BEYOND THE 'EMPLOYEE OR ENTREPRENEUR' DICHOTOMY:

ACCOUNTING FOR ENTREPRENEURIAL PERSISTENCE AND TYPE

by

Andrew E. Burke[1] (University of Warwick)

Felix R. FitzRoy[2] (University of St Andrews)

and

Michael A. Nolan[3] (University of Hull)

December 2003

Abstract

The paper makes three contributions to the economics literature on entrepreneurship. We offer a new measure of entrepreneurship which accounts for variations in persistence in self-employment. We outline an econometric methodology to account for this approach and find that it is superior to probit/logit models which have dominated the literature. While our results indicate that this existing literature is good at explaining an individual's propensity for self-employment, we find that entrepreneurial persistence is determined by a different model and unearth some new insights into the roles of early career experience, finance, role models, gender and the unemployment push effect.

Keywords: Self-employment, count data.

JEL classifications: J23, C25.

---

[1] Reader in Enterprise Finance & Economics, Centre for SMEs & Mercia Institute of Enterprise, Warwick Business School, University of Warwick, CV4 7AL. Email: andrew.burke@wbs.ac.uk
[2] Professor, Dept of Economics, University of St Andrews, KY16 9AL. Email: frf@st-andrews.ac.uk
[3] CORRESPONDING AUTHOR. Lecturer in Economics, University of Hull, Cottingham Road, Hull, HU6 7RX, UK. Telephone: +44 (01482) 465855. Fax: +44 (01482) 466216. EMAIL: M.A.Nolan@hull.ac.uk

Introduction

Extensive research over the past 25 years has looked for characteristics that drive individuals to choose self-employment instead of wage employment. It has been typical to use cross-sectional data, viewing individuals only at a particular point in time or comparing their behaviour between two points in time, so that dichotomous choice models can be used to identify characteristics associated with either state. This approach has its theoretical foundations in discrete models of career choice (such as Kihlstrom and Laffont 1979 and Evans and Jovanovic 1989), and has been driven at least in part by the ease with which probit and logit analysis can be used to predict probable career choice between these two alternatives (for example, Evans and Leighton 1989, and Blanchflower and Oswald 1998).

As a consequence, the much more rich (and realistic) dynamics of individual career choice have been largely neglected. It is well known that many entrepreneurs gain some prior experience in paid employment before starting their own business. Similarly, others use wage work as a means of saving finance for a new venture. Most new start-ups then fail within a few years, and many of the (formerly) self-employed who suffered from bad luck or bad judgement then switch to temporary or lasting employment, perhaps after spells of unemployment or inactivity. Likewise, some individuals are serial entrepreneurs and some of these spend temporary periods in employment between ventures; especially those who are unsuccessful. This depiction contrasts with the methodology of discrete models and empirical analysis where individuals are forced into mutually exclusive categories comprising pure wage-workers and pure entrepreneurs – 'pure' in the sense that individuals spend either all of their time in wage-work or self-employment. However, labour market data indicates that this is not only inaccurate but is actually very misleading. Over a nine year period, our data set illustrates that while self-employment is a minority career activity, within this category 'pure' entrepreneurs are outnumbered by individuals who mix their career with spells in both self-employment and wage-work. The aim of this paper is to try to move beyond the dichotomous

depiction of entrepreneurship and wage-work, and begin to explore the implications of entrepreneurial persistence.

Longitudinal analysis of individual careers and transitions, in terms of cognitive and behavioural characteristics and dynamic optimisation by agents with bounded rationality, as well as the random shocks that often seem to override "the best laid plans of mice and men", raises formidable problems in terms of estimation and modelling. In this paper we shall only take a modest step beyond the static, binary choice approach by considering a reduced-form, *ex post* result of sequentially 'optimal' decisions, namely the total time spent in self-employment over a period of approximately 9 years. In reality, this summary measure can obviously include multiple spells and thus does not address spell durations as such. The complementary measure "time not self-employed" includes employment, unemployment and inactivity – though for many (particularly among the males in our chosen cohort) it will be mainly employment; and, in any case, we do include some basic controls to provide crucial distinctions between those who are employed and those who are not.

Once we proceed beyond the confines of considering individuals only at a point in time, we could in principle go as far as to examine all individuals on a continuum between choosing to spend their entire career history in self-employment, or none of it, or any intermediate fraction. However, our data, the crucial elements of which are drawn from the sixth sweep of the UK National Child Development Study (NCDS), suggest that we can draw a natural distinction between individuals who (as pure wage-workers) never become self-employed, and those who have ever been self-employed for any positive duration(s). The latter group might be termed the *Entrepreneurial Types* (ETs*),* some of whom may only be very briefly self-employed. The cross-section of those who are self-employed at any point in time within our sample period will be a proper subset of the ETs, since some individuals may only be self-employed before or after the date of the cross-section. The more inclusive ET set should thus provide more insight into the fundamental determinants of a propensity for self-employment.

In the second stage of our analysis, we then estimate the *total time spent in self-employment* by ET individuals between the ages of 33 and 42 – a measure of what we shall call *Entrepreneurial Persistence* (EP). This summary measure is dictated largely by data availability, and is clearly not a direct measure of survival or spell duration because we do not consider all the details on economic activity status, transitions and durations that constitute each individual's event history over our 9 year period. However, our measure is an indicator of entrepreneurial performance that should be correlated with the other measures such as job creation that we studied previously in a cross-section of NCDS individuals at age 33 – see Burke, FitzRoy and Nolan (2000, 2002). We should be able to provide a more insightful perspective about the factors that determine entrepreneurship, as well as being able to assess the extent to which pre-existing discrete analyses – such as those drawing on the fifth sweep (1991) of the NCDS (Blanchflower and Oswald (1998) and Burke, FitzRoy and Nolan (2000, 2002)) – have been distorted by oversimplification of the empirical analysis.

Our two-regime approach allows us to distinguish between characteristics that encourage individuals to try self-employment (ET), and those that are associated with longer total times in self-employment (or greater EP); the relevance of this distinction is ultimately an empirical question that is indeed confirmed by our data. We begin by modelling the count of quarters spent self-employed, using (two-regime) zero-inflated count data models that allow for there being a fundamental difference between those individuals who might be self-employed at some time, and others who never would countenance such an activity. Results for an alternative modelling approach – a 'hurdle' model, based on the work of Cragg (1971), Lin and Schmidt (1984) and Jones (1989) – are also estimated. These provide some indications of robustness for our conclusions about the impacts, for a cohort of individuals, of various aspects of an individual's background, experience and characteristics in determining time spent self-employed during nearly a decade of mature adult life.

As we point out in Section 1, economic theory offers only limited guidance. We proceed, in Section 2, to give an account of the econometric methodology to be used. Section 3 describes the data, drawn from the NCDS. In Section 4, we describe our empirical results for gender-specific two-regime

models on time spent self-employed, and offer our interpretation of their meaning. Conclusions are summarised in Section 5.

Section 1: Economic Background

Our central motivation stems from the fact that the empirical tests underlying models of self-employment (such as Kihlstrom and Laffont 1979, Jovanovic 1982, Evans and Jovanovic 1989, de Meza and Southey 1996, Blanchflower and Oswald 1998 and Burke, FitzRoy and Nolan 2000) can be improved in order to provide a more accurate and insightful perspective on both the determinants of self-employment as a career choice, and the subsequent time spent in self-employment. The conceptual background for our reduced-form empirical models is the dynamic programming problem under uncertainty faced by individuals with differing preferences and abilities for employment and self-employment. It is assumed that utility is affected by both the pecuniary, and the non-pecuniary, dimensions of each form of economic activity. For wage work, this is a standard approach – but it is worth elaborating a little for the case of self-employment. Income in self-employment is related to entrepreneurial ability (itself, influenced by innate and acquired human capital), access to resources (including finance) and the competitive environment in which the venture operates. Non-pecuniary factors can include the enjoyment of realising a vision, non-financial effects (e.g. helping others, promoting a philosophy or point of view), working in a sector or being a manager. Of course, non-pecuniary effects may be negative – such as the disutility from effort (much like that experienced in wage work) or the negative side effects of pursuing the chosen strategy (e.g. family costs, job displacement in other firms, damage to the environment, etc.). The inherent risk attached to the (pecuniary and non-pecuniary) returns from an activity, and the individual's attitude to risk, are also relevant.

In the first period, optimal choice of activity in employment or self-employment maximises the present value of expected discounted lifetime utility, given current knowledge about the effect of the initial decision on later career prospects. This effect can arise in many ways, through accumulation of

4

human capital interacting with abilities and preferences. In the next period, random shocks are realised and new information is acquired, and the new best choice may differ from the Period 2 plan that was made initially with less information. Decisions are thus made sequentially, and the resulting sequence of activities can be summarised by the integer EP, defined as the (possibly zero) number of periods spent in self-employment by each individual. EP is thus a function of the identifying vector of individual characteristics x_i in our dataset, and of the realisations of all the random shocks over the person's career during our overall time window. This picture contrasts with the stark representation often found elsewhere – which implies that individuals are either 100% pure wage-workers, or totally committed entrepreneurs. Roughly 9.7% of NCDS individuals spent some of our 9-year sample window in self-employment and another part of it in wage work (it is worthwhile to compare this to the 6.8% that spent the whole window in self-employment).

Choice models based on static utility functions, say of expected income and 'job' satisfaction in employment or self-employment, will generate either corner solutions or a unique interior optimum as in consumption theory, under the appropriate concavity assumptions. However, in our dynamic context, this standard approach can easily be misleading. Apart from a few part-time entrepreneurs who also hold regular jobs, most people who switch between the two modes do so at discrete intervals. Planned transitions, such as learning skills in employment and then transferring human and other capital into an entrepreneurial venture, are the dynamic equivalent to the interior optimum in static choice. However, all these cases suggest a sequence of corner solutions with at least initially increasing returns to duration in any activity, and transitions motivated primarily by a combination of expectations and shocks.

This pattern may result from quite a number of sources. Firstly, in view of the learning costs involved in any new activity, very short spell duration may be involuntary – the result of bad luck, over-optimism or error resulting in bankruptcy or redundancy. In such cases, the expected outcome from self-employment is worse than the actual outcome. Whatever its origins, this form of over-optimism may result in a pattern where individuals only learn the true value of the venture after

actually starting it. Along the lines of Jovanovic (1982), they can reverse their decision to become self-employed if overly disappointed. Thus, factors causing over-optimism, such as evangelical entrepreneurial role models, may be expected to increase the probability of an individual being an entrepreneurial type but have either negative or no impact on their persistence in self-employment.

Secondly, it is also plausible that a specific type of entrepreneurial ability may have a high rate of economic depreciation. This is especially likely if the business opportunity is short lived, or if the specific skills/knowledge of an individual (such as knowledge relating to a technology) are superseded in economic importance by other varieties. Thirdly, the non-pecuniary vision or purpose of the venture (such as proving to family/friends that one is capable of running a business) may be realised quite quickly. As a result, a one-time entrepreneur may want to move on to other goals in life – and these may involve wage work. Fourthly, under uncertainty, a move into self-employment may be a means of signalling managerial or other skills to employers in order to secure wage work once the 'true' value of the skills have been recognised. For example, this type of entrepreneurial activity is very common in media industries such a music, film and literature where employers find it hard to select high quality employees in the absence of seeing some demonstrable market performance (usually demonstrated through a start-up). Finally, an individual may choose to become self-employed as a means of acquiring business experience (such as managerial skills or knowledge of a business sector) that it may not be possible to acquire in wage work. As we know, only a tiny fraction of employees ever get an opportunity to undertake learning by doing in the role of CEO. Yet this is exactly what every entrepreneur can do – albeit, usually in a smaller firm. Thus, an individual whose wage work career requires experience in a sector or senior managerial role may find that 'barriers to learning' are less in self-employment. Thus, in such a case, a career path involves a transitory initial spell in self-employment. For example, employers in the venture capital and private equity industries frequently seek individuals with a successful prior experience in entrepreneurship. In sum, when one moves from a single- to a multi-period perspective on career choice between wage work and entrepreneurship, the process not only becomes richer but the dichotomous view of pure entrepreneurs versus wage-workers

becomes misleading. Instead, self-employment and wage work are interrelated career options, frequently feeding off each other in terms of access to finance (for start-up), human capital and signalling. This career choice process is less about dichotomy and more about flexibility.

Many of the characteristics which determine career choice are in binary form, represented by dummy variables in estimation, and as usual there is unobserved heterogeneity between individuals, as well as the random influences on choices at each stage. If an individual has most of the characteristics associated with ET, but chooses zero EP (no self-employment), then it may be reasonable to ascribe this choice realisation to chance and classify the individual as a potential entrepreneur in an extended ET set; the details of this procedure will be discussed in Section 2 below on econometric methodology. The entrepreneurship literature has established that self-employment income is influenced by entrepreneurial ability (in turn, determined through elements such as education, previous work experience, family background, and innate ability), available business opportunities and the cost and availability of capital. However, as we have outlined above, many of these same variables also affect wage work income e.g. education, work experience, self-employment experience, etc. Likewise, many of these same factors potentially affect non-pecuniary income in both wage work and self-employment e.g. education, parents' career, personality type and work experience. Lazear (2002) suggests that pure entrepreneurs are 'jacks of all trades' – in contrast, we argue that those who mix spells of self-employment and wage work, over a period of time, may be the ultimate 'jacks of all jacks of all trades'[4].

A key issue is the distinction that can be drawn between the traditional dichotomous/discrete approach, and our two-regime (ET, EP) framework. As already noted, the discrete approach misses those ETs who were self-employed at other times, and is generally more prone to severely misclassifying individuals. It also loses the persistence dimension of entrepreneurship. Furthermore, a specification test of the appropriateness of two-regime econometric models will actually confirm our

---

[4] Such individuals would, thus, be the most versatile of all – followed by pure entrepreneurs; while entrenched wage-workers (ceteris paribus) would be placed last in the versatility league.

approach against the alternative of the oft-used traditional binary choice logit or probit approach to self-employment.

In the two-regime model, it is also possible for an individual characteristic or element of the x_i vector to have different predicted effects on the ET and EP components. To classify these possibilities, recall that a binary variable such as ET is modelled econometrically as the *probability* of being self-employed or belonging to ET, say Pr(ET). For example, we find that higher education reduces the probability of self-employment or Pr(ET) for males, but does not effect their total time spent in self-employment for those who do make this choice, so it is insignificant in estimates of EP. Other variables raise Pr(ET) but have no influence on EP. Conversely, there are characteristics that seem to be irrelevant for ET, but have positive or negative effects on EP. We discuss these in more detail in section 4, which deals with the interpretation of the results of the econometric analysis.

With this background, we turn next to the issues of estimating these relationships and the appropriate econometric methodology – prior to our description of the data, and then the results of our econometric estimation.

Section 2: Econometric Methodology

If the values taken by a dependent variable are non-negative integers, perhaps with plenty of zeros, it is possible to improve on the simple least squares regression framework. For such count data (e.g., the number of workplace accidents in a year at a set of factories), the most straightforward alternative (see, for example, Greene (1997)) is the Poisson regression model. This is based upon the probability mass function of the Poisson distribution – a well-known discrete distribution with a single parameter, λ. The probability mass function for a Poisson random variable, Y, is:

$$P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}; \ y_i = 0, 1, 2, ... \tag{1}$$

For the Poisson regression model, it is usual to specify the natural logarithm of the λ parameter as a linear regression function, so that $\lambda_i = \exp(x_i' \beta)$, and estimation is usually by the method of Maximum

Likelihood (ML). However, this model suffers from the restricting underlying assumption that the mean of a Poisson variate is equal to its variance (and given by λ_i) – a restriction that is seldom satisfied in actual data, with overdispersion (variance exceeds mean) being more common in practice.

The negative binomial model, also estimated by ML, allows for the possibility of the variance and the mean of a dependent variable being unequal. In addition, it can accommodate heterogeneity among observed individuals (or entities, like factories) that is not captured by the regressors included in the $\boldsymbol{\beta}$ vector. This is an important feature of the model, since it is all too common for available regressors to have insufficient capacity to capture variations in a dependent variable. Defining $\mu_i = \lambda_i u_i = \exp(\mathbf{x}_i'\boldsymbol{\beta} + \ln(u_i))$, we can write a close parallel to (1) above as:

$$P(Y_i = y_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}; \quad y_i = 0,1,2,\dots \tag{1'}$$

These probabilities condition on both $\mathbf{x}_i$ and u_i here – rather than just on $\mathbf{x}_i$, as in (1). It is common to assume that u_i follows the gamma distribution, with an assumed mean of unity for identification if $\boldsymbol{\beta}$ includes a constant. The conditional mean of Y is again λ_i, but the variance depends on the parameter, θ, of the gamma distribution – and, using the definition $\kappa = 1/\theta$, the variance is $\lambda_i(1 + \kappa.\lambda_i)$.

However, in considering a case such as individual self-employment over a sample period, there may be two types of person – one that would never (seriously) consider becoming self-employed; and the other that would be willing, but was not observed to be self-employed at any point in the observed sample period. Such "hurdle" models are commonly (as in STATA 7.0) – labelled as the zero-inflated Poisson and negative binomial models. A binary choice model (logit or probit) is used to capture the difference between those who would never choose to be self-employed (thus inflating the number of zeros observed for the dependent variable), and those who might do so at least sometimes. For the zero-inflated Poisson model, with a logit component to inflate the number of zeros:

$$P(Y_i = 0) = \frac{1}{1+e^{w_i'\alpha}}\left(e^{w_i'\alpha} + e^{-\lambda_i}\right) = \frac{1}{1+e^{w_i'\alpha}}\left(e^{w_i'\alpha} + \exp(-\exp(\mathbf{x}_i'\boldsymbol{\beta}))\right), \tag{2a}$$

AND
$$P(Y_i = y_i) = \frac{1}{1+e^{w_i'\alpha}}\left(\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right); \quad y_i = 1,2,\dots \tag{2b}$$

For the zero-inflated negative binomial model:

$$P(Y_i = 0) = \frac{1}{1 + e^{w_i'\alpha}}\left(e^{w_i'\alpha} + e^{-\mu_i}\right) = \frac{1}{1 + e^{w_i'\alpha}}\left(e^{w_i'\alpha} + \exp\left(-\exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \ln(u_i)\right)\right)\right),$$ (3a)

AND $$P(Y_i = y_i) = \frac{1}{1 + e^{w_i'\alpha}}\left(\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}\right); \quad y_i = 1, 2, \dots$$ (3b)

The zero-inflated models are also estimated using ML, and it is usual to use robust standard errors when reporting results for these models.

Prediction of the mean of the dependent variable is straightforward. For example, under the zero-inflated negative binomial model:

$$E[Y_i] = 0.\frac{e^{w_i'\alpha}}{1 + e^{w_i'\alpha}} + \sum_{y_i=0}^{\infty}\left[\frac{y_i}{1 + e^{w_i'\alpha}}\left(\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}\right)\right].$$ (4b)

Accordingly, estimated marginal effects for each included regressor can be calculated – usually using estimated coefficients and the sample means of the regressors[5].

The question arises of which model is the most appropriate one to choose. The Poisson regression model is nested within the negative binomial model, and the extra restriction imposed by it may be tested straightforwardly by means of the standard Likelihood Ratio test. If there is overdispersion, this test will favour the negative binomial model. However, another possible source of excess zeros is the scenario where there are two types of person – so that zero-inflated models are appropriate. Vuong (1989) provided a two-sided test applicable when choosing between a pair of non-nested models – either Poisson versus zero-inflated Poisson, or negative binomial versus zero-inflated negative binomial. Asymptotically, the Vuong test statistic has a standard normal distribution.

We also consider an alternative modelling approach, which is not specifically designed for the case of count data, as a comparator. Although a Tobit model (censored regression) would be more applicable than standard least squares regression – given the impossibility of observing negative self-

---

[5] The fact that a two-regime model is being used means that a change of size Δ in the sample mean of a regressor within the $\boldsymbol{\beta}$ vector yields a change of $(\Delta n_2 / n)$ in the sample mean of that same regressor within the $\boldsymbol{\alpha}$ vector, where n_2 is the number of individuals of the self-employed type and n is the total number of individuals in the sample. However, we do not know n_2 (although we could estimate it via the estimated probabilities derived from the logit model).

employment durations – it can readily be improved upon if we accept the view that there may be a separate process determining zero and non-zero values of the self-employment duration (similarly to the zero-inflated models in a count data context) within some sample period. Let us use a probit model for the individual's choice between some self-employment and no self-employment and a Tobit model for the subsequent choice of non-zero time spent self-employed:

$$d_i^* = \mathbf{w}_i'\boldsymbol{\alpha} + u_i; \quad d_i = \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5a}$$

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + v_i; \quad y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5b}$$

Lin and Schmidt (1984) consider a model proposed by Cragg (1971), using equations like (5a) and (5b). If an individual chooses no self-employment via the probit model (5a), then equation (5b) – a truncated regression (with truncation to the left of zero) – is not relevant for that individual. Summation of the respective log-likelihoods of the univariate probit model and the truncated regression model yields the overall log-likelihood for the two-regime model. This combination reduces to the log-likelihood of a Tobit model if $\alpha = (\beta/\sigma)$, where σ is the standard deviation of the disturbance term v in (5b). While Lin and Schmidt (1984) derive an LM test of the Tobit model against the two-regime Cragg model, Greene (1997) points out that a simple Likelihood Ratio test is possible (as an asymptotically equivalent alternative).

In a paper on cigarette smoking by individuals, Jones (1989) considers several alternative model structures. These differ with respect to the independence (or otherwise) of the disturbance terms in equations like (5a) and (5b); and in whether the participation decision 'dominates' – such that a participant cannot be observed at the zero cut-off point (and only non-participants in self-employment can be observed with no time spent in self-employment during the sample). Slightly confusingly, the 'Cragg model' we have drawn from Lin and Schmidt (1984) is described by Jones (1989) as the 'First hurdle dominance' model. Its log-likelihood function, across a sample of n individuals, is as follows – using $\Phi(.)$ to denote the cumulative distribution function of the standard normal:

$$\ell = \sum_{i=1}^{n}\left[(1-d_i)\ln\Phi\left[-\mathbf{w}_i'\boldsymbol{\alpha}\right]+d_i\left[\ln\Phi\left[\mathbf{w}_i'\boldsymbol{\alpha}\right]-\ln\Phi\left[(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)\right]-\ln\left(\sigma\sqrt{2\pi}\right)-\tfrac{1}{2}\sigma^{-2}\left(y_i-\mathbf{x}_i'\boldsymbol{\beta}\right)^2\right]\right]. \quad (6)$$

The predicted unconditional mean number of quarters spent in self-employment comes from the following expression – using ϕ to denote the probability density function of the standard normal:

$$0.\Phi\left[-\mathbf{w}_i'\boldsymbol{\alpha}\right]+\left(\mathbf{x}_i'\boldsymbol{\beta}+\sigma\frac{\phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}{\Phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right)}\right)\Phi\left[\mathbf{w}_i'\boldsymbol{\alpha}\right] \quad (7)$$

Estimated marginal effects for each regressor can again be calculated using estimated coefficients and the sample means of the regressors[6].

Jones (1989) uses the 'Cragg model' title for a model in which the disturbance terms from (5a) and (5b) are independent[7]. Given the discussion underlying the various model structures in Jones (1989), we need to emphasise here that there are aspects of self-employment that enable a qualitative distinction to be drawn between those who would never be self-employed and those who might be, independent of how long a period in self-employment might last. Although cigarette smoking and self-employment are very different in many respects, attitudes to self-employment within society do also vary considerably – and this variation itself differs to an extent by social classification sub-group.

Jones (1989) also points out that participation in an activity at a point in time is determined by the fact that an individual has previously decided to commence that activity, and has also not quit from it – forming the basis for sample separation models that explicitly model the individual's decisions to commence and/or to quit. Blundell, Ham and Meghir (1987) note that such an approach should improve the efficiency of estimates. However, previous studies of individual self-employment at a particular time, such as Blanchflower and Oswald (1998) and Burke, FitzRoy and Nolan (2000, 2002), have not used this method. In any event, however, this paper needs a different approach – since we are examining self-employment over quite a long period of time (rather than at some given instant), and multiple cycles of starting and quitting are observed for some individuals over a nine-year period.

---

[6] Care is again needed because a change of size Δ in a regressor sample mean among the self-employed will lead to a smaller change in that regressor's sample mean across the whole sample – of $(\Delta n_2 / n)$.

[7] This model also has the Tobit model nested within it, provided the typo is corrected (as $\alpha_0 \to \infty, p(v > -\alpha'z) \to 1$).

Section 3: Data Description

The data used for our empirical analysis are taken from the National Child Development Study (NCDS). The NCDS has obtained information about a cohort of individuals born in the week from 3rd March, 1958 to 9th March 1958 inclusive and living in Great Britain. Following an initial study in 1958, a series of surveys has been undertaken at irregular intervals – in 1965, 1969, 1974, 1981, 1991 and 2000/01. This paper focuses on the number of *years, quarters or months* (complete or incomplete) spent by individuals in self-employment in the period between sweeps 5 and 6 of data collection. Nonetheless, we consider many regressors for inclusion that refer to the characteristics and background of the individual over the entire period of their life up to 2000/01.

In Table 1, below, we summarise the distribution of the number of quarters (periods of three months) of self-employment undertaken by NCDS individuals between March 1991 (NCDS5) and the time of the NCDS interview in 1999/2000 (variable between 102 and 113 months later). Note that incomplete quarters are counted – so that, for example, a period of 4 months is recorded as 2 quarters, while a period of 50 months is recorded as 17 quarters:

Table 1: Distribution of quarters of self-employment from 1991 – by gender.

	All		*Males*		*Females*	
Quarters Category	*Frequency*	*%*	*Frequency*	*%*	*Frequency*	*%*
0	9289	81.76	4200	75.07	5089	88.26
1 – 4	184	1.62	103	1.84	81	1.40
5 – 8	183	1.61	113	2.02	70	1.21
9 – 12	177	1.56	102	1.82	75	1.30
13 – 16	146	1.29	87	1.55	59	1.02
17 – 20	151	1.33	96	1.72	55	0.95
21 – 24	122	1.07	71	1.27	51	0.88
25 – 28	124	1.09	77	1.38	47	0.82
29 – 32	103	0.91	73	1.30	30	0.52
33 – 38	882	7.76	673	12.03	209	3.62
TOTAL	11361	100.00	5595	100.00	5766	100.00

It is important to note that Table 1 results from NCDS6 variables concerning main economic activity – so our data do not necessarily refer to a sole economic activity at a particular point in time. The frequency distributions display the key features we would expect – a substantial majority of individuals who are not self-employed at any stage over a period of close to a decade; more self-

employment among males than females; and a fairly small core of individuals who were self-employed throughout (less than half of those with some experience of self-employment in the years between NCDS5 and NCDS6):

Regressors to be considered for inclusion can be split into several categories, as follows:

1. *General controls* – a gender dummy (where the sample is not split by gender); a dummy for self-employment at age 23 (NCDS4); eight English region dummies (SW England is the base region) and separate dummies for Scotland and Wales, to capture NCDS5 region of residence data and control for variations in costs (particularly housing) and regional demand conditions.

2. *Family background* – a dummy captures non-white ethnicity; up to four dummies are used to capture the social class (class I, the base case, is top) of the cohort member's father in 1965 (NCDS1); several dummies are used to capture the occupation of the cohort member's father in 1965 (NCDS1); another dummy reflects family financial difficulties (NCDS1); a dummy is used to indicate use of the English language at home (NCDS2); another dummy indicates the cohort member's father to be self-employed in 1969 (NCDS2); two grouped variables from NCDS3 indicate the age at which the cohort member's father and mother left full-time education; another grouped variable indicates, for the cohort member's 1974 school, the percentage of male parents in a non-manual job; a dummy (NCDS5) indicates whether the cohort member's parents ever permanently separated or divorced.

3. *Education, ability and training* – there are dummies to indicate highest academic qualification (CSE, O level, A level, first degree or higher degree); four pairs of dummies capture performance in separate reading and maths tests at age seven (NCDS2) and age sixteen (NCDS3). For each test, a dummy is used to indicate a score definitively (not tied) in the top quintile of the cohort and another indicates a score in the bottom quintile – leaving the middle 60% (plus ties) of each ability distribution as the base case. A dummy variable captures embarkation by the cohort member on an apprenticeship by 1981; three others denote (respectively) receipt of vocational, professional and nursing qualifications by 1991.

14

4. *Non-cognitive attributes* – several psychological measures are included as discrete scores. Creativity comes from NCDS1 (1965) – a zero value denoting no creativity, and other values rescaled to a maximum of 0.4; while unforthcomingness, withdrawal, depression, anxiety acceptance and hostility towards (other) children are taken from NCDS2 (1969) – each with a zero minimum; and caution, moodiness, timidity, sociability and laziness measures are derived from NCDS3 (1974) – varying in the range [–2,+2]. There is a dummy for fear of new situations (1974). A number of dummies indicate the aspect that the cohort member regarded, in 1981 (NCDS4), as being most important when choosing a job. These include promotion, being in charge, being one's own boss, lack of responsibility, job security and good pay (cohort members responding with some other job characteristic form the base group).

5. *Financial* – real terms value of inheritance received by 1991 may enter both linearly (scaled in units of £10000) and quadratically (scaled by a factor of 10^{-10}), or as a dummy variable (above a threshold value level); the year in which inheritance was received (subtracting 1900 from the actual year, and then dividing by 100).

6. *Other* – a regressor is defined as the number of spells of unemployment undergone between March 1981 and being surveyed in 1991 (NCDS5); a dummy captures not having at least one child by 1991 – an alternative, related, measure being the number of children by 1981; another dummy indicates membership of a union or staff association in 1991 (NCDS5).

7. *Missing value dummies* – for some individual regressors, and some groups of regressors, an extra dummy is used to indicate that relevant data were missing, and as a (rather limited) control for this fact. The effects of sample attrition are more important if the attrition is non-random.

Section 4: Empirical Results

This section will present the main results. There is a need to consider predicted marginal effects, as well as the estimated coefficients. It is important to point out that background variables have substantial explanatory power in explaining labour market history up to the age of 33. Thus, it is unsurprising that the background variables themselves are often not separately significant in

(unreported) specifications that include a regressor capturing the percentage of the period from the end of education up to the 33rd birthday that is spent in employment. The explanatory power of such a regressor in determining labour market status between age 33 and age 42 is, itself, likely to be beyond dispute – and, thus, not particularly interesting. On the other hand, the effects of background characteristics from further back in an individual's life – which, to a considerable extent, feed into their developing labour market history – are rather less obvious, and correspondingly more intriguing.

4.1 Male self-employment:

Using the zero-inflated negative binomial model, it is worth noting firstly that there is an under-prediction of male self-employment – with only 78.04% of males predicted to have no period of self-employment between NCDS5 and NCDS6 (actually 75.07%) and a predicted average of 5.04 quarters self-employed (actually 6.30). As is shown in Table A1 (see Appendix 1), there is an unsurprisingly strong element of persistence in the tendency to be self-employed – with self-employment status from ten years prior to NCDS5 making 'zero inflation' less likely, and having a positive impact on the number of quarters spent self-employed between NCDS5 and NCDS6 by ET males. For an otherwise average male, the predicted number of quarters self-employed is 4.58 for an individual not self-employed at age 23, but 18.85 for one who was. The final two columns of Table A1 show that the actual conditional mean value of this regressor is nearly three times as high among those with some self-employment between NCDS5 and NCDS6, compared to the male group as a whole.

Other regressors that are predicted to have statistically significant effects in favour of both being an ET, and EP, include having a father who was an employee-manager in farming, having an apprenticeship by age 23, being creative (as measured way back at age 7) and having expressed the view at age 23 that being one's own boss is the most important aspect of a job. It should be pointed out that this model's predictions are of limited value for an individual with all of these attributes (but who is otherwise average) – since the predicted mean of 53.18 quarters exceeds the length of the

period between NCDS5 and NCDS6. Notably, only the control for union membership at age 33 has a statistically significant negative effect on being an ET, and EP.

The remaining regressor effects found for males in this model impact significantly on being an ET, or on EP, but not both. Thus, a male is more likely to be of ET if his father was the manager of a small firm, or if he had (himself) received by age 33 an inheritance of more than £30000 (in 1991 terms). He is predicted as less likely to be of ET if he possesses a first degree as his highest academic qualification, if he has a vocational qualification, if he was timid back at age 16, or if he viewed either promotion or job security as being the most important aspects of a job when asked in 1981 (aged 23). There are only two regressors that are statistically significant in only increasing EP – namely having had a father who was self-employed, and the value of an inheritance received by age 33. Of these effects, the magnitude of the latter is small – only 0.12 quarters for an otherwise average male in receipt of an inheritance worth £30000. The three regressors that correspondingly only reduce EP are having an O-level equivalent as highest academic qualification, being flexible back at age 16 (presumably too flexible) and having the view at age 23 that lack of responsibility is the most important aspect of a job.

The first hurdle dominance model confirms the expected strong effect of early self-employment on both being of ET, and EP through the period between NCDS5 and NCDS6. This model predicts the (unconditional) mean number of quarters for male self-employment more accurately than the zero-inflated binomial model, at 5.63 quarters (actually 6.30). For an otherwise average male, it also predicts the number of quarters self-employed as 4.96 for an individual not self-employed at age 23, and 19.32 for one who was – both of these predictions exceed their counterparts from the zero-inflated binomial model.

Results are also similar for the other regressors that were previously found to have positive effects on both aspects of self-employment – but we do find that, for the last two (creativity, and the desire to be one's own boss), the effect on EP is now statistically significant only at the 10% level. In another change relative to the zero-inflated negative binomial model, the value of an inheritance

received, in linear form, has a positive effect on both the probability of being of ET, and EP. However, the effect on EP is only significant at the 10% level, while the effect on the probability of being of ET is part of a broader effect of inheritance that also involves a negative quadratic effect and a negative effect of more recent inheritance[8]. The previously reported strong negative effect of union membership in 1991, on both aspects of self-employment between NCDS5 and NCDS6, remains in this model.

We again find that a male is more likely to be of ET if his father was the manager of a small firm, but we also now see a similar type of effect if his father was a farmer with his own account, or if he was viewed as lazy when aged 16. However, predictors for less likelihood of being of ET are the same as before, except that the effect of timidity at age 16 has disappeared. There is, again, a statistically significant positive effect on EP of having a self-employed father. Having an O-level equivalent as highest academic qualification and being flexible back at age 16 are still predicted to reduce EP, but the only other variables with a similar statistically significant effect in this model are high maths ability at age 16 and the number of unemployment spells suffered up to the age of 23.

Interpretation of results for male self-employment

Our two part econometric approach proves to be superior to the simple logit/probit and therefore, supports our reclassification of pure wage-workers and self-employed into pure wage-workers and entrepreneurial types. However, what is reassuring in terms of the existing trajectory of the empirical work is that our estimation of entrepreneurial types arrives at a specification which is broadly similar to previous studies – in the case of this data set to that of Blanchflower and Oswald (1998) and Burke, FitzRoy and Nolan (2000). In this sense it means that we can have a large degree of confidence in the existing literature in terms of understanding what makes someone of ET. Nevertheless, our results do unearth some substantial differences in terms of the final specification of ETs compared to EP. In fact, our results very clearly indicate that being of ET is driven by a quite different set of forces from EP. It follows that our results indicate that using probit/logit analysis to

---

[8] The t-ratio for this effect is -1.95 – so that it very narrowly fails to be statistically significant at the 5% level.

determine 'what makes an entrepreneur' is misleading if one is interested in persistent entrepreneurs rather than individuals who hope to get around to it some day or who try it only fleetingly. The significance of this difference is underlined by the observation earlier in the paper that, over the sample period, the population of individuals who move between self-employment and wage work is greater than those who are classified as entirely self-employed over the sample period. The results provide some interesting insights into this distinction – which, we believe, enriches our understanding of the process of what makes entrepreneurs.

Throughout our results for males, we note a high degree of consistency between the results from the zero-inflated and hurdle models (this feature also remains for females). In each form of estimation there is strong path dependence in terms of career choice early in each male's life. We find that if a male was self-employed at the age of 23 that he is not only more likely to be of ET over the ages 33-42 but is also more likely to persist in self-employment too. The results raise some issues for entrepreneurship education as they suggest that awareness of self-employment as a career path early in a male's life may be a key influence on an economy's long-term enterprise base. Likewise, it may also indicate that 'learning by doing' in self-employment early in a career can be a useful driver of entrepreneurial human capital and/or its specific nature may lock an individual into this form of career path.

Family background highlights some interesting intergenerational effects. A father who is self-employed or a manager of a small firm has a positive effect on his son being of ET. However, only a self-employed father has a positive effect on a son persisting in entrepreneurship. In terms of a human capital/mentor interpretation, this might indicate that there are valuable entrepreneurial skills – distinct from small business management skills – that only a self-employed father can pass onto a son. An alternative interpretation stems from a role model or 'influenced expected utility' effect where a father who is a manager of a small firm (without real experience of self-employment) may cause over-optimistic expectations (of the kind identified by de Meza and Southey (1996)) of utility from self-employment among their sons. If this is the case among a significant number of sons then they will not

persist in self-employment thereby generating insignificance (perhaps negating positive effects of mentoring by a father who is or was a small business manager) of the 'dad manager of a small firm' variable in the EP estimation. In contrast, self-employed fathers have real experience of self-employment and hence may pass on more realistic expectations of utility from self-employment to their sons – in which case the EP estimation is not affected by an outflow of those whose expected utility needed serious downward revision. Outside of these effects, the results seem to indicate that a dad who works in the farming sector has a positive impact on the son being of ET and persisting in self-employment. This effect may be due to sons of farmers being more likely to enter the farming sector than non-farmers' sons. In this case, with the high prevalence of small firms and self-employment in the farming sector one might well expect this pattern of econometric results.

The education variables, while different in composition, broadly reflect the interpretation of previous logit/probit estimates of the same dataset provided by Blanchflower and Oswald (1998) and Burke, FitzRoy and Nolan (2000). In general, higher levels of education are not associated with entrepreneurial types but low levels of education are negatively related to EP (or performance as in the case of Burke, FitzRoy and Nolan 2000). The same type of observation applies to the role of 'creativity' among psychological profiles of males. We find that creativity has both a positive effect on an individual being of ET (similar to Blanchflower and Oswald 1998, and Burke, FitzRoy and Nolan, 2000) and persisting in entrepreneurship (as found in the second stage estimation of performance in Burke, FitzRoy and Nolan, 2000). However, some other psychological profiles show an interesting distinction between ETs and EP. Notably, 'being cautious' is found to be a positive attribute of persistence in entrepreneurship which would make sense in terms of the impact of risk aversion on sample selection. However, this result is more interesting in light of the recent theory posited by Bhide (2000) who, on the basis of case study evidence, argued that entrepreneurs who ran high growth ventures were not typically risk prone but were more characterised by a 'heads I win, tails I do not lose very much' approach. In this light, our results provide some statistical support for Bhide's case study evidence. Less easy to interpret is the finding that 'being flexible' (usually believed to be of the

essence of entrepreneurship) appears to be negatively related to EP. This may reflect an inverse effect, namely that inflexible individuals may be more focused/persistent types who might be willing to see a venture through 'thick and thin' hence giving rise to the negative relationship between flexibility and persistence in entrepreneurship.

The role of finance as depicted through the exogenous measure of inheritance has similar effects to that outlined in studies such as Evans and Jovanovic (1989), Evans and Leigthton (1989), Blanchflower and Oswald (1998) and Burke, FitzRoy and Nolan (2000). Simply put, receipt of an inheritance increases the likelihood that a male will be of ET. In terms of persistence it is also found to be significant but the marginal effects show that it has only a minor role to play. This would seem to indicate that its effects are largely short term and are overtaken by other more pressing influences on the decision to persist in self-employment.

Finally, in terms of an auxiliary grouping of variables, some interesting results emerge. We find that having children seems to neither stimulate nor deter being of ETs and EP among males. Given that we find that we later find it has a negative effect on female EP, we think this is consistent with a view that despite changes in the labour market, females still bear the main economic burden of looking after children. Turning to the role of unemployment, we find that spells in unemployment do not appear to push individuals to become of ET and in fact appear to cause those who nonetheless choose to become self-employed to persist less in entrepreneurship. This result contrasts with the view originally put forward by Foreman-Peck (1985) – who, using UK data for the Interwar period, finds evidence of a push effect and speculates that these start-ups were more likely to be low quality. Our more direct evidence, for a more recent period, does not support the push hypothesis but does indicate that individuals with more early life experience of unemployment seem to have less staying power in entrepreneurship – being negatively related to EP. Thus, the results seem to indicate that unemployment has quite a disequilibrating role in the economy in terms of weakening the enterprise economy. This is in contrast to the push hypothesis literature (see Storey 1994 for an overview) which views unemployment as an equilibrating force in the economy.

Using the zero-inflated negative binomial model, there is also an under-prediction of female self-employment – with 90.42% of females predicted to have no period of self-employment (actually 88.26%) and a predicted average of 1.90 quarters self-employed (actually 2.45). The element of persistence in self-employment evident through the regressor 'self-employed at age 23' is also statistically strong here – though the effect is a little smaller in magnitude (1.86 quarters of self-employment predicted for an otherwise average female that was not self-employed aged 23, compared to 6.98 for an otherwise similar individual that was). There are fewer statistically significant effects in the case of females – and there are no other regressors that work significantly in favour of being of ET, and of EP. However, a nursing qualification acts against being of ET, and has a negative impact on EP (only 0.81 quarters of self-employment is predicted for an otherwise average female who was not self-employed at age 23, and possesses a nursing qualification).

A female is more likely to be of ET if her father was a manager of a small firm, a worker with his own account or an employee manager in farming. Her probability of being of ET is also positively linked to her father's age when leaving full-time education. Other regressors that have a similar effect include a professional qualification, an apprenticeship, being in the top quintile on mathematical ability at age 7 and viewing (at age 23) being one's own boss as the most important characteristic of a job. Unsurprisingly, a lower probability of being of ET is linked to union membership at age 33. Similar effects are also found for the English language being spoken at home (at age 11), being in the bottom quintile on reading ability at age 16, being cautious at age 16 and viewing job security as the most important aspect of a job (at age 23). The last two of these results in particular are very plausible, intuitively. The two regressors that only act to increase EP are a strong desire for a lack of responsibility in a job (not easily explained), and the timing of the receipt of an inheritance (closer to 1991, rather than less recently). We find four regressors that are linked to reduced female EP – namely, the mother's age of departure from full-time education, the number of unemployment spells endured since age 23, poor health at age 33 and the number of children borne by age 23.

As in the case of the zero-inflated negative binomial model, the first hurdle dominance model shows only past self-employment having a statistically significant positive effect on the probability of being of ET, and EP. Again, only the nursing qualification variable has a corresponding dual negative effect (although unforthcomingness comes quite close). This model again yields a higher predicted number of quarters of self-employment, at 1.99 (actually 2.45). It also predicts 1.93 quarters of self-employment predicted for an otherwise average female that was not self-employed aged 23, compared to 7.57 (a rather bigger difference from the zero-inflated negative binomial model's prediction) for an otherwise similar individual that was. Possession of a nursing qualification in conjunction with not being self-employed at age 23 yields a prediction of 0.81 quarters of self-employment (identical to the prediction from the zero-inflated model).

The set of seven regressors that predict females as more likely to be of ET is identical to its counterpart from the earlier model. So too is the corresponding set of five regressors that are linked to a lower probability of being of ET. Increased EP is again indicated by the receipt of an inheritance close to 1991, but it is also suggested for those females with a first degree as their highest academic qualification. We find that, in addition to the four regressors previously linked to reduced female EP, similar predicted effects are also present for being in Social Classes II, III and IV at age 7; and for having suffered a parental split (at some point up to the late boundary age of 33 years).

Interpretation of results for female self-employment

We find some stark differences between the female and male results – which, we believe, underlines the appropriateness of dividing the datasets (along the lines mooted in Burke, FitzRoy and Nolan, 2002). While males of ET and male EP can be largely explained within the confines of economics models of entrepreneurship augmented with psychological factors, the same approach is less satisfactory in explaining female entrepreneurship. Nonetheless, some generic features do emerge from the estimation process. As in the case of males, we find a strong degree of path dependence in terms of early career choice with females who were self-employed at the age of 23 also being more

likely to be of ET and persist in self-employment over the age 33-42. We deduce similar implications for entrepreneurship education to those outlined above for males.

In the case of the family background variables, we do not find that gender differences undermine the influence of the father's career on daughters. As in the case of males, we find that both fathers who are managers of small firms or self-employed (in the case of females only those who are 'worker own account') appear to have a positive impact on daughters being of ET. Moreover, as in the case of males, we find that the daughters of fathers who are managers of small firms are not any more persistent in self-employment than daughters without such a father. Thus, as before, we view this as either evidence of limited relevance of small firm managerial skills for persistence in entrepreneurship and/or evidence of a role model father causing over-optimism about self-employment utility among their daughters. However, the area where males and females diverge is that, unlike males, this same pattern also emerges for fathers who are self-employed in that their daughters do not appear to persist longer in self-employment than those who do not have a self-employed ('worker own account') father. As before, this might again be due to a father role model/mentor causing over-optimism (of the de Meza and Southey 1996 form) among daughters but it might also be due to key differences in human capital that are pivotal to typical male and female self-employment. Namely, the father's human capital may be less applicable to a daughter's career (compared to a son's) and hence mentoring by a father becomes less useful for females. An alternative viewpoint could be that the human capital transmission channel might be generally stronger from father-son than father-daughter. In other words, if fathers have closer and more communicative relationships with sons than daughters, then sons may receive a greater transfer of human capital from a father.

In the case of education, we note that higher levels of education – in the form of a first university degree – have a positive impact on EP. In this sense, the general pattern that education is good for persistence is similar to males. However, the pattern diverges in terms of determinants of being of ET – as university education has insignificant effects. We also find some polarised effects – with low levels of education (e.g. 'O level highest') appearing to be on the verge of significance in

terms of a stimulus to be of ET; while a high level of education, in the form of a professional qualification, does likewise.

In terms of psychology scores, creativity is not a driver of female entrepreneurship in the same way as it is for males. It is insignificant and on the verge of a negative effect in terms of EP. Cautious females tend to avoid self-employment, as do those who value job promotion. However, like males, a desire to be 'one's own boss' is positively related to being of ET – but, unlike males, it is not associated with EP.

In terms of the roles of finance and spells in unemployment, the difference between males and females only persists in the case of finance. Females are not stimulated to be of ET by receipt of an inheritance but are stimulated to persist longer in self-employment by such an event. Thus, the impact of an exogenous increase in access to finance appears to stimulate entrepreneurship among males and females, but in very different ways. In contrast, unemployment has similar effects in that it tends to decrease EP among both females and males. Poor health seems to constrain persistence in self-employment among females more than males while as we noted before, having children (by age 23) seems to only constrain persistence in female self-employment. Thus, overall we note some key areas of difference between female and males regarding both being of ET and EP. The extent of these gender differences justifies the treatment of male and female self-employment as distinct processes in separate equations – a practice not often observed in the previous literature on entrepreneurship.

4.3 Differences when self-employment is not disaggregated by gender:

In principle, it does not seem controversial to suggest that some detail will be lost under the traditional approach where data encompassed both males and females, and only a simple gender dummy was included to capture differences in outcome (as in Blanchflower and Oswald (1998) and Burke, FitzRoy and Nolan (2000)). Unreported results demonstrate that this is, indeed, true. In the majority of cases, but by no means all, the estimated coefficient for males seems to dominate. On the whole, then, one of the greater dangers of retaining aggregation across the genders is to incorrectly

suppose that certain regressors have an impact on aspects of female self-employment, when in fact they do not[9].

Section 5: Conclusion

The paper has aimed to contribute to the literature on entrepreneurial choice by moving beyond a dichotomous two-dimensional view of the labour market where individuals are classified wholly as either wage-workers or entrepreneurs. We note that the majority of entrepreneurs actually spend some of their career in wage work and hence we have distinguished between entrepreneurial types (individuals who either at some stage have been self-employed or if not, would consider self-employment as a career option) and entrepreneurial persistence. We outline how the theoretical considerations governing entrepreneurial choice become richer when this distinction is considered. We also offer an econometric approach to account for and test the appropriateness of this classification. Using the latest update to the NCDS dataset we then undertake analyses of the determinants of being of ET (individuals with an inclination for entrepreneurship), and of EP, for both males and females across a 9-year period – from age 33-42. The results indicate that this approach is indeed a superior specification to the simpler probit/logit dichotomous approach. The results have important ramifications because we find a very stark difference between factors that encourage individuals to try self-employment and those that determine persistence in self-employment. We find that the determinants of being of ET are similar to the results in the existing empirical literature based on a probit/logit estimation of self-employment choice. Given the superiority of our econometric approach, this finding is reassuring as it means that the pre-existing literature on self-employment choice is indeed a good guide to explaining what makes an individual of ET. However, our results for EP are very different and taking these results in conjunction with our results on entrepreneurial types provides an enriched understanding of what makes an entrepreneur.

---

[9] The obvious alternative of including gender interaction terms (for many regressors) is likely to suffer from being rather difficult to readily interpret. Results tables would also further increase in size, given that each of our models has two parts (one to capture whether an individual is of ET, and the other to gauge EP).

We find that male and female entrepreneurship are not governed by a common economic model – although there are some common determinants. One of these is an early career experience of self-employment – which is associated with path dependence and tends to encourage persistent entrepreneurship throughout the 33-42 career span. Similarly, higher (or not lower) levels of education tend to be associated with EP among both males and females. In terms of access to finance, we find that it encourages being of ET among males with only marginal effects on EP. In the case of females it only has the effect of increasing EP. In terms of fathers acting as mentors and role models we also find some path dependence as self-employed fathers tend to encourage more entrepreneurial types and more persistence among their sons. In the case of females, this effect only occurs in terms of encouraging ET. We believe the difference may be due to closer relationships between fathers and sons than fathers and daughters – perhaps, indicating more relevance of a father's human capital for a son's (rather than a daughter's) business. We also find that fathers who are managers of small firms encourage both sons and daughters to be of ET but have negligible effects on EP. We believe this may be due to a misalignment of skills for self-employment and small business management and/or small business managers causing their offspring to have over-optimistic views (of the kind identified by de Meza and Southey (1996)) of prospects in self-employment. We believe that this effect does not occur for self-employed fathers in the case of males because, as one might expect, these fathers have real experience of self-employed and hence might be expected to provide their sons with a more realistic perspective. This realism might be overtaken by a lack of active mentoring in the case of females – hence undermining the positive effect on persistence.

Our results are also consistent with a view of children acting as a greater hindrance on a female's entrepreneurial career than that of a male. Having children by the age of 23 has no statistically significant effect on male entrepreneurship but is found to be negatively related to persistence in self-employment among females. We also find that persistence in self-employment is hindered by poor health among females more than males.

Finally, our analysis sheds some interesting light on the role of unemployment on entrepreneurship. It provides a much more disequilibrating and negative view of unemployment than previously depicted in the entrepreneurship literature. We find that spells in unemployment do not swell the number of those of ET among males and females and actually causes a decrease in EP. Thus, our research does not support the view of there being an 'unemployment push' into self-employment, thereby raising the supply of entrepreneurs. In contrast, our analysis indicates that the number of entrepreneurs is not increased and arguably their quality declines as early life experience in unemployment reduces EP.

In sum, we believe the paper has provided a richer and more insightful depiction of the entrepreneurial process through new empirical findings based on new data, a new theoretical perspective and an econometric methodology to test and investigate it. We believe this is only a first step beyond the dichotomous pure wage-worker versus pure entrepreneur logit/probit approach towards a more multi-dimensional and dynamic view of the determinants of what makes an entrepreneur.

References

Bhide, A. (2000), *The Origin and Evolution of New Businesses*, Oxford University Press, New York.

Blanchflower, D. and Oswald, A. (1998), What makes an entrepreneur? *Journal of Labor Economics*, 16: 26-60.

Blundell, R., Ham, J. and Meghir, C. (1987), Unemployment and female labour supply, *Economic Journal*, 97 (Conference Supplement), 44-64.

Burke, A.E., FitzRoy, F.R. and Nolan, M.A. (2000) When less is more: distinguishing between entrepreneurial choice and performance, *Oxford Bulletin of Economics and Statistics*, 62: 565-587.

Burke, A.E., FitzRoy, F.R. and Nolan, M.A. (2002) Self-employment wealth and job creation: the roles of gender, non-pecuniary motivation and entrepreneurial ability, *Small Business Economics,* 19(3): 255-270.

Cragg, J.G. (1971) Some statistical models for limited dependent variables with application to the demand for durable goods, *Econometrica*, 39: 829-844.

de Meza, D. and Southey, C. (1996), The borrower's curse: Optimism, Finance and Entrepreneurship, *Economic Journal*, vol. 106: 375-386.

Evans, D. and Leighton, L.S., (1989), Some Empirical Aspects of Entrepreneurship, *American Economic Review*, 79: 519-535.

Evans, D. and Jovanovic, B. (1989), An Estimated model of Entrepreneurial Choice under Liquidity Constraints, *Journal of Political Economy*, vol. 97: 519-535.

Foreman-Peck, J.S. (1985), Seedcorn or Chaff? New Firm Performance in the Interwar economy, *Economic History Review*, 38(3): 402-422.

Greene, W.H. (1995) *User's manual, LIMDEP, version 7.0 (for PC under DOS, & for mainframe)*, Econometric Software, Inc.

Greene, W.H. (1997) *Econometric analysis*, New Jersey, Prentice-Hall (3rd edition).

Greene, W.H. (2002) *Reference Guide and Econometric Modeling Guide, LIMDEP, version 8.0 (for PC under Windows)*, Econometric Software, Inc.

Heckman, J.J. (1979) Sample selection bias as a specification error, *Econometrica*, 47: 153-161.

Jones, A.M. (1989) A double-hurdle model of cigarette consumption, *Journal of Applied Econometrics*, 4: 23-39.

Lin, T-F. and Schmidt, P. (1984) A test of the Tobit specification against an alternative suggested by Cragg, *The Review of Economics and Statistics*, 66: 174-177.

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

Storey, D. J. (1994), Understanding the Small Business Sector, Thomson Learning, London.

Vuong, Q. (1989) Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, 57: 307-334.

White, H. (1978) A heteroskedasticity consistent covariance matrix and a direct test for heteroscedasticity, *Econometrica*, 46: 817-838.

Table A1: Zero-inflated negative binomial model for males.

Variable	INFLATION (logit)		COUNT (negbin)		Mean	Mean, EP > 0
	Estimate	Est./S.E.	Estimate	Est./S.E.	Mean	EP > 0
Constant	0.938	4.54	3.084	67.22	1.000	1.000
GENERAL CONTROLS (10 region dummies are also included in the inflation part only)						
Self-employed at age 23	-1.895	-12.97	0.282	10.36	0.054	0.155
FAMILY BACKGROUND						
Non-white ethnicity	-0.443	-1.64	0.128	1.57	0.015	0.021
Dad manager, small firm	-0.388	-3.34			0.087	0.122
Dad professional employee	-0.246	-1.53			0.044	0.047
Dad worker, own account	-0.362	-1.76			0.025	0.035
Dad farmer employee manager	-0.753	-2.61	0.182	2.93	0.010	0.022
Dad self-employed			0.148	3.12	0.043	0.061
Mum's time in education	-0.055	-1.79			2.840	2.867
EDUCATION, ABILITY AND TRAINING						
O Level highest			-0.072	-2.34	0.316	0.331
A Level highest	0.188	1.39			0.086	0.082
First degree highest	0.358	2.97			0.114	0.097
Higher degree highest	0.513	1.80	-0.174	-1.17	0.019	0.015
Vocational qualification	0.184	2.17			0.399	0.399
Professional qualification			0.078	1.59	0.095	0.093
Other qualification	0.123	1.35			0.202	0.188
Apprenticeship (by 1981)	-0.157	-3.32	0.083	5.24	0.544	0.598
Maths Low aged 7	0.244	2.10			0.120	0.098
Maths Low aged 16	-0.358	-3.18			0.107	0.118
NON-COGNITIVE ATTRIBUTES (MEASURED IN THE PAST)						
Creativity	-1.428	-3.04	0.404	2.21	0.165	0.170
Unforthcomingness	0.034	1.81			1.365	1.226
Caution			0.024	1.40	0.183	0.117
Flexibility			-0.048	-2.35	0.164	0.148
Timidity	0.112	2.13			0.013	-0.044
Promotion	0.385	2.71			0.074	0.060
Being one's own boss	-0.481	-4.25	0.074	2.16	0.093	0.158
Lack of responsibility			-1.367	-7.53	0.003	0.001
Job security	0.278	3.12			0.253	0.194
FINANCIAL						
Inheritance > £30000	-0.486	-2.47			0.028	0.039
Value of inheritance (linear)			0.003	2.37	0.469	0.742
Year of inheritance	0.159	1.47			0.197	0.194
OTHER						
Union at age 33	1.260	13.89	-0.186	-4.20	0.303	0.134
Unemployed spells, 1981-1991			-0.023	-1.65	0.402	0.396
Poor health at age 33	0.878	2.06	-0.308	-1.40	0.012	0.007
Childless aged 33			-0.038	-1.18	0.269	0.266
ln (κ)	Not applicable		-1.191	-6.44		

There are 5595 observations, of which 4200 have no time spent in self-employment. The unconditional mean of the dependent variable is 6.301, while the conditional mean is 25.271. The maximum of log-likelihood, using a logit for the inflation model, is –8341.664. The Vuong test has a test statistic of 23.93. Removal of the region dummies from the logit yields a test statistic of 32.99, significant at the 1% level.

Table A2: Zero-inflated negative binomial model for females.

Variable	INFLATION (logit) Estimate	Est./S.E.	COUNT (negbin) Estimate	Est./S.E.	Mean	Mean, EP > 0
Constant	1.807	5.00	3.204	28.92	1.000	1.000
GENERAL CONTROLS (10 region dummies are also included in both parts of the model)						
Self-employed at age 23	1.326	5.10	0.229	2.58	0.016	0.059
FAMILY BACKGROUND						
Social class V (1965)	0.485	1.85			0.052	0.025
Family financial problems			-0.251	-1.83	0.062	0.041
English spoken at home	0.526	2.12			0.809	0.821
Dad manager, small firm	-0.319	-3.28			0.087	0.130
Dad professional employee	-0.277	-1.42			0.036	0.059
Dad worker, own account	-0.694	-3.09			0.026	0.046
Dad farmer employee manager	-0.725	-2.07			0.010	0.019
Mum's time in education			-0.045	-2.63	2.893	3.170
Dad's time in education	-0.063	-2.12			2.785	3.164
Male parents non-manual	-0.033	-1.42			2.837	3.242
EDUCATION, ABILITY AND TRAINING						
O Level highest	-0.179	-1.68			0.379	0.391
A Level highest	-0.209	-1.40			0.102	0.129
First degree highest			0.135	1.83	0.110	0.139
Professional qualification	-0.390	-2.55			0.068	0.105
Nursing qualification	0.561	2.32	-0.359	-2.29	0.059	0.031
Other qualification	-0.158	-1.46			0.230	0.229
Apprenticeship (by 1981)	-0.319	-3.28			0.076	0.146
Maths High aged 7	-0.270	-2.39			0.167	0.236
Maths Low aged 7	0.225	1.44			0.134	0.087
Maths High aged 16			0.101	1.49	0.134	0.176
Maths Low aged 16	-0.242	-1.61			0.143	0.121
Reading High aged 16	-0.222	-1.62			0.102	0.152
Reading Low aged 16	0.363	2.04			0.129	0.074
NON-COGNITIVE ATTRIBUTES (MEASURED IN THE PAST)						
Creativity	-0.913	-1.54	-0.472	-1.44	0.174	0.192
Unforthcomingness	0.042	1.67			0.180	0.160
Hostility to other children			0.078	1.78	1.274	1.075
Caution	0.156	2.71			0.184	0.108
Flexibility	-0.108	-1.68			0.194	0.270
Promotion	0.389	1.75			0.046	0.040
Being one's own boss	-0.624	-3.24			0.028	0.072
Lack of responsibility			0.326	2.47	0.005	0.004
Job security	0.408	2.77	0.114	1.44	0.139	0.096
FINANCIAL						
Year of inheritance			0.154	2.54	0.199	0.244
OTHER						
Union at age 33	0.979	7.20			0.212	0.105
Unemployed spells, 1981-1991			-0.109	-3.05	0.295	0.298
Poor health at age 33			-0.900	-2.71	0.015	0.013
# Children when aged 23			-0.118	-3.23	0.412	0.321
ln (κ)	Not applicable		-0.870	-5.20		

There are 5766 observations, of which 5089 have no time spent in self-employment. The unconditional mean of the dependent variable is 2.446, while the conditional mean is 20.829. The maximum of log-likelihood is –4526.469.

Table A3: First hurdle dominance model for males.

Variable	Participation (probit) Estimate	Est./S.E.	EXTENT (truncreg) Estimate	Est./S.E.	Mean	Mean, EP > 0
Constant	-0.643	-5.24	21.218	16.51	1.000	1.000
GENERAL CONTROLS						
Self-employed at age 23	1.110	13.00	8.067	8.02	0.054	0.155
FAMILY BACKGROUND						
Non-white ethnicity	0.271	1.72	3.264	1.27	0.015	0.021
Dad manager, small firm	0.252	3.66			0.087	0.122
Dad professional employee	0.158	1.65			0.044	0.047
Dad worker, own account	0.240	1.98	-4.227	-1.47	0.025	0.035
Dad farmer employee manager	0.460	2.54	5.097	2.14	0.010	0.022
Dad farmer own account	0.662	3.44			0.008	0.018
Dad self-employed			6.429	2.90	0.043	0.061
Mum's time in education	0.029	1.65			2.840	2.867
EDUCATION, ABILITY AND TRAINING						
O Level highest			-1.660	-2.04	0.316	0.331
First degree highest	-0.155	-2.23			0.114	0.097
Higher degree highest	-0.207	-1.30			0.019	0.015
Vocational qualification	-0.102	-2.07			0.399	0.399
Professional qualification			2.348	1.84	0.095	0.093
Other qualification	-0.069	-1.31			0.202	0.188
Apprenticeship (by 1981)	0.101	3.72	2.159	4.84	0.544	0.598
Maths High aged 7			1.424	1.50	0.194	0.202
Maths Low aged 7	-0.136	-2.06			0.120	0.098
Maths High aged 16			-2.126	-1.98	0.195	0.182
Maths Low aged 16	0.191	2.89			0.107	0.118
NON-COGNITIVE ATTRIBUTES (MEASURED IN THE PAST)						
Creativity	0.885	3.24	9.767	1.93	0.165	0.170
Unforthcomingness	-0.017	-1.60			1.365	1.226
Caution			0.972	1.97	0.183	0.117
Flexibility			-1.438	-2.36	0.164	0.148
Laziness	0.042	2.14			-0.107	-0.027
Moody			0.517	1.31	-0.441	-0.396
Sociability	0.030	1.36	0.560	1.24	0.446	0.490
Promotion	-0.221	-2.74			0.074	0.060
Being one's own boss	0.281	4.20	1.690	1.71	0.093	0.158
Lack of responsibility	-0.841	-1.76	-28.332	-1.21	0.003	0.001
Job security	-0.157	-3.07			0.253	0.194
FINANCIAL						
Value of inheritance (linear)	0.031	3.81	0.082	1.62	0.469	0.742
Value of inheritance (squared)	-0.011	-3.87			0.337	0.487
Year of inheritance	-0.118	-1.95			0.197	0.194
OTHER						
Union at age 33	-0.710	-14.43	-5.036	-4.61	0.303	0.134
Unemployed spells,1981-1991			-0.912	-2.27	0.402	0.396
Poor health at age 33	-0.457	-2.08	-7.747	-1.70	0.012	0.007
# Children aged 23	0.054	1.32			0.181	0.195
Childless aged 33			-1.175	-1.37	0.269	0.266
σ	Not applicable		12.226	43.01		

Respective maxima for log-likelihoods are -2752.385 and -5340.912. If a common set of regressors is used in both parts, the Likelihood Ratio test statistic against the Tobit is 736.4.

Table A4: First hurdle dominance model for females.

Variable	Participation (probit)		EXTENT (truncreg)		Mean	Mean, EP > 0
	Estimate	Est./S.E.	Estimate	Est./S.E.	Mean	EP > 0
Constant	-1.081	-5.62	33.855	8.90	1.000	1.000
GENERAL CONTROLS (including 10 region dummies in the participation probit only)						
Self-employed at age 23	0.798	5.32	6.032	2.29	0.016	0.059
FAMILY BACKGROUND						
Social class II (1965)			-7.548	-2.79	0.131	0.199
Social class III (1965)			-5.923	-2.33	0.472	0.448
Social class IV (1965)			-7.328	-2.45	0.150	0.133
Social class V (1965)	-0.235	-1.86			0.052	0.025
Family financial problems			-6.471	-1.76	0.062	0.041
English spoken at home	-0.277	-2.06			0.809	0.821
Dad manager, small firm	0.177	2.38			0.087	0.130
Dad professional employee			-6.095	-1.89	0.036	0.059
Dad worker, own account	0.377	3.03			0.026	0.046
Dad farmer employee manager	0.391	1.98			0.010	0.019
Mum's time in education			-1.345	-2.59	2.893	3.170
Dad's time in education	0.035	2.21			2.785	3.164
Male parents non-manual	0.020	1.61			2.837	3.242
Parental split by 1991			-4.553	-2.25	0.141	0.131
EDUCATION, ABILITY AND TRAINING						
O Level highest	0.092	1.63	2.081	1.37	0.379	0.391
A Level highest	0.115	1.43			0.102	0.129
First degree highest			4.391	1.97	0.110	0.139
Professional qualification	0.196	2.31			0.067	0.105
Nursing qualification	-0.271	-2.27	-11.285	-2.69	0.059	0.031
Other qualification	0.084	1.47			0.230	0.228
Apprenticeship (by 1981)	0.182	3.30			0.076	0.146
Maths High aged 7	0.154	2.53			0.167	0.236
Maths Low aged 7	-0.118	-1.51			0.134	0.087
Maths Low aged 16	0.135	1.75			0.143	0.121
Reading High aged 16	0.111	1.47	2.750	1.30	0.102	0.152
Reading Low aged 16	-0.178	-2.01			0.129	0.074
NON-COGNITIVE ATTRIBUTES (MEASURED IN THE PAST)						
Creativity	0.410	1.29	-13.651	-1.46	0.174	0.192
Unforthcomingness	-0.023	-1.76	-0.602	-1.62	1.274	1.274
Caution	-0.098	-3.17			0.184	0.184
Flexibility	0.048	1.39			0.194	0.194
Laziness	-0.042	-1.68			-0.355	-0.042
Promotion	-0.215	-1.89			0.046	0.040
Being one's own boss	0.359	3.19			0.028	0.072
Job security	-0.200	-2.70			0.139	0.096
Good pay			-4.289	-1.70	0.084	0.078
FINANCIAL						
Year of inheritance			4.017	2.30	0.199	0.244
OTHER						
Union at age 33	-0.516	-7.70			0.212	0.105
Unemployed spells, 1981-1991			-3.303	-3.17	0.402	0.298
Poor health at age 33			-22.880	-2.36	0.015	0.013
# Children when aged 23			-3.264	-2.36	0.412	0.321
σ	Not applicable		13.743	24.24		

Respective maxima for log-likelihoods are -1893.949 and -2568.746. If a common set of regressors is used in both parts, the Likelihood Ratio test statistic against the Tobit is 271.6.