	(58.55)	0.030	(41.45)	0.073
Marital status					
Single	0.043	(80.45)	0.004	(19.55)	0.047
Married	0.043	(77.80)	0.012	(22.20)	0.055
Separated/widowed/divorced	0.043	(79.08)	0.011	(20.92)	0.054
Number of preschool children					
None	0.043	(80.45)	0.009	(19.55)	0.052
1	0.043	(78.06)	0.012	(21.94)	0.055
2	0.043	(73.74)	0.015	(26.26)	0.058
Number of school age children					
None	0.043	(80.45)	0.008	(19.55)	0.051
1	0.043	(78.92)	0.011	(21.08)	0.054
2	0.043	(73.50)	0.015	(26.50)	0.058
3	0.043	(67.45)	0.021	(32.55)	0.063
Education					
No formal qualifications	0.043	(80.45)	0.010	(19.55)	0.052
GCSE	0.043	(80.66)	0.010	(19.34)	0.053
Further education	0.043	(83.44)	0.008	(16.56)	0.051
Higher education	0.043	(79.45)	0.011	(20.55)	0.054
Housing Tenure					
Rent LA	0.043	(95.86)	0.002	(4.14)	0.045
Rent	0.043	(80.31)	0.010	(19.69)	0.053
Mortgage	0.043	(80.19)	0.011	(19.81)	0.053
Owned	0.043	(65.04)	0.023	(34.96)	0.066
Household income					
25th percentile (£345)	0.043	(51.91)	0.040	(48.09)	0.082
50th percentile (£522)	0.043	(71.99)	0.017	(28.01)	0.059
75th percentile (£771)	0.043	(90.50)	0.004	(9.50)	0.047

Table 8: Self-employment Predicted Probabilities - continued

	Contract Heterogeneity		Individual Heterogeneity		Total Probability
	Probability	%	Probability	%	
Region					
Northern Ireland	0.043	(76.96)	0.013	(23.04)	0.056
Wales	0.043	(87.24)	0.006	(12.76)	0.049
Scotland	0.043	(61.96)	0.026	(38.04)	0.069
North England	0.043	(78.80)	0.012	(21.20)	0.054
Midlands	0.043	(81.86)	0.009	(18.14)	0.052
South England	0.043	(90.83)	0.004	(9.17)	0.047
Industry					
Retail	0.061	(85.86)	0.010	(14.14)	0.071
Manufacturing	0.053	(83.42)	0.011	(16.58)	0.063
Services/public	0.027	(74.82)	0.009	(25.18)	0.036
Hours worked					
35	0.008	(77.88)	0.002	(22.12)	0.010
40	0.034	(80.73)	0.008	(19.27)	0.043
45	0.099	(83.26)	0.020	(16.74)	0.119
50	0.174	(85.57)	0.029	(14.43)	0.203
55	0.227	(87.62)	0.032	(12.38)	0.259
60	0.267	(89.37)	0.032	(10.63)	0.298
Ethnicity					
White	0.042	(80.45)	0.010	(19.55)	0.052
Non white	0.067	(89.47)	0.008	(10.53)	0.075

Figure 3: Predicted Probabilities: Hours Worked



in the contribution of the individual and employment type heterogeneity parameters to the three employment outcomes. This is illustrated in Figure 3.

Fixed wage workers are most likely to be observed working a standard working week (approximately 40 hours), PRP workers have a slightly shorter working week (35 hours per week) and the self-employed have longer working hours (the probability of being self-employed increases as hours increases). Captivity variables dominate the probability of being employed on a fixed wage contract for low hours of work, however, their contribution to the total probability diminishes rapidly as the number of hours worked increases. The same pattern is observed for those employed under PRP contracts; whilst the converse is true for the self-employed. The self-employed are much more likely to be observed working long hours and logically the employment type heterogeneity parameters dominate this probability.

Figure 4: Predicted Probabilities: Ethnicity

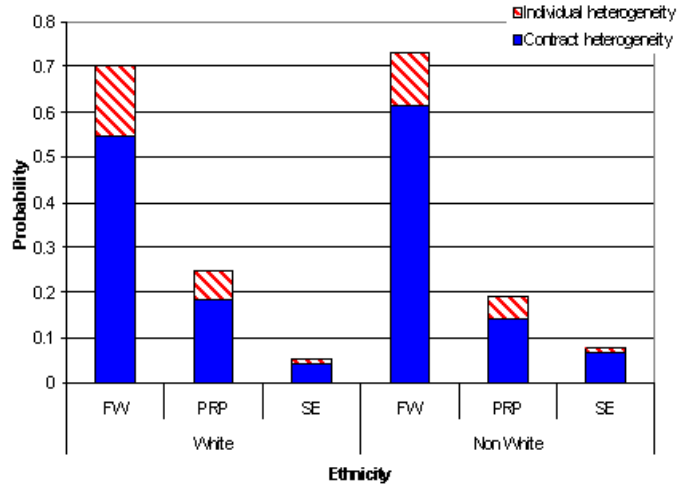


Figure 4 shows how employment type varies with ethnic origin. There have been a number of studies that have shown that ethnic minorities have a higher probability of self-employment (see, Clark and Drinkwater 2000, Brock and Evans 1986), and this finding is reflected in our results. The graph shows that non-whites are much more likely to be self-employed than the other ethnic groups, Borooah and Hart (1999) find similar evidence. In contrast, employment type heterogeneity plays a large role in determining the probability of being a fixed wage worker for whites. The significance of ethnic origin as a captivity parameter is consistent with the hypothesis that labour market discrimination pushes non-whites into self-employment.

In summary, our results show that employment type heterogeneity is important in modelling employment outcomes. Moreover, simply including demand side variables as standard regressors in a model of individual heterogeneity leads to relatively poor predictive power when compared to

a specification that explicitly and simultaneously controls for the captivity impact of the demand side factors.

6 Conclusions

Economists have in general found it hard to predict who are the self-employed. We show that whilst individual characteristics are important identifiers of employment type, aspects of employment type heterogeneity also play an important role. Modelling employment type heterogeneity as individual heterogeneity (i.e. as in the multinomial logit framework) leads to poor predictions when compared to the Parameterised DOGEV model that allows us to separate the effects of individual and employment type heterogeneity. Moreover, our results suggest that the existing literature may have drawn conclusions from potentially misspecified models. Furthermore, we have found evidence of ordering across employment types suggesting an additional source of misspecification in the existing literature which has largely focused on unordered discrete choice models.

We have shown that the self-employed are captive to self-employment due to their ethnicity, long work hours and the industry in which they work. Moreover, we show that the probabilities of being in any particular employment type are also driven by supply side factors. Our results concerning individual heterogeneity are consistent with the existing literature in as much as the self-employed are more likely to be older and have housing equity.

References

- BERNHARDT, I. (1994): “Comparative Advantage in Self-Employment and Paid Work,” *Canadian Journal of Economics*, 27, 273–289.
- BLANCHFLOWER, D. G., AND A. J. OSWALD (1998): “What Makes an Entrepreneur?,” *Journal of Labor Economics*, 16(1), 26–60.
- BORJAS, G. I. (1986): “The Self-Employment Experience of Immigrants,” *The Journal of Human Resources*, 21, 485–506.
- BORJAS, G. I., AND S. G. BRONARS (1989): “Consumer Discrimination and Self-Employment,” *Journal of Political Economy*, 97, 581–605.
- BOROOAH, V. K., AND M. HART (1999): “Factors Affecting Self-Employment Among Indian and Black Caribbean Men in Britain,” *Small Business Economics*, 13(2), 111–129.
- BROCK, W. A., AND D. S. EVANS (1986): “The Economics of Small Business: Their Role and Regulation in the U.S. Economy,” *New York: Holmes and Meier*.
- CALVO, G., AND S. WELLISZ (1980): “Technology, Entrepreneurs and Firm Size,” *Quarterly Journal of Economics*, 85, 663–678.
- CLARK, K., AND S. DRINKWATER (2000): “Pushed Out or Pulled In? Self-Employment Among Ethnic Minorities in England and Wales,” *Labour Economics*, September, 603–28.

- DE WIT, G., AND V. F. A. A. M. WINDEN (1989): “An Empirical Analysis of Self-Employment in the Netherlands,” *Small Business Economics*, 1, 263–272.
- DUNCAN, A., AND M. HARRIS (2002): “Simulating the Effect of Welfare Reforms Among Sole Parents in Australia,” *The Economic Record*.
- DUNN, T., AND D. HOLTZ-EAKIN (2000): “Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links,” *Journal of Labor Economics*, 18(2), 282–305.
- EVANS, D. S., AND L. S. LEIGHTON (1989): “Some Empirical Aspects of Entrepreneurship,” *American Economic Review*, 79, 519–535.
- EVANS, M. D. R. (1989): “Immigrant Entrepreneurship: Effects of Ethnic Market Size and Isolated Labor Pool,” *American Sociological Review*, 54, 950–962.
- FRY, T., R. BROOKS, B. COMLEY, AND J. ZHANG (1993): “Economic Motivations for Limited Dependent and Qualitative Variable Models,” *Economic Record*, 69, 193–205.
- FRY, T., AND M. HARRIS (2004): “The Dogit Ordered Generalized Extreme Value Model,” *Forthcoming in Australian and New Zealand Journal of Statistics*.
- GAUDRY, M., AND M. DAGENAIS (1979): “The Dogit Model,” *Transportation Research - Part B*, 13B, 105–112.

- HOUT, M., AND H. ROSEN (2000): "Self-Employment, Family Background and Race," *Journal of Human Resources*, 35(4), 670–92.
- JOHANSSON, E. (2000): "Self-Employment and Liquidity Constraints: Evidence from Finland," *Scandinavian Journal of Economics*, 102(1), 123–34.
- KIDD, M. P. (1993): "Immigrant Wage Differentials and the Role of Self-Employment in Australia," *Australian Economic Papers*, 32, 92–115.
- LE, A.-T. (1999): "Empirical Studies of Self-Employment," *Journal of Economic Surveys*, 13(4), 381–416.
- MADDALA, G. S. (1983): *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, U.K.
- MANSKI, C. (1977): "The Structure of Random Utility Models," *Theory and Decision*, 8, 229–254.
- REES, H., AND A. SHAH (1986): "An Empirical Analysis of Self-Employment in the UK," *Journal of Applied Econometrics*, 1, 95–108.
- SCHILLER, B., AND P. CREWSON (1997): "Entrepreneurial Origins: A Longitudinal Inquiry," *Economic Inquiry*, 25, 523–531.
- SMALL, K. (1987): "A Discrete Choice Model for Ordered Alternatives," *Econometrica*, 55, 409–424.
- TAYLOR, M. (1996): "Earnings, Independence or Unemployment: Why Become Self-Employed?," *Oxford Bulletin of Economics and Statistics*, 30, 194–204.