



# 0706

**Education and Regional Job Creation by the  
Self-Employed: The English North-South Divide**

by

**Andrew E. Burke**  
Cranfield University School of Management and  
Max Planck Institute of Economics

**Felix R. FitzRoy**  
University of St. Andrews and the IZA Bonn

**Michael A. Nolan**  
University of Hull

**Number of Pages: 30**

The *Papers on Entrepreneurship, Growth and Public Policy* are edited by the  
Group Entrepreneurship, Growth and Public Policy, MPI Jena.  
For editorial correspondence,  
please contact: [egppapers@econ.mpg.de](mailto:egppapers@econ.mpg.de)

ISSN 1613-8333  
© by the author

Max Planck Institute of Economics  
Group Entrepreneurship, Growth and  
Public Policy  
Kahlaische Str. 10  
07745 Jena, Germany  
Fax: ++49-3641-686710

**EDUCATION AND REGIONAL JOB CREATION BY THE SELF-EMPLOYED:  
THE ENGLISH NORTH-SOUTH DIVIDE**

*Andrew E. Burke (Cranfield University School of Management, UK and the Max Planck Institute of  
Economics ( Entrepreneurship Group), Germany)*

*Felix R. FitzRoy (University of St. Andrews, UK and the IZA (Bonn), Germany)*

*and Michael A. Nolan (University of Hull, UK)*

March 2006

**Abstract**

Using decomposition analysis, the paper investigates the reasons why Northern England has less but higher performing self-employed businesses than the South. It finds the causes are mainly structural differences rather than due to regional variation in people's characteristics. The paper also unearths a regional dimension behind the impact of education on entrepreneurial job creation. It finds that, in the less developed North, education boosts self-employment job creation by enhancing performance per venture (quality). In the South, it reduces it by having no effect on quality alongside a negative effect on the number of people who become self-employed (quantity).

Keywords: Self-employment, job creation, North-South divide, decomposition.

JEL classifications: J23, R11, R23

*Acknowledgements*

The authors are also grateful to Peter Shepherd of City University (London) and Kathy Sayer of the ESRC's Data Archive for generous assistance with requests for data and information. The authors would like to thank David Audretsch for suggesting that they investigate the regional dimension of their earlier work. In addition, the authors would also like to thank Jon Atkins, Martin Robson and participants at the Work and Pensions Economics Group Annual Conference, University of York, July 2005. A disclaimer applies.

**Section 1: Introduction:**

The regional dimension of the British economy has been well documented. Regional variation in economic performance is revealed in a North-South divide where the South typically has better economic performance than the North with lower unemployment and higher GDP per capita. This divergence in performance tends to be fairly persistent and has been associated with an economic policy response (for example, see Lewis and Townsend (1989), Fothergill (2001) and Gudgin (1996)). In particular, a policy response to the high and persistent regional divergences in unemployment (for example, see Gray (2004)) or non-employment (see Anyadike-Danes (2004)) has been to seek job creation through self-employment (Shutt and Sutherland (2003)). The work of Robson (1998), and that of Georgellis and Wall (2000), each include a theoretical framework to underlie a relationship between the regional self-employment rate and a variety of characteristics of the regional economy. Thus, there are some good research papers on the impact of structural differences on self-employment performance. What has received much less attention is regional variation in the average characteristics of people in the North versus the South. In other words, factors affecting the *ability* and *predisposition* of individuals to exploit available self-employment opportunities. These factors include individual characteristics such as skill, experience, education, psychological make-up and culture which may be important in making the most of available profit opportunities. Variation in regional self-employment rates and job creation can therefore be attributed to both structural disparities and differences in average individual characteristics.

In this paper we aim to develop the trajectory of research on the North-South divide in self-employment performance in England by accounting for compositional as well as structural influences. Through the use of longitudinal data on individuals, we can investigate the impact of both aspects. Thus, the contribution of this paper is to investigate – with regard to the numbers of the self-employed as indicated by the probability of self-employment, and also the performance measure of job creation by the self-employed – whether there is evidence of a North-South divide in English self-employment. This has not been done before.

Moreover, in neglecting this issue, previous work risks not properly identifying the impact of some characteristics on self-employment – if, for example, a statistically significant effect for one region opposes an effect in the opposite direction for another region. Of particular interest to us is the role of post secondary education which at a national UK level has been found to reduce the number of people who choose to become self-employed but increase the job creation capability of people who nonetheless opt to be self-employed. Burke, FitzRoy and Nolan, (2000) and Cowling, Taylor and Mitchell (2004) find that the net effect of these opposing effects effect is positive so that education increases the number of jobs created by the self-employed sector. Here we examine the regional dimension of this nationwide result and uncover some regional variation in terms of how this ‘less is more’ result comes about. We use data from the National Child Development Study (NCDS) – males and females being considered separately, as was justified previously in the work of Burke, FitzRoy and Nolan (2002). The gender split in that paper had as its background the ‘self-employed female underperformance hypothesis’, as discussed by Rosa, Carter and Hamilton (1996), and Du Rietz and Henrekson (2000).

In addition, we perform some decomposition analysis, in order to investigate the extent to which the variation in self-employment between Northern and Southern England results from fundamental differences in the characteristics of the inhabitants (on average) of those regions, as opposed to differential responses to given sets of characteristics. Our work follows on from the work in papers such as Reimers (1983), Cotton (1988), Neumark (1988), Oaxaca and Ransom (1994), and Fairlie (1999, 2003) – which utilised and developed the seminal analysis on decomposition by Blinder (1973) and Oaxaca (1973).

The paper proceeds with Section 2, in which the data used in our analysis are described. Section 3 proceeds to discuss the methodology we use, focusing particularly on the application of decomposition techniques to this topic. Section 4 lays out and discusses our results. It focuses initially on analysis of the probability of self-employment, and decomposing the differences between Southern England and Northern England into their compositional and structural parts, and

then proceeds to crucially complementary analysis of male self-employment job creation. There then follows a concluding section.

## Section 2: Data Description

The National Child Development Study (NCDS) – the data source used for our empirical analysis – has, over a spell of more than thirty years, been periodically obtaining information about a cohort of individuals born in the week 3rd March, 1958 to 9th March 1958 inclusive and living in Great Britain. Following an initial study in 1958, surveys undertaken at irregular intervals – in 1965, 1969, 1974, 1981 and 1991 – formed the basis of work done on self-employment by Blanchflower and Oswald (1998) and Burke, FitzRoy and Nolan (2000, 2002). The first of these papers only considers the self-employment decision, while the second and third also analyse measures of entrepreneurial performance, and the contribution of the third study is to disaggregate by gender.

The precise extent of self-employment indicated by the fifth sweep NCDS data from 1991 (NCDS5 hereafter) depends upon the exact definition that is chosen. Blanchflower and Oswald (1998) choose to define as self-employed the 1,279 (out of 11,369) individuals who indicate self-employment to be their main economic activity – some of whom are only part-time self-employed. We follow the broader definition used by Burke, FitzRoy and Nolan (2000, 2002) – including some people for whom part-time self-employment is not their main economic activity, and yielding a total of 1,558 self-employed. The maximum sample size, for our analysis of the probability of choosing self-employment rather than being solely in some other form of economic activity<sup>1</sup>, is 11,113. This indicates a self-employment probability of 0.140 across males and females. Although very similar in size to self-employment proportions given for 1991 in Blanchflower and Oswald (1998), our sample includes the unemployed in the ‘not self-employed’ cohort (whereas theirs does not). It should be

---

<sup>1</sup> We include the unemployed and, for example, housewives – many of whom are often considered to be economically inactive. In principle, these individuals could make a decision to become self-employed – provided appropriate incentives are offered so that self-employment provides them with greater utility than any feasible alternative.

noted that the 11,113 individuals can be split into 5,432 males and 5,681 females. The self-employment probability for males is 0.195, while that for females is only 0.088.

Table I (below) shows in more detail how the self-employment probability differs by region in the NCDS5 data<sup>2</sup>. The last four rows of the table calculate self-employment probabilities for composite regions comprising at least two of the standard Government Office Regions (GORs). This paper focuses particularly on Southern England (a combination of Greater London, South East England and South West England) and Northern England (a combination of Yorkshire & The Humber, North West England and the North of England GOR). The table shows clearly that, for the NCDS data, there is a noticeably higher self-employment rate in Southern England than in Northern England – 23% against 17% among males, and 11% against 8% among females. It is also clear – as expected – that self-employment rate is substantially higher for males than females (about 19½% against 9%).

*{Table I near here}*

Turning from self-employment rates to performance, we use a measure of employment by the self-employed which is also provided by NCDS5 – where each self-employed cohort member indicates how many employees he/she has. Some summary statistics are shown in Table II below. Table II deals with 1526 self-employed individuals that report a value (quite often nil) for job creation. The higher job creation rate by the self-employed in Northern England is an interesting feature to emerge from the table. This is true both for males – 3.529 jobs per self-employed individual on average, compared to 2.652 jobs for Southern England – and for females (3.477 jobs per self-employed individual, versus 3.079). Another insight offered by Table II is that there is less of a job creation gap between self-employed men and self-employed women than there is a gap between the gender-specific self-employment probabilities. Indeed, there is almost no difference by gender in the job creation rate for the self-employed of Northern England.

---

<sup>2</sup> It is beyond the scope of this paper to speculate on how the North-South divide has developed over time, or on how it might differ across population age cohorts.

*{Table II near here}*

We now turn our attention to the range of factors likely to play a role in the determination of self-employment choice and performance. These regressor variables used in our estimation of self-employment probability probits, and/or job creation tobits are drawn from the NCDS. They are based on the theoretical underpinnings of the previous papers on self-employment choice using this dataset, namely Blanchflower and Oswald (1998) and Burke, FitzRoy and Nolan (2000, 2002) and are as follows:

1. *Ability, education and training* – there are dummies to indicate whether the highest academic qualification achieved is O level (or equivalent), A level, first degree or higher degree; up to 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 on an apprenticeship by the cohort member by 1981; another denotes receipt of a vocational qualification by 1991.
2. *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, flexibility, 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. Included are 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. This set of dummies serves to capture the individual's primary motivational priority across a range of pecuniary and non-

pecuniary dimensions – and the potential relevance of both pecuniary and non-pecuniary motivation has been emphasised in the model shown in Burke, FitzRoy and Nolan (2000).

3. *Family background* – a dummy reflects family financial difficulties (NCDS1); another denotes use of the English Language at home in 1969 (NCDS2); a series of dummies are used to indicate the occupation of the cohort member's father in 1969 – including employee manager of small firm, employee manager of large firm, professional self-employed, professional employee, foreman (manual work), skilled manual, worker with own account, farmer employee-manager and farmer with own account; 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.
4. *Current family* – a dummy captures not having at least one child by 1991; we also investigate the interaction of this dummy with higher level qualification (at least A level). Being childless may be relevant because people with children face extra obligations and, thus, time constraints. Burke, FitzRoy and Nolan (2002) found that both males, and less qualified females, with children are more likely to be self-employed (perhaps due to self-employment's potential flexibility in working time). Those authors also found that highly qualified males with children, once self-employed, undertook greater hiring of workers than otherwise similar childless entrepreneurs. To capture exogenous finance, three variables are also constructed (NCDS5) to capture the size (linearly and quadratically<sup>3</sup>) and timing (year) of any inheritance received by the NCDS cohort member – in order to investigate the existence of liquidity constraints.
5. *Region* – we construct four regions from the 11 GORs of Great Britain<sup>4</sup>, of which two (Southern England and Northern England) are our concern in this paper. Even within the composite regions, there may be some variation in costs (particularly housing) and demand conditions. In our probits

---

<sup>3</sup> These enter in present value (1991) form. The linear term is divided by 10000 (yielding a mean, across all cases with specified region and gender, of 0.5321) and the quadratic term is divided by  $1.0 \times 10^{10}$  (which gives a mean across all cases of 0.4996). The timing control indicates the year in which the inheritance was received (subtracting 1900 from the year in question, and then dividing by 100).

<sup>4</sup> In grouping the 11 Government Office Regions of Great Britain into 4 – of which two are the north and the South (of England) – we use a similar classification to Burke, FitzRoy and Nolan (2002).



of self-employment, we include two GOR dummies – with South-West England GOR being the base part of our Southern England composite region, and North of England GOR being the base within our broader definition of Northern England. In the self-employment tobit equations, we include the average GOR unemployment rate as a control, rather than pairs of GOR dummies.

6. *Aspects of self-employment* – we include a control for the length in years by 1991 of a spell of self-employment ongoing at NCDS5. We also use a dummy to distinguish those among the self-employed who are not full-time.
7. *Missing value dummies* – for some individual regressors, and some groups of regressors, an extra dummy is used to indicate missing data, and as a (rather limited) control for this fact. This approach is quite common, and has been used previously by papers that analyse NCDS data.

### **Section 3: Empirical methodology**

We perform some basic decomposition analysis on our results in a bid to determine whether there is evidence that the differences in self-employment probability between Southern England and Northern England are primarily a result of ‘compositional’ differences or ‘structural’ differences. Our interpretation of these terms follows previous literature – so that compositional differences reflect between-region differences (on average) in the characteristics of individuals; and structural differences refer to there being between-region differential responses to given characteristics. This approach is useful to investigate one of the core motivations of the paper which is to assess the extent to which the self-employed sector differs between Northern and Southern England due to variations in the composition of the regional economies (particularly in the characteristics of the workforce) and to differences in the behaviour patterns of individuals (responses to given characteristics).

The initial work done on decomposition by Oaxaca (1973) and Blinder (1973) pertained to linear regressions of the logarithm of wages. However, since probit and logit specifications also typically involve a linear regression specification, Gomulka and Stern (1990) and Fairlie (1999,

2003) have noted that it is feasible to undertake such decompositions for these models. When appropriately transformed, the estimated coefficients from probit and logit models are typically similar – which is unsurprising, given the shapes of the normal and logistic distributions that underlie the respective models. For our decompositions, we switch from the probit model to the logit model – because a useful feature of the logit model is that the predicted probability of a given outcome is identical to the actual probability, not only for the whole sample, but also for sub-samples where a given dummy variable takes a particular value<sup>5</sup>.

The basic format of the decomposition is as follows:

$$\bar{Y}_S - \bar{Y}_N = \left[ \sum_{i=1}^{n_S} \frac{F(\mathbf{X}'_{iS} \hat{\boldsymbol{\beta}}_S)}{n_S} - \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iN} \hat{\boldsymbol{\beta}}_S)}{n_N} \right] + \left[ \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iN} \hat{\boldsymbol{\beta}}_S)}{n_N} - \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iN} \hat{\boldsymbol{\beta}}_N)}{n_N} \right], \quad (1)$$

where  $\bar{Y}$  indicates the mean of the dependent variable in a logit model,  $F(\cdot)$  is the Cumulative Distribution Function of the logistic distribution, the  $S$  and  $N$  subscripts indicate (throughout) Southern England and Northern England respectively,  $n$  denotes the number of individuals in a particular area,  $\mathbf{X}'_i$  is an individual's vector of characteristics and  $\hat{\boldsymbol{\beta}}$  is the vector of estimated coefficients from the logit model. The estimated coefficients vectors have the subscript  $S$  or  $N$  attached because they are generated through the separate estimation of the logit model for those from Southern England, and for those from Northern England. The version of the decomposition shown in equation (1) is split into a compositional effect (the term in the first square bracket) and a structural effect. The compositional term looks at the average predicted probability of self-employment that would be generated if the individuals from Northern England were subject to the response coefficients resulting from the estimation of a logit model for Southern England, relative to the average predicted probability of self-employment in Southern England. The structural term measures the influence on the self-employment probability of the difference between the response coefficients for Southern England and those for Northern England, together with the impact of

<sup>5</sup> This is true approximately for the probit model, and empirically we found this approximation to be quite close.

unobserved regional differences in characteristics (which, by definition, cannot be captured within the logit estimation).

The following expression for the decomposition is equivalent:

$$\bar{Y}_S - \bar{Y}_N = \left[ \sum_{i=1}^{n_S} \frac{F(\mathbf{X}'_{iS} \hat{\boldsymbol{\beta}}_N)}{n_S} - \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iN} \hat{\boldsymbol{\beta}}_N)}{n_N} \right] + \left[ \sum_{i=1}^{n_S} \frac{F(\mathbf{X}'_{iS} \hat{\boldsymbol{\beta}}_S)}{n_S} - \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iS} \hat{\boldsymbol{\beta}}_N)}{n_S} \right]. \quad (2)$$

The interpretation of equation (2) involves the opposite movement to equation (1) where individuals now move from the buoyant South northwards to a part of England which is relatively demand-deficient by comparison<sup>6</sup>.

Even and Macpherson (1990, 1993) noted that the decomposition component attributable to differences in characteristics can be split on a variable by variable basis (or for a group of variables, where this is more appropriate). The numerator for the required ratio is given by the size of the difference in sample means across the two groups (regions in our case) for the single variable, weighted by its estimated coefficient. The denominator is the difference in sample means across the two groups for *all* variables (each difference being weighted by the corresponding estimated coefficient). Hence the contribution to the probability gap by regressor  $r$  is as follows:

$$\left[ \sum_{i=1}^{n_S} \frac{F(\mathbf{X}'_{iS} \hat{\boldsymbol{\beta}}_S)}{n_S} - \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iN} \hat{\boldsymbol{\beta}}_S)}{n_N} \right] \left[ \frac{(\bar{X}_{rS} - \bar{X}_{rN}) \hat{\beta}_{rS}}{(\bar{X}'_S - \bar{X}'_N) \hat{\boldsymbol{\beta}}_S} \right]. \quad (3)$$

Before we proceed to our estimation results, we should undertake a little further discussion of the issue of the appropriate decomposition. In its applications to wages, the question of what constitutes the ‘no discrimination’ distribution is often discussed. In our case too, we might expect that norm coefficients should be something other than either those for the South (as in equation (1)) or those for the North (equation (2)). While Reimers (1983) suggested the use of the arithmetic mean of the two sets of coefficients, and Cotton (1988) proposed the use of a weighted mean (the relative weights being determined by relative sample sizes) it has become more usual to consider the

<sup>6</sup> Naturally, we would expect there to be significant knock-on effects of substantial one-way migration within England. Nonetheless, we might abstract to consider simultaneous balanced migration of some number of “typical” individuals.

coefficients resulting from pooled estimation across the groups under examination – as in Neumark (1988) and Oaxaca and Ransom (1994), where it is demonstrated that the coefficients from a pooled regression can be written as a weighted sum of the regression coefficients for the two groups. In the case of the logit model, the decomposition can be written as follows:

$$\left[ \sum_{i=1}^{n_S} \frac{F(\mathbf{X}'_{iS} \hat{\boldsymbol{\beta}}_P)}{n_S} - \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iN} \hat{\boldsymbol{\beta}}_P)}{n_N} \right] + \left[ \sum_{i=1}^{n_S} \frac{F(\mathbf{X}'_{iS} \hat{\boldsymbol{\beta}}_S)}{n_S} - \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iS} \hat{\boldsymbol{\beta}}_P)}{n_S} \right] + \left[ \sum_{i=1}^{n_S} \frac{F(\mathbf{X}'_{iN} \hat{\boldsymbol{\beta}}_P)}{n_N} - \sum_{i=1}^{n_N} \frac{F(\mathbf{X}'_{iN} \hat{\boldsymbol{\beta}}_N)}{n_N} \right], \quad (4)$$

where the  $P$  subscript refers to the pooled sample of Southern England plus Northern England. The first term reflects the compositional effect (viewing the pooled coefficients as applying to individuals in both Southern England and Northern England). The second and third terms each reflect how the regional coefficients depart from the pooled norm.

#### Section 4: Estimation results

We begin our econometric work by considering the factors which influence the probability of an individual being self-employed. The well-known probit model provides a straightforward method of estimating this probability and Table III, below, shows probit maximum likelihood estimates<sup>7</sup> for males – in Southern England and Northern England separately. The regressors in Table III are those remaining after a general-to-specific process based principally on at least some weak evidence of statistical significance in one or both regions<sup>8</sup>. A number of differences are apparent between the two ends of England, beyond the higher mean level of self-employment in Southern England indicated previously in Table I. Some of these differences are shown by the simple means of the regressors in the third and sixth columns of numbers within Table III. For example, among NCDS cohort males, a higher proportion reach first degree level in Southern

<sup>7</sup> These estimates, and those elsewhere in this paper, were generated using the Limdep package (see Greene (2002)). It should be noted that Greene argues against the general usage of ‘robust standard errors’ across a variety of non-linear models. We were not successful in attempting to estimate a probit model with a particular specification for heteroscedasticity – non-convergence resulted when estimation of such models was tried.

<sup>8</sup> The two region dummies for which estimates are reported near the foot of Table III refer respectively to the Greater London and South-East England GORs in the case on Southern England (the South-West England GOR forming the base); and to Yorkshire and the Humber and the North-West England GOR in the case of Northern England (the North of England GOR being the base here).

England (16%, versus 10% in Northern England). Nor is this North-South divide on education confined to the NCDS generation itself – since the regressor means for the (grouped) variables on parental education are both noticeably higher for the South than the North.

The previously evident negative link between post-compulsory qualifications and the probability of male self-employment (see Burke, FitzRoy and Nolan (2002)) shows through clearly for Southern England – particularly for those with children – but is notable by its absence (even at the 10% significance level, and especially for those with children) for the northern region. This may be a result of the less buoyant economy of Northern England having lower wage work opportunities for the relatively well educated (since these opportunities would normally tend to draw them away from self-employment). There is a similar result for vocational qualifications, but apprenticeship has a positive association with male self-employment probability in both regions. The significant positive coefficient (Southern England) on the dummy identifying low reading ability at age 16 might indicate these individuals have relatively poor employment prospects given the generally higher levels of education in the South and are pushed into self-employment.

*{Table III near here}*

Creativity, which was found by Burke, FitzRoy and Nolan (2000) to be positively – though weakly – linked to the self-employment probability across both genders, is shown above to be almost significant for males in Southern England only. Another finding in this category is that those in the North rated (at school) as relatively depressed are more likely to take up self-employment. The fact that being rated as lazy when aged 16 is positively linked to self-employment in the South – but not (significantly) in the North – would appear to indicate that the employers in the South feel able to reject lazy individuals, who are then pushed into self-employment. Lazy individuals may also find employment more arduous than self-employment (where they are their own boss). Being rated as sociable when aged 16 is also positively related to the self-employment probability for males in Southern England only – which might be an indication of the sometimes observed North-South gap in ways of working and trading, or a tendency towards a difference in the self-

employment activities undertaken in the two regions. The finding of a negative relation between being timid aged 16 and self-employment at age 33 for Northern England only interestingly coincides with a lower mean for the timidity regressor there compared to the South (perhaps a timid individual stands out a bit more in the North – as unsuitable for self-employment). An expressed desire to be one's own boss at age 23 is associated, as expected, with a higher probability of self-employment in the North and the South – and there is also the expected negative link of self-employment with the desire for job security (although it is notable that a higher proportion of males in the North rated job security as the most important job characteristic in 1981 – when unemployment was particularly high, especially in the North).

Having a male parent who was the manager of a small firm is less common in Northern England than in the South, but this aspect of family background only appears to have a significant positive impact on the self-employment probability in the North. Parental education lasts about 0.3 years longer on average for each parent in the South, and its links with self-employment probability also seem to exhibit a North-South divide – with a statistically weak positive link from father's education (only) in Southern England, and a stronger positive link with mother's education (only) in the North.

The effect of inheritance on the self-employment probability found by Blanchflower and Oswald (1998) and by Burke, FitzRoy and Nolan (2000, 2002) shows evidence of a non-linear component. Only in the last of those studies are males considered separately. The effect of inheritance on the self-employment probability found there – a positive linear effect of strong statistical significance and a rather weak negative quadratic effect, plus a weak link with timing whereby recent inheritance comes with less chance of self-employment – is altered when the regional dimension is considered. This paper finds a positive linear effect of inheritance for Southern England only, and no evidence of a quadratic effect. There is also a statistically significant link between recent inheritance and lower self-employment probability in the South, but no evidence of any effect of the magnitude or timing of inheritance for Northern England.

Decomposition results:

To calculate our decompositions, we have dropped all dummies to capture Government Office Region (GOR) or a wider region – taking the rather severe position that, under the ‘no discrimination’ distribution, there is assumed to be no purely spatial aspect to variation in the self-employment probability. We switch across to the logit model, so that predicted probabilities reflect actual frequencies precisely. Estimation results for corresponding logit models in this case are quite similar to their (appropriately transformed) probit analogues in Table III. The decompositions are shown in Table IV, below – for five alternative approaches. The compositional term is negative in four out of five instances, but differs in sign if equation (2) is used. This serves to re-emphasise the importance of the choice that is made for the ‘no discrimination’ distribution. In each case, however, the compositional term is dwarfed by the structural part(s) of the overall gap in self-employment probability – indicating that male self-employment probability appears to be subject to a substantial North-South divide that is not accounted for by the characteristics of individuals.

*{Table IV near here}*

Notwithstanding the limited magnitude of the compositional term found above, we now investigate – using equation (3) following Even and Macpherson (1993) – the contributions of the various categories of regressor laid out previously<sup>9</sup> in Section 2. The second term in equation (3) can be used to break down the coefficient-weighted difference in regressor sample means, in accordance with each of our reported methods of decomposition in turn. The results are shown in Table V, below<sup>10</sup>. A general point should be noted about the relationship between rows 1, 2 and 4 in Table V – namely, that the definition of the second term of equation (3) combines with the Reimers (1983) definition of the ‘no discrimination’ coefficients (as the simple arithmetic mean of the coefficient vectors for Southern England and Northern England) so that each element in row 4 is

<sup>9</sup> Categories 5 (regions) and 6 (characteristics of self-employment) do not apply to our decomposition of the probability of self-employment.

<sup>10</sup> Each row has six columns of numbers. Each of the first five give the numerator for the second term in equation (3) for that particular category of regressor. The last column (which is the sum of the first five) is the denominator for the second term in equation (3). The second row illustrates an important problem in the use of equation (3) – where the denominator of the second term is very small relative to four of the numerators.

equal to the simple average of the corresponding elements from rows 1 and 2. Similarly, there is a relationship between rows 1, 2 and 5 – whereby each element in row 5 is equal to the weighted average of the corresponding elements from rows 1 and 2 (as in the Cotton (1988) definition of the ‘no discrimination’ coefficients).

*{Table V near here}*

One of the key features of Table V is that the first category (ability, education and training) of regressors has a robust effect across the five forms of the decomposition – whereby the extra ability, education and training of an average individual in Southern England leads them to be less likely to be self-employed than the average individual in Northern England (probably through the extra opportunities as an employee which higher levels of education and training usually afford). However, the gap in family background characteristics (category 3) is shown in Table V to have a robust effect in the opposite direction – so that the type of family background enjoyed by the average individual in Southern England differs from that of his counterpart in Northern England in ways that, *ceteris paribus*, make self-employment more likely.

#### Female self-employment:

Table VI shows self-employment probit estimates for females – separately for Southern England and Northern England. In addition to the higher self-employment rate in Southern England already noted in Table I, the sample means shown in Table VI indicate some interesting regional differences – although the fact that there are fewer regressors that are found to have a statistically significant effect on the self-employment probability means that this table is shorter than Table III.

*{Table VI near here}*

As for males, females in the South are more highly educated on average than those in the North. For females, in contrast to males, there is little evidence of academic qualifications having an impact on the probability of self-employment. However, there is some indication that females in the South with A-levels as their highest qualification may be more likely to be in self-employment at age 33.

---



No corresponding effect is evident for females in Northern England, although high reading ability at age 7 is associated with a greater chance of being self-employed in NCDS5.

We briefly summarise some other results on self-employment probability that differ from the findings of Burke, FitzRoy and Nolan (2000, 2002), or exhibit a North-South divide. Previous depression (positive) and timidity (negative) have impact only in the South. So too does the desire at age 23 to be one's own boss (positive) – this was not apparent in earlier work<sup>11</sup>. The link between having a father working with his own account and subsequent self-employment of the child is now shown to stem from females in Northern England. The positive association between childlessness and the female self-employment probability is significant only at the 10% level, and only disappears among highly qualified females for Northern England. Magnitude of inheritance is only of statistical significance for females in the South – where the linear effect is statistically strong, but the quadratic (negative) effect is only significant at the 10% level.

#### Decomposition results:

The decompositions are again shown for the same five approaches. In contrast to the results for males, the compositional term for females is positive in every case – although the overall gap is of the same sign as it was for males. For equations (1) and (4), the compositional term accounts for about half the overall gap – and, for all five decomposition approaches, there is a greater relative importance for the regional differences in average characteristics in determining female probability of self-employment than was found for male self-employment probability. See Table VII, below:

*{Table VII near here}*

Viewing the respective balances of the compositional and structural elements in rows 1 and 2 as those for two opposite extremes in terms of the form of decomposition, the element balances for the other ('intermediate') forms of decomposition do fall in between. It was initially felt by some

---

<sup>11</sup> Nor was the lower incidence of such females in Northern England.

previous authors that this must be the case – but Oaxaca and Ransom (1994) gave an example where it was not.

The breakdown of the coefficient-weighted difference in regressor sample means is shown in Table VIII, below – although it should be noted that the effects of inheritance on female self-employment probability in Northern England (measured by statistically insignificant, but quite large coefficients) seem to distort the results for category 4 and the overall breakdown in rows 2, 4 and 5. One key difference between these results for females and the corresponding results for males in Table V is the effect of regressors from category 1. Although, there is evidence of more education and training on average for females in Southern England, this has rather limited impact on the self-employment probability (and the effect is positive for most decompositions). There also seems to be more of a role for differences in non-cognitive characteristics than was evident for men. The effect of family background regressors from category 3, on the other hand, is rather smaller.

*{Table VIII near here}*

#### Job creation by self-employed males:

Table IX, below, shows the results of censored (Tobit) regression for males – again, comparing Southern England and Northern England. Among the male self-employed, job creation is positively associated with academic qualifications – but while having an A-level as highest seems significant in the South, it is the lower O-level (or equivalent) and the more advanced first degree that are significant at the 5% level or better for Northern England. Meanwhile, the positive effect of a professional qualification – found elsewhere in more aggregated samples – is preserved across the two separate regions. The negative link shown between job creation and high maths ability (at age 7) may be an indication that non-pecuniary considerations are important for some able entrepreneurs – whereby they may choose a self-employed activity that does not create a maximum number of jobs. Unforthcomingness was found by Burke, FitzRoy and Nolan (2000) to have a statistically very weak negative link to job creation across all the self-employed. Table IX illustrates that this

relationship is statistically significant at the 5% level for males in Southern England. The classification of a job characteristic as most important does not appear to be a good indicator of self-employment job creation – although there is a weak positive link for the promotion characteristic in the North. Timidity is negatively linked to job creation in the South – whereas, in the North, it is negatively linked to the self-employment probability. Having a father who was a professional employee is associated with greater job creation by self-employed males in Southern England, but not for the North (negative but insignificant here) – whereas Burke, FitzRoy and Nolan (2000) found a statistically weak positive effect aggregated across self-employed males and females. Another notable effect found in this family background category is the strong positive relationship between the cohort member's father's education and job creation in the South only.

The estimates on the inheritance regressors superficially appear different between the two regions – but the differences are not statistically significant. The positive linear coefficients are, themselves, statistically significant at the 10% (South) and 5% levels (North). The negative quadratic estimate for the South is significant at the 10% level, while that for the North is not quite significant at even that level. The other noticeable distinction is that the regressor means are quite different between the two regions: those for the South are substantially higher, and given the greater gap for the quadratic regressor, this is an indication of some rather large inheritances having been received among the South's self-employed males.

*{Table IX near here}*

## **Section 5: Conclusions**

This paper focuses on the issue of whether there is a North-South divide in England, regarding the factors that determine the level of self-employment and job creation by the self-employed. The background to this is the known differences in the regional economies of the two areas. Males and females are studied separately, because previous work has indicated that the processes underlying self-employment decisions may differ by gender. We also undertake some

decomposition analysis to provide insights into the distinction between regional structural effects and the effects of regional differences in the average characteristics of individuals.

On the probability of male self-employment, we find that the negative link with post-compulsory education is clear only for Southern England – with more employment opportunities in the South, this effect probably reflects more highly educated individuals choosing wage work rather than self-employment. In the North, fewer job opportunities may tend to push a well educated individual at the margin into self-employment. We also find that in the South there are some indications of self-employment being associated with lower ability and motivation.

In terms of evidence of finance gaps for the self-employed, Blanchflower and Oswald (1998) note that inheritance is a good proxy for the impact of exogenous availability of finance – and should have a positive and significant effect on self-employment if finance constraints exist. Our results for Southern England include a positive effect from the magnitude of any inheritance received, and from it being received less recently – but neither of these effects appear for Northern England. This result is interesting because, if anything, one would normally expect more finance to be available to the self-employed in the South compared to the North. Thus, the regional variation which we find may indicate that greater business opportunities in the South means greater amounts of finance are needed in order to realise optimal scale so that finance constraints are more likely to exist in the South. In addition, we note that there is a higher incidence of self-employment in the South compared to the North, so that one might expect greater competition for small firm finance and hence a greater incidence of finance gaps in the South. Thus, the analysis uncovers some interesting regional differences which underlie findings by previous studies which use the same database at a national level and find positive concave (Blanchflower and Oswald (1998), and Burke, FitzRoy and Nolan (2000, 2002)) effects of inheritance on self-employment. Obviously, further research is needed but the results so far raise an issue of whether the type of finance support for the self-employed ought to vary regionally.

The higher probability of self-employment in Southern England is shown, through decomposition, to exist in spite of a small negative compositional effect – which itself includes a rather larger negative component arising out of the higher education, ability and training of males in Southern England, and a notable opposing effect resulting from differences in family background. None of these findings could have been made apparent through a traditional aggregated study of male self-employment. Our results on male job creation indicate there are some differences by region in which qualifications are associated with the creation of more jobs. Having a father who was a professional employee only seems to help job creation for those in the South; and job creation in this region (only) is also linked to paternal education.

For the female self-employment probability, there is much less evidence of education having an impact – although there is a post-compulsory qualification level (A level) that may have a push effect into self-employment for Southern England. Inheritance magnitude is found to have no effect on self-employment probability among females in the North – and this is another regional contrast. The higher probability of self-employment for females in Southern England is shown, through decomposition, to result – to some extent – from a compositional effect. This is contributed to by both education and family background, but the situation is made harder to read by the impact of inheritance – particularly through a small number of large inheritances to females in Northern England (where the effect of inheritance on the self-employment probability is found to be statistically insignificant).

Finally, the male results give some new insights into the ‘less is more’ hypothesis as well as explaining why self-employment is higher in the Southern England but more productive in Northern England. In terms of the latter, the short answer is that the difference is due to differences in the regional economies and not the characteristics of the average individual in each region. Thus, for example, the fact that post-compulsory education is lower in Northern England than Southern England would actually lead to higher levels of self-employment if the Northern economy mirrored that of the South. The fact is that the Northern economy is found to be structurally different to the

South so that the lower rate of self-employment is generated *despite* post-compulsory education being lower. Likewise, in terms of performance, the lower levels of post-compulsory education in the North would typically have no effect on performance if the Northern economy was a mirror of the South but by contrast the results show that it is *only* in the Northern economy that post-compulsory education has a positive effect on performance. Again the North-South divide in self-employment is being driven by differences in the regional economic structures rather than the composition of the male population. These observations then enlighten a core question about the regional dimension to the ‘less is more’ hypothesis. The above results imply that in the South there is an unambiguous negative effect of post-compulsory education on the performance of the self-employed sector as it reduces the *quantity* but has no effect on the *quality* (performance) of the self-employed. By contrast, in the North post-compulsory education has an unambiguous positive effect as it has no effect on the numbers of self-employed (*quantity*) but enhances *quality*. Therefore, the result shows that the net positive result generated in Burke, FitzRoy and Nolan (2000, 2002) which uses the same dataset was not as a result of a general effect across the UK but stemmed from two different regional economies responding in opposite directions in terms of an impact of post-compulsory education on self-employment. It indicates that the same might be true for the net positive ‘less is more’ effect reported in Cowling, Taylor and Mitchell (2004) which uses British Household Panel Survey data. This new insight not only unearths a more complex regional dimension in the performance of the self-employed sector but also highlights that the ‘less is more’ result cannot be assumed as we have provided evidence here how structural differences can give rise to other effects from increased post-compulsory education – namely, ‘less is less’ (the South) and ‘same is more’ (the North). This indicates that the ‘less is more’ result appears to be the consequence of aggregating different regional economies in which post-compulsory education draws a different response from the self-employed sector; and regional differences regarding factors such as economic development, industrial structure, unemployment rate and culture lead to differences in the opportunities available to the well qualified. This new insight highlights the

importance of identifying how regional economies differ in terms of the determinants of job creation by the self-employed.

---

**References**

- Anyadike-Danes, M. (2004) The Real North-South Divide? Regional Gradients in UK Male Non-employment, *Regional Studies*, 38(1): 85-95.
- Blanchflower, D. and Oswald, A. (1998), What Makes an Entrepreneur? *Journal of Labor Economics*, 16: 26-60.
- Blinder, A. S. (1973), Wage Discrimination: Reduced-Form and Structural Estimates, *The Journal of Human Resources*, 8(4): 436-455.
- 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: 255-270.
- Cotton, J. (1988), On the Decomposition of Wage Differentials, *Review of Economics and Statistics*, 70(2): 236-243.
- Cowling, M., Taylor, M. and Mitchell, P. (2004), Job Creators, *The Manchester School*, 72(5): 601-617.
- Du Rietz, A. and Henrekson, M. (2000), Testing the Female Underperformance Hypothesis, *Small Business Economics*, 14(1): 1-10.
- Evans, D. and Jovanovic, B. (1989), An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints, *Journal of Political Economy*, 97: 808-827.
- Even, W.E. and Macpherson, D.A. (1990), Plant Size and the Decline of Unionism, *Economics Letters*, 32(4): 393-398.
- Even, W.E. and Macpherson, D.A. (1993), The Decline of Private-Sector Unionism and the Gender Wage Gap, *Journal of Human Resources*, XXVIII(2): 279-296.
- Fairlie, R.W. (1999), The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-employment, *Journal of Labor Economics*, 17(1): 80-108.
- Fairlie, R.W. (2003), An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models, *Economic Growth Center Discussion Paper No. 873, Yale University*.
- Fothergill, S. (2001) The true scale of the regional problem in the UK, *Regional Studies*, 35(3): 241-246.
- Georgellis, Y. and Wall, H.J. (2000), What Makes a Region Entrepreneurial? Evidence from Britain, *Annals of Regional Science*, 34(3): 385-403.
- Gomulka, J. and Stern, N. (1990), The Employment of Married Women in the United Kingdom 1970-83, *Economica*, 57(2): 171-199.
- Gray, D. (2004), Persistent Regional Unemployment Differentials Revisited, *Regional Studies*, 38(2): 167-176.
- Greene, W.H. (2002), *LIMDEP, version 8.0 for the PC*, Econometric Software, Inc.
- Gudgin, G. (1996) Regional problems and policy in the UK, *Oxford Review of Economic Policy*, 11(2).
- Lewis, J and Townsend, A. (1989), *The North-South Divide: Regional Change in Britain in the 1980s*, edited volume, Paul Chapman, London.
- Neumark, D. (1988), Employers' Discriminatory Wage Behavior and the Estimation of Wage Discrimination, *The Journal of Human Resources*, 23: 279-295.



Oaxaca, R. (1973), Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review*, 14(3): 693-709.

Oaxaca, R.L. and Ransom, M.R. (1994), On discrimination and the decomposition of wage differentials, *Journal of Econometrics*, 61: 5-21.

Reimers, C.W. (1983), Labor Market Discrimination Against Hispanic and Black Men, *Review of Economics and Statistics*, 65(4): 570-579.

Robson, M.T. (1998), Self-employment in the UK Regions, *Applied Economics*, 30: 313-322.

Rosa, P., Carter, S. and Hamilton, D. (1996), Gender as a Determinant of Small Business Performance: Insights from a British Study, *Small Business Economics*, 8(4): 463-478.

Shutt, J. and Sutherland, J. (2003) Encouraging Transition into Self-employment, *Regional Studies*, 37(1): 97-103.

**Table I: The self-employment probability by region – males and females separately.**

<i>Region</i>	<b>MALES</b>			<b>FEMALES</b>		
	<i>Total</i>	<i>S/E</i>	<i>Prob</i>	<i>Total</i>	<i>S/E</i>	<i>Prob</i>
<b>Greater London</b>	358	72	0.201	397	53	0.134
<b>South East England</b>	1302	274	0.210	1351	140	0.104
<b>South West England</b>	464	133	0.287	499	50	0.100
<b>East Anglia</b>	182	32	0.176	219	14	0.064
<b>East Midlands</b>	295	55	0.186	301	28	0.093
<b>West Midlands</b>	512	101	0.197	501	41	0.082
<b>Wales</b>	417	82	0.197	352	29	0.082
<b>Yorkshire &amp; The Humber</b>	555	101	0.182	582	39	0.067
<b>North West England</b>	581	107	0.184	628	53	0.084
<b>North of England GOR</b>	285	34	0.119	287	23	0.080
<b>Scotland</b>	462	64	0.139	529	25	0.047
<i>Unknown</i>	19	4	0.211	35	4	0.114
<b>GREAT BRITAIN</b>	5432	1059	0.195	5681	499	0.088
<b>Southern England</b>	2124	479	0.226	2247	243	0.108
<b>Central England</b>	989	188	0.190	1021	83	0.081
<b>Northern England</b>	1421	242	0.170	1497	115	0.077
<b>Wales &amp; Scotland</b>	879	146	0.166	881	54	0.061

**Table II: Job creation by the self-employed by region – males and females separately.**

<i>Region</i>	<b>MALES</b>			<b>FEMALES</b>		
	<i>Jobs</i>	<i>S/E</i>	<i>Mean</i>	<i>Jobs</i>	<i>S/E</i>	<i>Mean</i>
<b>Greater London</b>	164	67	2.448	391	50	7.820
<b>South East England</b>	762	265	2.875	265	139	1.906
<b>South West England</b>	302	131	2.305	80	50	1.600
<b>East Anglia</b>	119	32	3.719	23	13	1.769
<b>East Midlands</b>	315	55	5.727	61	27	2.259
<b>West Midlands</b>	305	100	3.050	60	40	1.500
<b>Wales</b>	179	82	2.183	155	29	5.345
<b>Yorkshire &amp; The Humber</b>	398	101	3.941	109	37	2.946
<b>North West England</b>	349	106	3.292	236	52	4.538
<b>North of England GOR</b>	100	33	3.030	41	22	1.864
<b>Scotland</b>	669	62	10.790	87	25	3.480
<i>Unknown</i>	2	4	0.500	6	4	1.500
<b>GREAT BRITAIN</b>	3664	1038	3.530	1514	488	3.102
<b>Southern England</b>	1228	463	2.652	736	239	3.079
<b>Central England</b>	739	187	3.952	144	80	1.800
<b>Northern England</b>	847	240	3.529	386	111	3.477
<b>Wales &amp; Scotland</b>	848	144	5.889	242	54	4.481

**Table III: Male probits of the self-employment probability – South versus North.**

<i>Variable</i>	<b>Southern England</b>			<b>Northern England</b>		
	<i>Estimate</i>	<i>Est./S.E.</i>	<i>Mean</i>	<i>Estimate</i>	<i>Est./S.E.</i>	<i>Mean</i>
<b>A level is highest</b>	-0.532	-3.71	0.121	0.044	0.23	0.080
<b>First degree is highest</b>	-0.415	-2.86	0.164	-0.196	-0.93	0.104
<b>Higher degree is highest</b>	-0.807	-2.82	0.028	-0.689	-1.60	0.019
<b>Professional qualification</b>	0.122	1.11	0.123	0.210	1.41	0.108
<b>Vocational qualification</b>	-0.219	-2.81	0.460	-0.093	-0.96	0.483
<b>Apprenticeship 1981</b>	0.171	3.61	0.466	0.186	3.20	0.645
<b>Maths High Aged 7</b>	0.026	0.29	0.201	0.158	1.35	0.182
<b>Maths Low Aged 7</b>	-0.197	-1.73	0.107	0.070	0.54	0.142
<b>Reading High Aged 16</b>	-0.215	-1.92	0.162	0.003	0.02	0.121
<b>Reading Low Aged 16</b>	0.299	2.63	0.097	-0.211	-1.50	0.141
<b>Creativity</b>	0.919	1.89	0.165	0.248	0.41	0.165
<b>Depression</b>	0.003	0.11	0.877	0.074	2.70	0.986
<b>Caution</b>	-0.081	-1.55	0.178	0.043	0.65	0.173
<b>Laziness</b>	0.123	3.26	-0.167	0.069	1.42	-0.091
<b>Moodiness</b>	0.060	1.48	-0.454	0.034	0.68	-0.432
<b>Sociability</b>	0.101	2.33	0.449	0.038	0.74	0.404
<b>Timidity</b>	-0.009	-0.14	0.025	-0.181	-2.31	0.004
<b>Own boss important 1981</b>	0.426	4.09	0.102	0.304	2.16	0.092
<b>Job security important 1981</b>	-0.262	-2.98	0.217	-0.225	-2.06	0.289
<b>Dad manager of small firm</b>	0.150	1.45	0.114	0.376	2.34	0.067
<b>Dad professional employee</b>	-0.099	-0.62	0.056	0.402	1.68	0.033
<b>Dad worker own account</b>	0.176	1.12	0.038	0.352	1.43	0.025
<b>Dad farmer employee-manager</b>	0.439	1.49	0.010	1.083	2.58	0.007
<b>Dad farmer own account</b>	0.940	2.55	0.007	2.318	3.75	0.004
<b>Dad's years of education</b>	0.043	1.60	2.912	-0.033	-0.73	2.637
<b>Mum's years of education</b>	0.005	0.15	2.952	0.100	2.00	2.653
<b>No children</b>	-0.198	-2.36	0.353	-0.005	-0.04	0.274
<b>No children * higher quals</b>	0.248	1.61	0.142	-0.211	-0.88	0.072
<b>Inheritance</b>	0.047	3.02	0.687	0.103	1.41	0.243
<b>Inheritance squared</b>	-0.012	-0.54	0.184	-0.276	-0.98	0.018
<b>Year of inheritance</b>	-0.235	-2.55	0.268	-0.117	-0.82	0.192
<b>Sub-region 1</b>	-0.266	-2.49	0.169	0.269	2.14	0.391
<b>Sub-region 2</b>	-0.262	-3.35	0.613	0.305	2.43	0.409
<b>Constant</b>	-0.842	-3.19	1.000	-2.003	-4.55	1.000
Log-likelihood	-1019.283			-589.172		
Sample size	2124			1421		
Mean of dependent variable	0.22552			0.17030		

**Table IV: Male logit – decomposition of the self-employment probability gap.**

<i>Version</i>	<i>Compositional Term</i>	<i>Structural 1</i>	<i>Structural 2</i>	<i>Overall gap</i>
<b>Equation (1)</b>	$0.2255 - 0.2384 = -0.0128$	NIL	0.0680	0.0552
<b>Equation (2)</b>	$0.1788 - 0.1703 = +0.0085$	0.0467	NIL	0.0552
<b>Equation (4)</b>	$0.2030 - 0.2040 = -0.0011$	0.0226	0.0337	0.0552
<b>Reimers</b>	$0.1949 - 0.1974 = -0.0026$	0.0306	0.0271	0.0552
<b>Cotton</b>	$0.1997 - 0.2045 = -0.0047$	0.0258	0.0342	0.0552

**Table V: Male logit of the self-employment probability – breakdown of term 2 in (3).**

<i>Version</i>	<i>Regressor category (see Section 2)</i>					<i>Overall</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>7</i>	
<b>Equation (1)</b>	-0.1625	0.0157	0.0444	0.0038	0.0142	-0.0844
<b>Equation (2)</b>	-0.0536	0.0001	0.1162	-0.0482	-0.0143	+0.0004
<b>Equation (4)</b>	-0.1099	0.0121	0.0692	0.0025	0.0110	-0.0151
<b>Reimers</b>	-0.1080	0.0079	0.0804	-0.0222	-0.0001	-0.0420
<b>Cotton</b>	-0.1188	0.0095	0.0732	-0.0171	0.0028	-0.0504

**Table VI: Female probits of the self-employment probability – South versus North.**

<i>Variable</i>	Southern England			Northern England		
	<i>Estimate</i>	<i>Est./S.E.</i>	<i>Mean</i>	<i>Estimate</i>	<i>Est./S.E.</i>	<i>Mean</i>
<b>O level equivalent is highest</b>	0.144	1.41	0.409	0.039	0.30	0.444
<b>A level is highest</b>	0.285	1.93	0.123	-0.115	-0.48	0.081
<b>First degree is highest</b>	0.155	1.01	0.149	-0.261	-1.06	0.101
<b>Professional qualification</b>	0.073	0.55	0.081	0.135	0.69	0.066
<b>Vocational qualification</b>	-0.027	-0.32	0.338	-0.026	-0.23	0.304
<b>Apprenticeship 1981</b>	0.212	2.37	0.083	0.247	1.93	0.068
<b>Reading High Aged 7</b>	0.084	0.85	0.212	0.255	1.98	0.219
<b>Reading Low Aged 16</b>	-0.125	-0.82	0.093	0.212	1.32	0.145
<b>Depression</b>	0.058	1.97	0.701	-0.011	-0.24	0.687
<b>Caution</b>	-0.091	-1.50	0.152	-0.002	-0.03	0.229
<b>Timidity</b>	-0.144	-2.02	0.069	-0.080	-0.88	0.100
<b>Promotion important 1981</b>	-0.236	-1.28	0.054	0.073	0.30	0.041
<b>Own boss important 1981</b>	0.921	5.77	0.033	0.399	1.35	0.021
<b>Job security important 1981</b>	-0.199	-1.53	0.123	-0.138	-0.88	0.145
<b>Family financial difficulties</b>	-0.007	-0.04	0.045	-0.373	-1.47	0.068
<b>Dad manager of small firm</b>	0.169	1.40	0.107	0.180	0.98	0.084
<b>Dad professional self-emp</b>	-0.120	-0.30	0.010	1.531	2.26	0.003
<b>Dad worker own account</b>	0.116	0.56	0.034	0.645	2.29	0.023
<b>Dad farmer employee-mngr</b>	0.742	2.07	0.007	0.571	1.25	0.008
<b>No children</b>	-0.234	-1.85	0.249	-0.309	-1.77	0.206
<b>No children * higher quals</b>	0.179	0.97	0.112	0.551	1.94	0.078
<b>Inheritance</b>	0.052	2.81	0.748	0.203	1.16	0.585
<b>Inheritance squared</b>	-0.041	-1.86	0.410	-3.768	-0.99	2.351
<b>Sub-region 1</b>	0.144	1.19	0.177	-0.100	-0.69	0.389
<b>Sub-region 2</b>	0.055	0.58	0.601	0.040	0.29	0.420
<b>Constant</b>	-1.407	-4.92	1.000	-1.071	-2.29	1.000
Log-likelihood	-709.28			-375.06		
Sample size	2247			1497		
Mean of dependent variable	0.10814			0.07682		

**Table VII: Female logit – decomposition of the self-employment probability gap.**

<i>Version</i>	<i>Compositional Term</i>	<i>Structural 1</i>	<i>Structural 2</i>	<i>Overall gap</i>
<b>Equation (1)</b>	0.1081 – 0.0923 = +0.0159	NIL	0.0154	0.0313
<b>Equation (2)</b>	0.0828 – 0.0769 = +0.0059	0.0253	NIL	0.0313
<b>Equation (4)</b>	0.1015 – 0.0868 = +0.0147	0.0067	0.0099	0.0313
<b>Reimers</b>	0.0894 – 0.0810 = +0.0084	0.0188	0.0041	0.0313
<b>Cotton</b>	0.0917 – 0.0826 = +0.0091	0.0164	0.0057	0.0313

**Table VIII: Female logit of the self-employment probability – breakdown of term 2 in (3).**

<i>Version</i>	<i>Regressor category (see Section 2)</i>					<i>Overall</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>7</i>	
<b>Equation (1)</b>	0.0500	0.0465	0.0132	0.0197	0.0170	+0.1464
<b>Equation (2)</b>	-0.0470	0.0257	0.0460	-2.7322	0.0325	-2.6752
<b>Equation (4)</b>	0.0217	0.0418	0.0228	0.1626	0.0310	+0.2799
<b>Reimers</b>	0.0015	0.0361	0.0296	-1.3563	0.0247	-1.2644
<b>Cotton</b>	0.0111	0.0381	0.0263	-1.0834	0.0232	-0.9846

**Table IX: Male tobits of self-employment job creation – South versus North.**

<i>Variable</i>	Southern England			Northern England		
	<i>Estimate</i>	<i>Est./S.E.</i>	<i>Mean</i>	<i>Estimate</i>	<i>Est./S.E.</i>	<i>Mean</i>
<b>O level equivalent is highest</b>	2.402	1.06	0.380	3.559	1.97	0.375
<b>A level is highest</b>	10.388	2.46	0.076	5.877	1.58	0.079
<b>First degree is highest</b>	-2.983	-0.66	0.110	12.213	2.91	0.092
<b>Higher degree is highest</b>	-13.474	-1.07	0.009	10.457	1.05	0.008
<b>Professional qualification</b>	11.164	3.21	0.093	7.847	2.71	0.117
<b>Vocational qualification</b>	-0.309	-0.13	0.467	-2.447	-1.32	0.496
<b>Apprenticeship 1981</b>	0.147	0.10	0.616	1.850	1.68	0.783
<b>Maths High Aged 7</b>	-5.402	-1.98	0.184	1.051	0.50	0.204
<b>Maths High Aged 16</b>	2.189	0.67	0.143	-5.073	-1.74	0.150
<b>Reading High Aged 16</b>	5.188	1.36	0.102	7.279	2.24	0.117
<b>Unforthcomingness</b>	-1.501	-2.03	1.238	0.707	1.33	1.446
<b>Withdrawal</b>	0.076	0.05	0.317	-2.358	-1.82	0.400
<b>Hostility to (other) children</b>	-2.381	-1.54	0.257	0.926	1.17	0.392
<b>Caution</b>	3.915	2.61	0.022	-1.021	-0.77	0.088
<b>Moodiness</b>	-1.503	-1.43	-0.268	1.097	1.39	-0.325
<b>Timidity</b>	-3.906	-2.09	-0.076	1.895	1.15	-0.133
<b>Fear new situations</b>	1.103	0.37	0.143	-4.001	-1.38	0.117
<b>Promotion important 1981</b>	-1.014	-0.28	0.076	5.864	1.61	0.050
<b>Own boss important 1981</b>	-0.999	-0.37	0.168	1.493	0.62	0.179
<b>Job security important 1981</b>	-2.904	-1.04	0.173	1.967	0.86	0.213
<b>Dad professional employee</b>	14.679	2.82	0.039	-4.122	-0.88	0.046
<b>Dad professional self-emp</b>	-17.789	-1.82	0.011	29.182	2.27	0.004
<b>Dad worker own account</b>	-7.736	-1.61	0.058	3.125	0.80	0.038
<b>Dad farmer employee-mngr</b>	1.173	0.18	0.017	-13.100	-1.60	0.021
<b>Dad's years of education</b>	2.157	3.13	2.765	0.840	1.07	2.692
<b>No children</b>	-0.198	-0.08	0.289	-0.487	-0.23	0.254
<b>No children * higher quals</b>	-6.174	-1.23	0.089	-6.397	-1.22	0.058
<b>Inheritance</b>	2.270	1.77	1.290	4.503	2.17	0.397
<b>Inheritance squared</b>	-19.527	-1.78	0.651	-49.386	-1.56	0.044
<b>GOR unemployment rate</b>	-0.239	-0.15	6.527	0.761	0.87	10.995
<b>Years self-employed</b>	0.331	1.39	4.708	-0.021	-0.10	4.375
<b>Not full time</b>	0.508	0.18	0.168	-1.633	-0.75	0.225
<b>Constant</b>	-16.011	-1.40	1.000	-17.791	-1.65	1.000
<b>Sigma</b>	15.988	18.39	N/A	9.627	14.06	N/A
Log-likelihood	-898.87			-479.44		
Sample size	463			240		
Mean of dependent variable	2.65227			3.52917		