

What makes a die-hard entrepreneur? Beyond the ‘employee or entrepreneur’ dichotomy

Andrew E. Burke · Felix R. FitzRoy · Michael A. Nolan

Accepted: 24 October 2007 / Published online: 16 January 2008
© Springer Science+Business Media, LLC. 2008

Abstract The article 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 and as a result avoids the weakness of approaches which categorise an individual as an entrepreneur by observing their occupation at just one point in their career. We outline an econometric methodology to account for this approach and find, via a statistical test of model selection, 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 to *try* self-employment, we find that entrepreneurial *persistence* is determined by a different model and unearth some new insights.

Early self-employment encourages entrepreneurial persistence. For men, inheritance encourages persistence, and facilitates initial self-employment. Having a self-employed father as a role model makes sons persist longer. However, somewhat surprisingly, early experience of unemployment does not affect the probability of self-employment, while reducing persistence. The popular ‘unemployment push effect’ is thus rejected in our sample.

Keywords Self-employment · Entrepreneurial persistence · Count data

JEL Classifications J23 · C25 · L26

A. E. Burke
Cranfield University, Cranfield, Bedford MK43 0AL, UK

A. E. Burke
Max Planck Institute for Economics, Kahlaische Straße
10, Jena 07745, Germany

F. R. FitzRoy
University of St. Andrews, St. Andrews KY16 9AL, UK

F. R. FitzRoy
IZA, Bonn, Germany

M. A. Nolan (✉)
Department of Economics, University of Hull,
Cottingham Road, Hull HU6 7RX, UK
e-mail: m.a.nolan@hull.ac.uk

1 Introduction

Extensive research over the past 25 years has looked for characteristics that drive individuals to choose self-employment instead of wage work. Blanchflower and Oswald (1990, 1998) define the typical research question in this research area as ‘what makes an entrepreneur?’ Their later work notes that “The simplest kind of entrepreneurship is self-employment,” and we follow these authors, and many others, by using ‘entrepreneur’ and ‘self-employed’ interchangeably. In the literature 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), Evans and Jovanovic (1989) and Blanchflower and Oswald (1990), 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 is an area that still awaits detailed research. 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 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 (Holtz-Eakin et al. (1994)). Likewise, some individuals are serial entrepreneurs and some of these spend temporary periods in employment between ventures; especially those who are unsuccessful (for example, see Handy (1999) and Bridge (2006)). 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 all of their time in either wage-work or self-employment. However, labour market data indicates that this is not only inaccurate, but also actually very misleading. Over a 9-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. In fact, patterns in the data suggest three types of people (although we only directly consider a period of 8–9 years): those who never try self-employment, those who are ‘die-hard’ entrepreneurs and spend *their entire* career in self-employment, and those who move between wage work and self-employment. This pattern in the data has prompted us to delve deeper in to the question of what makes an entrepreneur in order to distinguish between the ‘die-

hards’ and the ‘less persistent’ entrepreneurs. Therefore, the aim of this article is to try to move beyond the dichotomous depiction of entrepreneurship and wage-work, and begin to explore the implications of entrepreneurial persistence. We want to differentiate the factors that make an individual *try* self-employment from those that make a persistent, dedicated or what we term a ‘*die-hard*’ entrepreneur. It is worth noting that even successful serial entrepreneurs may be associated with rapid change of ventures, giving rise to short-lived businesses alongside high persistence in self-employment. Similarly, relatively unsuccessful entrepreneurs who enjoy the non-pecuniary benefits of self-employment may still choose to spend long periods in self-employment. Therefore, at a theoretical level our concept of entrepreneurial persistence is not synonymous with firm survival.¹ However, at an empirical level survival and persistence are likely to be highly related as in this data nearly 90% of the time spent in self-employment is accounted for by a single continuous spell in self-employment. To our knowledge this is the first empirical analysis of the determinants of ‘die-hard’ entrepreneurs—although growth studies conditioned on the survival of a self-employed business fulfil a similar role. Before proceeding to outline the structure of the rest of the article we first spend a little time explaining terms associated with our reclassification of entrepreneurs from a single ‘pure’ type to a form which accounts for varying degrees of persistence (or ‘die-hardness’) in self-employment.

Not to overstate our contribution, here 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. This summary measure can obviously include multiple spells, and thus does not address spell durations as such, but as we have noted that is not the purpose of the article.² The complementary measure “time not self-employed” includes employment, unemployment

¹ Cressy (1996) uses the Evans and Jovanic (1989) model as a basis for examining the survival of start-ups—measured by the non-closure within four years of a business account opened in 1988.

² It is not a formal duration analysis of business survival of the sort undertaken by Van Praag (2003).

and inactivity—though for many (particularly among the males in our chosen cohort) it will be mainly employment; and, in any case, we include basic controls to provide crucial distinctions between those who are employed and those who are not.

Once we proceed beyond considering individuals only at a point in time, we could in principle examine all individuals on a continuum between choosing to spend their entire career history in self-employment, or none of it, or any fraction. However, the polarised nature of our data, the crucial elements of which are drawn from the sixth sweep of the UK National Child Development Study (NCDS), suggest that we ought to test whether a natural distinction exists between individuals who are pure wage-workers (no self-employment in their work history or likely future), and those who have ever been self-employed (or may be in the future). We label this latter group *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)—the highest levels of EP being for the die-hard entrepreneurs. This summary measure is dictated largely by data availability, and is clearly not a direct measure of survival or spell duration, although we note that for nearly 90% of the sample this is indeed the case. 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 et al. (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 et al. (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 (be of ET), and those that are associated with longer total times in self-employment (or greater EP)—the longest being the ‘die-hard’ entrepreneurs. 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 a fundamental difference between those individuals who might be self-employed at some time, and others who never would. 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 their background, experience and characteristics in determining time spent self-employed during nearly a decade of mature adult life.

There have been some other recent methodological advances. The work of Constant and Zimmerman (2006) uses a three-stage estimation approach, culminating in a structural probit that allows for the endogeneity of the role of earnings. Fraser and Greene (2006) employ heteroscedastic probit estimation of occupational choice, on the basis that entrepreneurial optimism diminishes with length of self-employment experience. Henley (2004) analyses panel data for self-employment by means of a two-step method proposed by Orme (1999)—which takes into account the ‘initial conditions’ problem, and identifies genuine state dependence by explicitly modelling unobserved heterogeneity. Parker and Belghitar (2006) utilise a multinomial logit model to investigate nascent entrepreneurship, and include a selectivity term to control for possible non-random attrition bias.

As we point out in Sect. 1, economic theory offers only limited guidance on modelling the likely determinants of self-employment. We proceed, in Sect. 2, to describe the data, drawn from the NCDS. Section 3 gives an account of the econometric methodology to be used. In Sect. 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 Sect. 5.

2 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 et al. (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.³ Income in self-employment is related to entrepreneurial ability (itself, influenced by innate and acquired human capital), access to resources (including finance)⁴ and the competitive environment in which the venture operates. Non-pecuniary factors could 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 (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 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% 'die-hards' that spent the whole window in self-employment because in effect this is the group most people have in mind when they think of a person who is an entrepreneur. However, to our knowledge this is the first article to attempt to isolate and estimate what makes this particular ('die-hard') type of entrepreneur.

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.

³ This inter-temporal approach is similar in spirit to the work of Ghatak et al. (2001).

⁴ There has been considerable debate about the role of access to finance for entrepreneurs—for example, see the comprehensive review by Parker (2004).

This pattern may result from quite a number of sources. First, 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.⁵ 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.

Second, 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. Third, 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 might involve wage work. Fourth, 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 that 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 Sect. 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 (2004) suggests that pure entrepreneurs are ‘jacks of all trades’—in contrast, we argue that those who mix

⁵ In our dataset, only for 8.8% of 752 periods of full-time self-employment was firm closure the primary reason given for leaving self-employment—and only for 3.2% of 216 periods of part-time self-employment.

spells of self-employment and wage work, over a period of time, may be the ultimate ‘jacks of all trades’.⁶

A key issue is the distinction that can be drawn between the traditional dichotomous and 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 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 $\text{Pr}(\text{ET})$. For example, we find that higher education reduces the probability of self-employment or $\text{Pr}(\text{ET})$ for males, but does not affect 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 $\text{Pr}(\text{ET})$, 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 Sect. 4, which deals with the interpretation of the results of the econometric analysis.

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

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 1999/2000. This article 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 1999/2000.

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).⁷ 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 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 die-hard individuals who were self-employed throughout (less than half of those with some experience of self-employment in the years between NCDS5 and NCDS6). Figure 1 provides an illustration of the relative frequencies and provides some motivation for our central research question. It shows three groups, comprising a group who have mixed their career between wage work and self-employment walled in at either tail by die-hard entrepreneurs on the right and pure wage workers on the left who have never tried self-employment.

Another issue that needs to be addressed is the number of spells of self-employment that each individual undertook to accumulate their observed number

⁶ 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.

⁷ This is only minor variation—solely resulting from the spread of times of the NCDS6 interviews, since all cohort members were born in the same week—and supports the validity of a count data approach for analysis.

Table 1 Distribution of quarters of self-employment from 1991—by gender^a

Quarters category	All		Males		Females	
	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

^a The table refers to the period from NCDS5 (1991) to NCDS6, and the NCDS6 interviews took place in 1999 and 2000

of quarters of self-employment. For example, were those who spent a modest amount of time self-employed repeatedly entering and exiting from self-employment? Were those occupying self-employment for long enough to be in the top (33–38 quarters) grouping doing so through a single spell in almost all cases? Table 2, below, allows us to answer these questions. Less than 20% of the individuals in the NCDS sample were clearly of Entrepreneurial Type (ET) through having some period of self-employment between NCDS5 and NCDS6. Of those undertaking some self-employment, nearly 90% had only one spell and 98% had two spells or fewer. Thus, empirical observation shows that our measure of Entrepreneurial Persistence (EP) is more useful in practice than we might have feared: we do not have to worry too much about controlling for the distinction between two individuals of ET with similar numbers of quarters self-employed, but very different numbers of spells of self-employment. This is a fortunate situation, since any regressor that simply measures the number of self-employment spells is definitionally forced towards a strong positive correlation with the number of quarters self-employed, since, when the latter is zero, the former must be zero also. However, conditional on some time having been spent self-employed, the number of self-employment spells might be negatively correlated with the number of quarters self-employed, if some individuals exhibit frequent transitions into, and out of, self-employment; while others undertake a single lengthy period in self-employment.

Further context is provided by Table 3—which demonstrates the mix between self-employment and

other main economic activity (or inactivity) states between NCDS5 and NCDS6, and largely the expected sorts of variations by gender:

We take the ‘general-to-specific’ approach for our estimation—starting by using data on as many available variables as possible that we might expect to be relevant in determining individual self-employment, but discarding some variables on the basis of the statistical evidence. 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; another dummy reflects family financial difficulties (NCDS1); 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 father⁸ in 1969 (NCDS2); a dummy is used to indicate use of the English language at

⁸ Worker own account, farmer own account and professional self-employed are summed to give (overall) ‘self-employed’. Self-employment data on the cohort member’s mother are less readily available—just 70 are recorded as ‘own account’.

Fig. 1 Relative frequencies of time spent self-employed, NCDS5-NCDS6

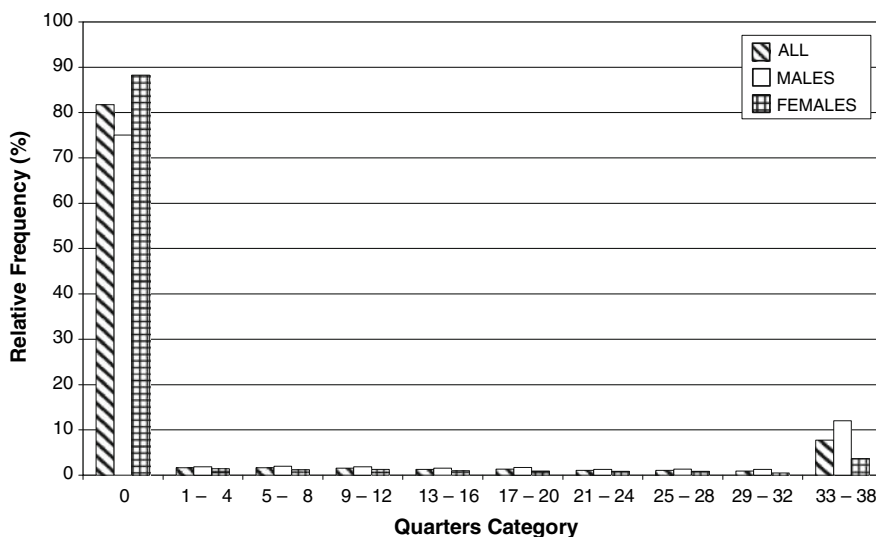


Table 2 Distribution of quarters of self-employment—by gender and number of spells

Quarters category	All			Males			Females		
	1 spell	2 spells	3+spells	1 spell	2 spells	3+spells	1 spell	2 spells	3+spells
1–4	177	6	1	98	4	1	79	2	0
5–8	168	15	0	101	12	0	67	3	0
9–12	161	15	1	97	4	1	64	11	0
13–16	127	14	5	74	10	3	53	4	2
17–20	128	16	7	80	11	5	48	5	2
21–24	97	21	4	54	16	1	43	5	3
25–28	100	20	4	63	11	3	37	9	1
29–32	74	24	5	52	20	1	22	4	4
33–38	793	73	16	625	40	8	168	33	8
Total	1825	204	43	1244	128	23	581	76	20

home (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,⁹ 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.

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

⁹ CSE indicates Certificate of Secondary Education. Like the more difficult O (Ordinary) level papers, CSEs were traditionally taken at age 16. A (Advanced) level papers were typically taken at age 18.

Table 3 Months of self-employment and other states of economic activity—by gender

Activity state	All (11361)		Males (5595)		Females (5766)	
	Mean	SD	Mean	SD	Mean	SD
F/T self-employment	11.02	29.75	18.07	36.83	4.18	18.24
P/T self-employment	1.83	11.73	0.58	6.52	3.05	15.07
F/T employee	60.23	47.79	79.41	41.90	41.62	45.76
P/T employee	16.28	32.55	0.99	7.55	31.10	39.83
Unemployment	2.08	10.74	3.03	12.77	1.17	8.21
Home/family care	10.78	27.06	0.61	6.22	20.64	34.76
Permanent sick/disabled	2.84	14.90	3.00	15.44	2.68	14.36
F/T education	0.85	5.54	0.49	4.12	1.19	6.62
Other	1.03	7.87	0.97	7.55	1.08	8.17
Total	106.93	5.41	107.15	4.29	106.72	6.30

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). See Burke et al. (2000), Taylor (2001), and Hurst and Lusardi (2004) for justification of non-linear effects. Although self-employment income is a potential determinant of EP, the NCDS data suffer from too many missing values (in addition to the guaranteed missing values for those who are not of ET, and possible measurement error); and also from the fact that income data are not available to cover the full NCDS5-NCDS6 period.
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. The issues of attrition and non-response at a particular sweep (wave), and the extent to which they are non-random in the case of the NCDS, are investigated in some detail by Hawkes and Plewis (2006). This builds on the more descriptive account by Plewis et al. (2004).

4 Econometric methodology

If the values taken by a dependent variable are non-negative integers, 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, 2002), Maddala (1983)) is the Poisson regression model—while an extension is offered by the negative binomial model. Negative binomial and Poisson random

variables each have a single parameter, and their discrete probability mass function can be written as:

$$P(Y_i = y_i) = e^{-\lambda_i} \lambda_i^{y_i} / y_i!; \quad y_i = 0, 1, 2, \dots \quad (1)$$

In each case, it is usual to specify the natural logarithm of the parameter as a linear regression function and estimate by the method of Maximum Likelihood (ML). The negative binomial model uses $\lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta} + \ln(u_i))$, where the u_i is often assumed to follow a unit-mean gamma distribution with parameter θ , and accommodates heterogeneity not captured by the $\boldsymbol{\beta}$ vector of regressors. The conditional mean of Y is λ_i , and the variance is $\lambda_i(1 + \kappa \cdot \lambda_i)$, where $\kappa = 1/\theta$. It is common in actual data for the variance to exceed the mean, and the negative binomial model allows this, whereas the mean and variance of a Poisson variate are both λ_i , where $\lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$, and unobserved heterogeneity is not addressed by the Poisson model.

In considering individual self-employment over a sample period, there may be two types of person not observed as self-employed—one that would never (seriously) consider becoming self-employed; and the other that would be willing, but was not self-employed at any point in the observed sample period. The *zero-inflated* Poisson and negative binomial models reflect this possibility—with a binary choice model (logit or probit) 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. Negative estimates here indicate a greater chance of being of ET. For both the zero-inflated negative binomial model and the zero-inflated Poisson model, with a logit component to inflate the number of zeros:

$$P(Y_i = y_i) = \left(1/1 + e^{w'_i \boldsymbol{\alpha}}\right) \left(e^{-\lambda_i} \lambda_i^{y_i} / y_i!\right); \quad (2a)$$

$$y_i = 1, 2, \dots$$

For the case of $Y_i = 0$, the negative binomial probability is shown below—and the appearance of the Poisson probability is different only in omitting the final $\ln(u_i)$ term:

$$P(Y_i = 0) = \left(1/1 + e^{w'_i \boldsymbol{\alpha}}\right) \left(e^{w'_i \boldsymbol{\alpha}} + e^{-\lambda_i}\right)$$

$$= \left(1/1 + e^{w'_i \boldsymbol{\alpha}}\right) \left(e^{w'_i \boldsymbol{\alpha}} + \exp(-\exp(\mathbf{x}'_i \boldsymbol{\beta} + \ln(u_i)))\right). \quad (2b)$$

The zero-inflated models are also estimated using ML, and it is usual to use robust standard errors (see White (1980), for example) 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:

$$E[Y_i] = 0 \cdot \left(e^{w'_i \boldsymbol{\alpha}} / 1 + e^{w'_i \boldsymbol{\alpha}}\right) + \sum_{y_i=0}^{\infty} \left[\left(y_i / 1 + e^{w'_i \boldsymbol{\alpha}}\right) \left(e^{-\lambda_i} \lambda_i^{y_i} / y_i!\right)\right]. \quad (3)$$

Estimated coefficients and regressor sample means can be used to provide estimated marginal effects.¹⁰

The Poisson regression model is nested within the negative binomial model, and the extra restriction imposed by it may be tested 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) is more applicable than standard least squares regression—since negative self-employment durations cannot be observed—it can readily be improved upon if a separate process determines zero and non-zero values of the self-employment duration (like the zero-inflated models for count data). Let us use a probit model for the individual's choice of whether to be self-employed and a Tobit model for the subsequent choice of non-zero time spent self-employed:

¹⁰ For a two-regime model with n_2 individuals of ET and n individuals in the whole sample, 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. The estimated probabilities from the logit model could be used to estimate n_2 , which is unknown.

$$d_i^* = \mathbf{w}'_i \boldsymbol{\alpha} + u_i; \quad d_i = \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (4a)$$

$$y_i^* = \mathbf{x}'_i \boldsymbol{\beta} + v_i; \quad y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (4b)$$

Lin and Schmidt (1984) consider a model proposed by Cragg (1971), using equations like (4a) and (4b). For an individual who chooses no self-employment via the probit model (4a), equation (4b)—a truncated regression (with truncation to the left of zero)—is not relevant. 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 $\boldsymbol{\alpha} = (\boldsymbol{\beta}/\sigma)$, where σ is the standard deviation of the disturbance term v in (4b). 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 (4a) and (4b); and in whether the participation decision ‘dominates’, in which case, in our example, only those not of ET could be observed spending zero time in self-employment. Slightly confusingly, the ‘Cragg model’ we have drawn from Lin and Schmidt (1984) is described by Jones (1989) as the ‘First hurdle dominance’ model. With $\Phi(\cdot)$ as the cumulative distribution function of the standard normal, its log-likelihood function, across a sample of n individuals, is as follows:

$$\ell = \sum_{i=1}^n \left[(1 - d_i) \ln \Phi[-\mathbf{w}'_i \boldsymbol{\alpha}] + d_i \left[\ln \Phi[\mathbf{w}'_i \boldsymbol{\alpha}] - \ln \Phi[\mathbf{x}'_i \boldsymbol{\beta}/\sigma] - \ln \left(\sigma \sqrt{2\pi} \right) - 1/2 \sigma^{-2} (y_i - \mathbf{x}'_i \boldsymbol{\beta})^2 \right] \right]. \quad (5)$$

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:

$$0. \Phi[-\mathbf{w}'_i \boldsymbol{\alpha}] + (\mathbf{x}'_i \boldsymbol{\beta} + \sigma(\phi(\mathbf{x}'_i \boldsymbol{\beta}/\sigma)/\Phi(\mathbf{x}'_i \boldsymbol{\beta}/\sigma))) \Phi[\mathbf{w}'_i \boldsymbol{\alpha}]. \quad (6)$$

Use of estimated coefficients and regressor sample means again leads to estimated marginal effects.¹¹

Jones (1989) uses the ‘Cragg model’ title for a model in which the disturbance terms from (4a) and (4b) are independent.¹² Given the discussion underlying the various model structures in Jones (1989), here we need to emphasise that certain aspects of self-employment enable a qualitative distinction between those who would never be self-employed and those who might be, even if only briefly. Although cigarette smoking and self-employment differ in many respects, attitudes to both exhibit considerable variation within society, including—to an extent—by social classification sub-group.

Jones (1989) also points out that participation in an activity at a point in time means that an individual has previously decided to commence it, and has also not quit from it. This is the basis for sample separation models explicitly modelling the individual’s decisions to commence and/or to quit. Blundell et al. (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 et al. (2000, 2002), have not used this method. This article needs a different approach, since we are not examining self-employment at some given instant, and multiple cycles of starting and quitting are observed for some individuals over our 9-year period.

5 Empirical results

This section will present the main results, summarising and discussing the content of Tables 4–7. We focus a fair bit on the intriguing (but less obvious) effects of background characteristics from the early lives of members of our 1958 cohort. The effects of including a few additional controls are discussed briefly in an appendix. Some statistical diagnostics

¹¹ 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 $(\Delta n_2/n)$.

¹² This model also has the Tobit model nested within it, provided the typo is corrected (as $\alpha_0 \rightarrow \infty, p(v > -\alpha'z) \rightarrow 1$).

Table 4 Zero-inflated negative binomial model for males

Variable	Inflation (logit)		Count (negbin)		Mean,	
	Estimate	Est./SE	Estimate	Est./SE	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

Table 4 continued

Variable	Inflation (logit)		Count (negbin)		Mean	Mean, EP > 0
	Estimate	Est./SE	Estimate	Est./SE		
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 5 First hurdle dominance model for males

Variable	Participation (probit)		Extent (truncreg)		Mean	Mean EP > 0
	Estimate	Est./SE	Estimate	Est./SE		
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

Table 5 continued

Variable	Participation (probit)		Extent (truncreg)		Mean	EP > 0
	Estimate	Est./SE	Estimate	Est./SE		
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

for our models, and a few predictions of time spent in self-employment (from Eq. 3) are also included there. However, to summarise, a two-regime approach is found to be justified throughout, for both males and females.

5.1 Male self-employment

There is an unsurprisingly strong element of persistence in the tendency to be self-employed—with self-employment status at age 23 making some self-employment between NCDS5 and NCDS6 more likely, and likely to last longer (shown in both Tables 4 and 5). Other factors found to favour 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. Notably, only the control for union membership at age 33 has the opposite effect on both being an ET, and EP.

A male is more likely to be of ET if his father was the manager of a small firm, if his father was a worker or a farmer with his own account (Table 5), if

he was lazy at age 16 (Table 5), or due to the receipt, timing and/or value of an inheritance received by age 33. 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 (Table 4) 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). Regressors that raise EP only include having had a father who was self-employed, and the value of an inheritance received at age 33 (but this effect is small in magnitude). Those that just reduce EP are having an O level equivalent as highest academic qualification, being flexible back at age 16 (presumably too flexible), having the view at age 23 that lack of responsibility is the most important aspect of a job (Table 4), high maths ability at age 16 (Table 5) and the number of unemployment spells suffered up to the age of 23.

5.1.1 Interpretation of results for male self-employment

Our two-part econometric approach proves to be superior to the simple logit/probit and therefore,

Table 6 Zero-inflated negative binomial model for females

Variable	Inflation (logit)		Count (negbin)		Mean	Mean, EP > 0
	Estimate	Est./SE	Estimate	Est./SE		
Constant	1.807	2.12	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

Table 6 continued

Variable	Inflation (logit)		Count (negbin)		Mean, EP > 0	
	Estimate	Est./SE	Estimate	Est./SE	Mean	EP > 0
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 . The Vuong test has a test statistic of 15.73. Removal of the region dummies from the logit yields a test statistic of 33.84, significant at the 1% level

supports our reclassification of pure wage-workers and self-employed into pure wage-workers and entrepreneurial types. However, what is reassuring 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 et al. (2000). Thus we confirm the existing literature on determinants of ET. Nevertheless, our results clearly indicate that causes of ET differ from those of EP. It follows that our results indicate that using probit/logit analysis to determine ‘what makes an entrepreneur’ is misleading if one is interested in ‘die-hard’ 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 article 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. 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, he is not only more likely to be of ET from 33–42, but is also more likely to persist in self-employment. 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

Table 7 First hurdle dominance model for females

Variable	Participation (probit)		Extent (truncreg)		Mean	Mean, EP > 0
	Estimate	Est./SE	Estimate	Est./SE		
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	-1.894	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

Table 7 continued

Variable	Participation (probit)		Extent (truncreg)		Mean	Mean, EP > 0
	Estimate	Est./SE	Estimate	Est./SE		
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

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 et al. (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 et al. 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 et al. (2000)) and persisting in entrepreneurship (as found in the second stage estimation of performance in Burke et al. (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 lovers, but rather had 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 die-hard/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 Leighton (1989), Blanchflower and Oswald (1998) and Burke et al. (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 ET, and EP, among males. Given that we later find it has a negative effect on female EP, this suggests 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 weakens the enterprise economy. This is in contrast to the push hypothesis (see Storey (1994) for an overview).

5.2 Female self-employment

Persistence in self-employment is again evident—but the regressor ‘self-employed at age 23’ is the only one to have positive effects on being of ET, and EP. However, a nursing qualification acts against being of ET, and has a negative impact on EP.

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. Regressors that only act to increase EP are a strong desire for a lack of responsibility in a job (Table 6, not easily explained), having a first degree as highest academic qualification (Table 7) and the timing of the receipt of an inheritance (closer to 1991, rather than less recently). Reduced female EP only is linked to being in Social Classes II, III and IV at age 7 (Table 7), the mother’s age of departure from full-time education, having suffered a parental split (Table 7), the number of unemployment spells endured since age 23, the number of children borne by age 23 and poor health at age 33.

5.2.1 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 (Burke et al. (2002)¹³). While males of ET and male EP can be largely explained within the confines of economic 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

¹³ It should not be controversial that detail is lost when only a simple gender dummy is included to capture differences in outcome (as in Blanchflower and Oswald (1998) and Burke et al. (2000))—and this is confirmed by unreported results. Given how much more common male self-employment is than female self-employment, it is also unsurprising that differences from gender splitting the data are usually more apparent for females. We chose to split the data, rather than using gender interactions, because the latter generates coefficients which are less immediately interpretable, and even longer tables of estimated results.

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 being a die-hard entrepreneur.

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.

6 Conclusion

The article contributes to the literature on entrepreneurial choice by moving beyond a dichotomy between wage-workers and 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 have been self-employed or, if not, would consider self-employment as a career option) and entrepreneurial persistence including die-hard entrepreneurs. We outline how entrepreneurial choice becomes richer when this distinction is considered. We also offer an econometric approach to test the appropriateness of this classification. To our knowledge, this article is the first empirical analysis of the determinants of persistent or 'die-hard' entrepreneurs. Using a recent update to the NCDS dataset we explore ET (individuals with an inclination for entrepreneurship), and EP, for both males and females across a 9-year period—from age 33–42. Diagnostic tests indicate clearly that a two-regime approach is indeed a superior specification to the simpler probit/logit dichotomous approach. The results have important ramifications because of the differences we find 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 ET. However, our results for EP or

‘die-hards’ are quite different and taking these results in conjunction with those on entrepreneurial types provides an enriched understanding of what makes an entrepreneur.

We find that male and female entrepreneurship are distinct—although with some common determinants. One of these is an early career experience of self-employment—which tends to encourage persistent entrepreneurship throughout the 33–42 career span. Similarly, higher levels of education tend to be associated with EP among both males and females. Access to finance 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. We also find self-employed fathers tend to encourage more entrepreneurial types and more persistence among their sons, but only the former among their daughters. This may indicate more relevance of a father’s human capital for a son’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. This may be due to disparate skills for self-employment and small business management, and/or small business managers encouraging 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 sons of self-employed fathers because the latter’s actual experience should provide their sons with a more realistic perspective. However, there may be less mentoring in the case of daughters—feeding through less additional realism and hence undermining the positive effect on persistence.

Our results are also consistent with children being a greater hindrance to female entrepreneurship. Having children by the age of 23 has no statistically significant effect on male entrepreneurship but is negatively related to female EP. We also find EP is hindered by poor health among females more than males.

Finally, our analysis sheds some interesting light on unemployment and entrepreneurship, which contrasts with previous views in the literature. We find that spells in unemployment do not increase ET, and decrease EP. Thus, we do not find an ‘unemployment push’ into self-employment, but it appears that early life experience in unemployment may reduce self-employment quality (leading to the drop in EP).

In sum, we offer a new theoretical perspective and empirical findings based on new data. However, this is only a first step beyond the pure wage-worker versus pure entrepreneur logit/probit approach, towards a multi-dimensional and dynamic analysis of different kinds of entrepreneurship.

Appendix 1

Unreported results indicate that, if the number of part-time self-employment spells is controlled for, this has little impact on other estimated coefficients. These same results indicate that people with several spells of part-time self-employment are typically less persistent. The estimates do not change much if we exclude the modest fraction of individuals with more than one self-employment spell. Unreported results that control for time spent in long-term sickness or disability are also similar, although there is a noticeable effect on poor health estimates (through likely multicollinearity). As expected, longer periods in sickness or disability seem to reduce the chances of being of ET, and also to reduce EP (although not significantly for females). A control for operation of a self-employed business from home at NCDS5 indicates a positive effect on EP—but again leaves other results similar.

Table A.1 contains diagnostics which detail the bases for favouring the zero-inflated negative binomial model and the first hurdle dominance model for males and females—i.e. two-regime models.

The table goes on to demonstrate some of the characteristics of predictions generated using the zero-inflated negative binomial and first hurdle dominance models. In particular, both show a tendency to under-predict the overall average length of time spent self-employed. The prediction comparison between those self-employed at age 23 and those who were not should be set in the context of the final two columns of Table 4. These show that those with *some* self-employment between NCDS5 and NCDS6 are nearly three times as likely to have been self-employed at age 23, compared to the male group as a whole. By contrast, the effect of an inheritance worth £30000 on predicted self-employment duration—for an otherwise average male—is only 0.12 quarters.

There are a few (relatively minor) differences in the results from zero-inflated and first hurdle

Table A. 1 Diagnostics and predictions

	Males		Females	
	Test statistic	Favours	Test statistic	Favours
Test				
Likelihood ratio (LR) test (negative binomial vs. Poisson)	78320	neg bin	54174	neg bin
Vuong test (for zero-inflation)	23.9	<i>zlnb</i>	15.7	<i>zlnb</i>
Likelihood ratio (LR) test (1st hurdle dominance vs. Tobit)	736	1st hurdle	271	1st hurdle
Prediction				
	<i>zlnb</i>	1st hurdle	<i>zlnb</i>	1st hurdle
Quarters self-employed (not self-employed at age 23)	4.58	4.96	1.86	1.93
Quarters self-employed (not S/E at 23, nursing qualification)	N/A	N/A	0.81	0.81
Quarters self-employed (self-employed at age 23)	18.74	19.32	6.98	7.57
Percentage with no S/E quarters	78.06	72.34	90.42	86.77
Quarters spent self-employed	5.04	5.63	1.90	1.99
Actual				
Percentage with no S/E quarters	75.07		88.26	
Quarters spent self-employed	6.30		2.45	

Notes: Likelihood ratio tests are asymptotically $\chi^2_{(1)}$ under the null. The Vuong test is asymptotically standard normal under the null. The term *zlnb* denotes the zero-inflated negative binomial model

dominance models. For example, the effects on male EP of creativity and the desire to be one's own boss are statistically significant only at the 10% level in Table 5 (first hurdle dominance model). In addition, the value of an inheritance received, in linear form, has a positive effect on both the probability of being of ET, and EP in Table 5. 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.¹⁴

References

- Bhide, A. (2000). *The origin and evolution of new businesses*. New York: Oxford University Press.
- Blanchflower, D., & Oswald, A. (1990). *What makes a young entrepreneur?* Papers 373, London School of Economics – Centre for Labour Economics.
- Blanchflower, D., & Oswald, A. (1998). What makes an entrepreneur? *Journal of Labor Economics*, 16, 26–60.
- Blundell, R., Ham, J., & Meghir, C. (1987). Unemployment and female labour supply. *Economic Journal*, 97 (Conference Supplement), 44–64.
- Bridge, R. (2006). *My big idea: 30 successful entrepreneurs reveal how they found inspiration*. London and Philadelphia: Kogan Page.
- Burke, A. E., FitzRoy, F. R., & 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., & 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.
- Constant, A., & Zimmerman, K. F. (2006). The making of entrepreneurs in Germany: Are native men and immigrants alike? *Small Business Economics*, 26(3), 279–300.
- Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39, 829–844.
- Cressy, R. (1996). Are business start-ups debt-rationed? *Economic Journal*, 106, 1253–1270.
- De Meza, D., & Southey, C. (1996). The borrower's curse: Optimism, finance and entrepreneurship. *Economic Journal*, 106, 375–386.
- Evans, D., & Leighton, L. S. (1989). Some empirical aspects of entrepreneurship. *American Economic Review*, 79, 519–535.
- Evans, D., & Jovanovic, B. (1989). An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy* 97, 519–535.

¹⁴ The *t*-ratio for this effect is -1.95 —so that it very narrowly fails to be statistically significant at the 5% level.

- Foreman-Peck, J. S. (1985). Seedcorn or chaff? New firm performance in the interwar economy. *Economic History Review*, 38(3), 402–422.
- Fraser, S., & Greene, F. J. (2006). The effects of experience on entrepreneurial optimism and uncertainty. *Economica*, 73, 169–192.
- Ghatak, M., Morelli, M., & Sjostrom, T. (2001). Occupational choice and dynamic incentives. *Review of Economic Studies*, 68, 781–810.
- Greene, W. H. (1997). *Econometric Analysis*. (3rd ed.,) New Jersey: Prentice-Hall.
- Greene, W. H. (2002). *Reference guide and econometric modeling guide, LIMDEP, version 8.0 (for PC under Windows)*. Econometric Software, Inc.
- Handy, C. (1999). *The new Alchemists: How visionary people make something out of nothing*. London: Hutchinson.
- Hawkes, D., & Plewis, I. (2006). Modelling non-response in the National Child Development Study. *Journal of the Royal Statistical Society (Series A)*, 169(3), 479–491.
- Henley, A. (2004). Self-employment status: The role of state dependence and initial circumstances. *Small Business Economics*, 22(1), 67–82.
- Holtz-Eakin, D., Joulfaian, D., & Rosen, H. S. (1994). Sticking it out: Entrepreneurial survival and liquidity constraints. *Journal of Political Economy*, 102, 53–75.
- Hurst, E., & Lusardi, A. (2004). Liquidity constraints, household wealth, and entrepreneurship. *Journal of Political Economy*, 112(2), 319–347.
- Jones, A. M. (1989). A double-hurdle model of cigarette consumption. *Journal of Applied Econometrics*, 4, 23–39.
- Jovanovic, B. (1982). Selection and evolution of industry. *Econometrica*, 50, 649–670.
- Kihlstrom, R., & Laffont, J. J. (1979). A general equilibrium entrepreneurial theory of the firm formation based on risk aversion. *Journal of Political Economy*, 87, 719–748.
- Lazear, E. P. (2004). Balanced skills and entrepreneurship. *American Economic Review*, 94(2), 208–211.
- Lin, T-F., & 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.
- Parker, S. C. (2004). *The economics of self-employment and entrepreneurship*. Cambridge University Press.
- Parker, S. C., & Belghitar, Y. (2006). What happens to nascent entrepreneurs? *Small Business Economics*, 27(1), 81–101.
- Plewis, I., Calderwood, L., Hawkes, D., & Nathan, G. (2004). Changes in the NCDS and BCS70 populations and samples over time, Centre for Longitudinal Studies Technical Report.
- Storey, D. J. (1994). *Understanding the small business sector*. London: Thomson Learning.
- Taylor, M. (2001). Self-employment and windfall gains in Britain: Evidence from panel data. *Economica*, 68, 539–565.
- Van Praag, C. M. (2003). Business survival and success of young small business owners. *Small Business Economics*, 21(1), 1–17.
- Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–334.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.