

Curb Your Enthusiasm: Optimistic Entrepreneurs Earn Less

EUROPEAN ECONOMIC REVIEW (forthcoming)

Accepted for publication on 30 August 2018

Final author post-print version

Christopher Dawson*, David de Meza†, Andrew Henley‡, G. Reza Arabsheibani†

Abstract

This paper concerns the implications of biased beliefs on entrepreneurial earnings. Amongst self-employed business owners, income is decreasing in optimism measured whilst still an employee. Controlling for earnings in paid employment, self-employment earnings of those with optimism above the mean are some 30% less than those with optimism below the mean. For employees, it is optimists that have higher earnings. These and associated results suggest that mistaken expectations lead to entry errors. As a test of external validity, future divorcees turn out to be financial optimists, indicating our measure captures an intrinsic psychological trait associated with rash decisions.

Keywords: Financial optimism, expectations, entrepreneurs

JEL Classification: D84, M13

* School of Management, University of Bath; † London School of Economics; ‡ Cardiff Business School, Cardiff University

Acknowledgement: We are grateful to two referees, Thomas Åstebro, Jordi Blanes-i-Vidal, Hans Hvide, Ross Levine, Yona Rubinstein, Robert Sauer, Marcia Schafgans, an associate editor and participants in the 5th HEC Entrepreneurship Workshop for their very helpful comments.

“What wild imaginations one forms where dear self is concerned! How sure to be mistaken!”

Jane Austen

“The first principle is that you must not fool yourself and you are the easiest person to fool.”

Richard P. Feynman

1. Introduction

A common theme of self-help books, exemplified by Norman Vincent Peale’s influential “The Power of Positive Thinking” (1952), is “When you expect the best, you release a magnetic force in your mind which by a law of attraction tends to bring the best to you.”¹ There is now considerable evidence that beliefs matter for performance. Some of the studies are summarised in Compte and Postlewaite (2004), who argue that biased expectations may therefore be optimal. Nevertheless, there is a downside. Incorrect forecasts tend to result in mistaken decisions and hence worse outcomes. Self-belief may enhance performance but also result in participation in activities doomed to failure.

This paper examines how these forces play out in start-ups, a big decision with many uncertainties. Optimists overweight the upside, and so tend to self-select into self-employment, as an increasing number of studies find.² More optimistic individuals may mistakenly think they have identified good opportunities and, therefore, tend to switch too soon and into objectively poor projects.³ These are reasons why optimism may be associated with lower self-employment earnings. Nevertheless, if Hamlet is right and “..thinking makes it so.”, optimists may outperform.

¹ Peale was Donald Trump’s childhood pastor.

² For example, Arabsheibani et al. (2000), Cassar (2010), Puri and Robinson (2007, 2013) and Dawson et al. (2014).

³ The reasoning is that of de Meza and Southey (1996), Camerer and Lovello (1999) and Malmendier and Tate (2008) for why optimistic CEOs are more likely to make value-destroying acquisitions. Optimism has other implications for entrepreneurship such as a preference for debt (de Meza and Southey (1996) tested by Landier and Thesmar (2009)).

The main finding is that prior optimism is negatively associated with the subsequent earnings of the self-employed. Controlling for earnings as an employee, self-employed pessimists earn some 30% more than optimists. In contrast, amongst employees, optimists earn more.⁴ Entrepreneurial optimism implies that as far as private returns are concerned, entry into entrepreneurship is sometimes mistaken. This is a reason for caution in adopting policies that encourage start-ups. Our results provide tentative support that optimism does actually matter in this regard.

Two other papers look at how aspects of preferences affect entry into self-employment and subsequent earnings. In Hvide and Panos (2013), the taste parameter is risk preference, proxied by stock market participation and personal leverage. According to reduced form estimates, risk tolerance encourages entry but depresses earnings. The interpretation is a selection effect, that more risk tolerant types accept lower expected return projects. Hamilton et al. (2014) study the effect of the “big five” personality traits. Personality potentially affects relative earning power in paid and self-employment, as well as relative non-pecuniary attraction. A structural model is estimated using simulated maximum likelihood to identify these selection and treatment effects. Self-employment is found to be attractive to those open to new experience but lowers its expected financial returns. According to the model, the sign of selection and treatment effects on earnings is the same. Both of these papers invoke rational expectations. In our case, the effect of forecast bias is investigated. Unlike the other papers, where the explanatory variables are preference based, systematic error implies a potential case for policy intervention to offset the bias.

⁴ There is evidence that the economic return to self-employment is low. According to Hamilton (2000), the median self-employed worker earns less than they would as an employee. Similarly, Moskowitz and Vissing-Jorgenson (2002) find that the return on the equity invested in private businesses does not compensate for the risk involved. Levine and Rubinstein (2017) find that owners of incorporated businesses increase their gross earnings relative to paid employment. It is unincorporated businesses that do worse. Åstebro et al. (2013) find that sole proprietors suffer large income falls relative to their employee earnings. Optimism is a possible explanation for low earnings along with underreporting to evade tax (Astebro and Chen, 2014), preference for autonomy (Hurst, Li and Pugsley, 2014) and rational experimentation (Kerr et al., 2014). Åstebro et al. (2015) find some experimental evidence of skewness loving, possibly because large prizes are the most salient.

Hurst and Pugsley (2011) document that most businesses start small, remain small and do not innovate. This suits most owners, since the most common reason given for starting a business is desire for autonomy. If start-ups mostly represent “lifestyle” choices, it is argued that they create few positive externalities and therefore explicit and implicit subsidies should be eliminated. This conclusion is reinforced by overexpansion of self-employment due to opportunities for tax evasion and avoidance and the non-taxability of non-pecuniary benefits, explored in Hurst and Pugsley (2016). Although many of those setting up businesses may be knowingly foregoing expected income, our paper provides evidence that the fall in income is systematically underestimated. The implication is that if expectations were rational, there would be fewer start-ups.

The next section sets out the analytical issues. Section 3 describes the data and discusses the implementation of the method. Results follow in Section 4. As a test of the robustness, Section 5 examines whether financial optimists make rash decisions in other spheres. Optimists are more likely to make poor marriage matches resulting in divorce and to be heavy smokers. Finally, brief conclusions are drawn.

2. Optimism and Earnings: Theory

This section provides the theoretical underpinning for the empirical finding that the sign of the optimism effect on earnings of the self-employed is more negative than its effect on employees. As optimists are also more likely to be self-employed, this cannot easily be reconciled with the first observation if the optimism measure merely proxies for some unobserved productivity attribute. The analysis is developed in two steps. First, the pure selection effects on intrinsic optimism are established. Then the complications arising when optimism is estimated as forecast error are addressed.

Entry into self-employment can be considered as a choice based on perceived comparative advantage. In the spirit of Lazear (2004), suppose earnings in both paid and self-employment depends on unobserved entrepreneurial ability, z , and an observed attribute, x . An individual's expected earnings in self-employment are $w_s = S(z, x)$ and in paid employment are $w_p = P(z, x)$.⁵ It is likely that z has a smaller effect in paid employment. Assume that everyone starts out in paid employment, then has the opportunity to switch to self-employment. Also, for simplicity, that choices are made to maximize expected earnings. As illustrated in Figure 1, under rational expectations, conditional on the observable taking value \bar{x} , individuals with z above z^r become self-employed. Average earnings are higher in self-employment as the only reason to switch from paid employment is to boost earnings.

Optimists overestimate returns in self-employment relative to paid employment.⁶ Specifically, setting up a business gives more scope for optimism than continuing as an employee, as proposed by de Meza and Southey (1996). Evidence that the self-employed are indeed more optimistic than employees is provided by, amongst others, Arabsheibani et al. (2001), Puri and Robinson (2007, 2013) and Åstebro et al. (2007). Dawson et al. (2014) find that optimism predates self-employment but is increased by self-employment.

⁵ This formulation assumes that returns do not depend on the numbers choosing each option. For example, if more restaurants are opened, this may depress returns to all, as in de Meza and Southey (1996). As this paper is concerned with the effect of individual differences in optimism this crowding effect can be ignored.

⁶ Usage is not settled. Many economists (e.g. Hvide, 2002), including us, consider optimism to be a self-serving bias in an estimate whilst excessive precision in the estimate (an overly narrow confidence interval) is overconfidence. Overconfidence sometimes covers both meanings. For some optimism is sometimes reserved for bias in the estimation of own ability as opposed to of favourable external events. Bengtsson and Ekeblom (2014) find that in Sweden, the self-employed are more optimistic about macro-economic variables than employees. Psychologists typically do not regard optimism as a forecast error but an upbeat attitude or a belief that good things will happen (as in the LOT-R general optimism inventory). For some individuals, this is a rational expectation, in which case they are not optimists in the sense of making self-serving errors. Moore and Healy (2008) (see also Åstebro, Nande and Weber, 2014) distinguish between *overestimation* of the individuals own ability or performance, *overplacement* where individuals assess their ability rank too highly, and *overprecision*, excessively narrow confidence intervals (overconfidence in terms of the previous footnote). Åstebro, Jeffrey, and Adomdza (2007) find evidence that inventor-entrepreneurs exhibit greater overestimation than the general population over performance in general knowledge testing as well as in LOT-R style general optimism.

For now, suppose that there are just two types of expectation: rational or optimistic. The perceived returns of an optimist in self-employment are $w_s = O + S(z, x)$ with $O > 0$.⁷ What matters is that optimism makes self-employment attractive relative to paid employment. For simplicity, it will be assumed that optimism does not affect perception of returns in paid employment at all. In Figure 1, the threshold for an optimist to enter self-employment is $z^o < z^r$. If the distribution of z is the same for optimists and realists, controlling for observables, optimists are more likely to enter self-employment than realists but earn less on average.⁸

It is possible that the distribution of z is different for optimists and realists. Suppose, for example, the z distribution of optimists is shifted to the right relative to realists. This augments the tendency for optimists to enter self-employment, but tempers the tendency for optimists to earn less. The combination of high entry into self-employment and low earnings is unlikely to be generated by omitted-variable bias.

⁷ Whether the optimism boost is additive, multiplicative or some other form is immaterial for what follows.

⁸ An extension to the model is that the self-employment opportunity available to an individual may be the result of a stochastic draw. The individual's unobserved project quality can be included in the S function. Optimism now concerns project quality as well as own attributes but the implications are the same.

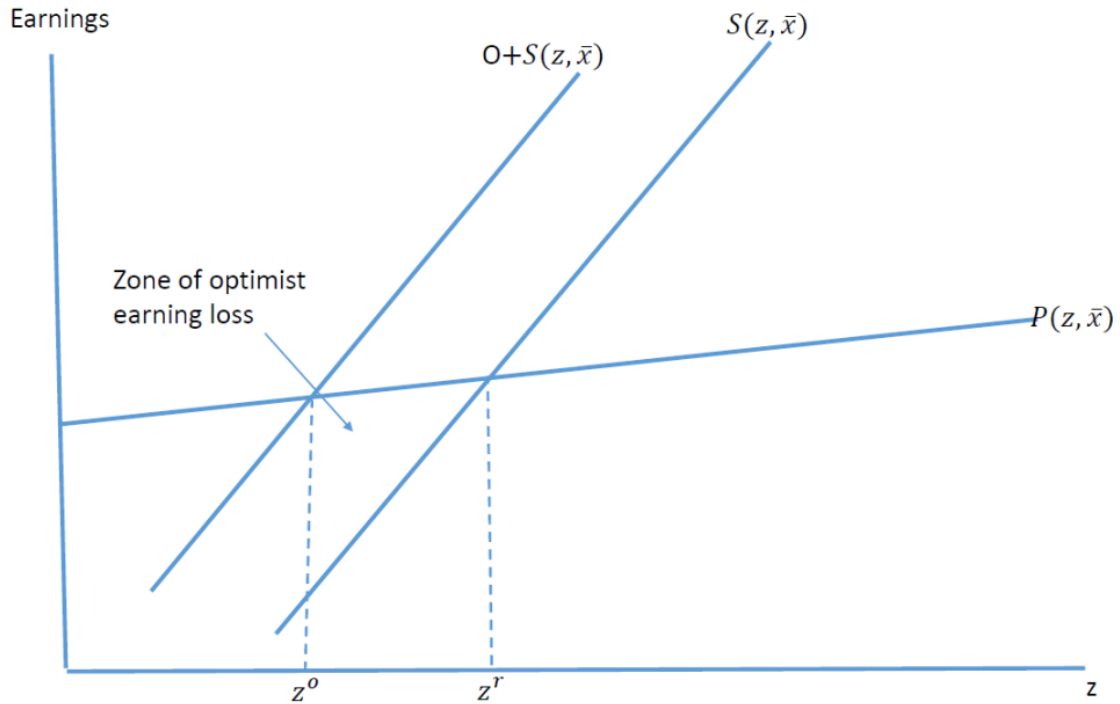


Figure 1: Selection in to self-employment

In bringing the model to the data, the complication is that optimism is not directly observed but estimated from earlier earnings forecast error. This has the advantage of directly concerning the relevant bias, but as earnings are a component of optimism, care must be taken to ensure that any relationship is not purely mechanical. Optimism is therefore measured in periods before its effects are estimated. Even so, the effects of measurement error must be accounted for. Unlucky income realizations raise measured optimism, but if the shock is transitory, optimism will be associated with higher future earnings, or unchanged future earnings if the shock is permanent. Measurement error is the only way optimism affects the earnings of those continuing in paid employment, and, as such, its coefficient will tend to be positive. In self-employment, the selection effect offsets measurement error, so the overall optimism effect could be negative.

Making these points more formally, the effect of optimism on forecast is

$$forecast_{it} = rational\ expectation_{it} + O_i \quad (1)$$

where the forecast is for period t made at $t - 1$ knowing the employment mode at t . O_i is intrinsic optimism. Realized earnings are

$$earnings_{it} = rational\ expectation_{it} + \varepsilon_{it} + p_{it} \quad (2)$$

where ε_{it} is a transitory income shock and p_{it} is a permanent shock. Measured optimism, as distinct from intrinsic optimism, is

$$measuredoptimism_{it} = forecast_{it} - earnings_{it} = O_i - \varepsilon_{it} - p_{it} \quad (3)$$

For an individual remaining in paid employment, rationally expected earnings only differ in each period by the income shocks,⁹ so from (1), (2) and (3)

$$\begin{aligned} earnings_{it} &= rational\ expectation_{it-1} + \varepsilon_{it} + p_{it} + p_{it-1} \\ &= earnings_{it-1} + measuredoptimism_{it-1} - O_i + \varepsilon_{it} + p_{it-1} \\ &\quad + p_{it} \quad (4) \end{aligned}$$

The earnings equation (4) is the basis for empirical estimation. Realized earnings are observable, as is measured optimism. Intrinsic optimism is not observable, but as it is a component of measured optimism with opposite sign to its direct appearance in the equation, its magnitude has no influence on earnings. The same is true of permanent shocks.¹⁰ However, measured optimism is decreasing in lagged transitory shocks, which do not otherwise appear in (4). For employees, future earnings are therefore increasing in past, measured optimism. This is not due to selection effects but measurement error. There is no selection effect because the unobserved variables that influence earnings in t had the same effect at $t - 1$, so are effectively controlled for in the future earnings equation. This is not the case for the self-employment earnings equation because the unobservables play a different role in the two modes.

⁹ It is possible that as careers develop some characteristics play different roles. In that case the effect of unobserved variables may not be perfectly captured by first-period earnings.

¹⁰ As in Gervais and O'Dean (2001), it may take time to adjust to a negative permanent shock during which time optimism prevails.

For those moving in to self-employment,

$$\begin{aligned}
 earnings_{it} &= rational\ expectation_{it-1} + f(\bar{x}, O_i) + \varepsilon_{it} + p_{it} + p_{it-1} \\
 &= earnings_{it-1} + measuredoptimism_{it-1} - O_i + f(z_i, \bar{x}, O_i) + \varepsilon_{it} + p_{it} \\
 &\quad + p_{it-1} \quad (5)
 \end{aligned}$$

In the first expression of (5), the rational expectation is for earnings at $t - 1$ when the individual is still in paid employment. To obtain self-employment earnings at t , an adjustment must be made to take into account that individuals earning the same in paid employment are differentially suited to self-employment. The extent to which adjustment is needed depends on the individual's z . This is unobservable, but due to selection the mean value of z is decreasing in optimism. Hence the adjustment function, $f(\bar{x}, O_i)$.

The properties of (5) with respect to measured optimism depend on the reason for variation. If measured optimism is high due to a transitory shock at $t - 1$, earnings will be higher at t , just as in the paid employment earnings function, (4). Higher O_i raises measured optimism but the earnings effect is no longer exactly cancelled out due to the $f(\bar{x}, O_i)$ term which imparts a negative effect.

The analysis has so far assumed that intrinsic optimism does not have a direct productivity effect. This is not necessarily the case. Incorrect expectations may mean that optimists take poor operating decisions, which may be particularly important for those running their own business. Optimism may also have positive effects. For example, as argued by Trivers (2000), optimism may have evolved to influence others. The best way to convince others of your competence is really to believe in it yourself. Self-deception begets effective deception. Some evidence of this role is provided by Adomdza et al. (2016) and Schwardmann and Van der Weele (2016). For employees, the target of influence is most obviously the boss, but could include customers or suppliers. The latter two influences are

also potentially relevant for the self-employed, as is the ability to persuade financiers. Effort choice may also be affected by optimism.¹¹ In principle, the effect could run in either direction. Optimists may apply more effort because they overestimate its marginal effectiveness or decrease effort because they believe that even with low effort success is assured, albeit with decreasing marginal returns.¹² For those remaining in paid employment, any productivity effects of optimism will be reflected in first period earnings. Given the second-period earnings function has first-period earnings as a control, it remains true that optimism only figures as a result of transitory shocks. As the productivity effect of optimism may be different in self-employment and paid employment, it may not be neutral in the self-employment earnings function. Denote self-employment earnings of an optimist as $w_s = S(z, x, O)$. In Figure 1, the earnings schedule of the optimistic self-employed is now distinct from the realistic. This will move the propensity to enter and earnings of optimists in opposite directions. Whatever the outcome, optimism still leads to excess entry.

In summary, the equilibrium relationships implied by self-selection and the optimism measure are as follows;

1. The optimism coefficient in the paid employment earnings function will be positive.
2. In the self-employment earnings function, the optimism coefficient will be lower than in paid employment and may be negative if selection effects are strong enough.
3. Optimists are more likely to be self-employed.

3. Data source and methods

¹¹ In the empirical work it is possible to control for variation in hours.

¹² Moscarini and Fang (2005) show that if optimism is good for incentives, employers may be better off preserving illusions by not tailoring offers to individual productivity. That is, optimism may lead to wage compression.

In outline, the method is to use the initial years of a large and long panel dataset, which includes forecasts and realizations, to estimate individual optimism and earnings capability in paid employment. Subsequent years of data are then used to determine how optimism impacts earnings controlling for past earnings and realizations.

3.1. The BHPS survey instrument

The data source for the analysis is the British Household Panel Survey (BHPS), a nationally representative longitudinal survey initiated in 1991 and funded by the UK Economic and Social Research Council as an internationally comparative multi-purpose research resource. A stratified random cluster sample of households, drawn from the population of British household postal addresses in Great Britain, is tracked annually.¹³ Each wave includes household and individual questionnaire instruments, the latter completed separately by all adult household members present at each wave. Follow-on rules ensure the tracking of any newly formed households involving originally enumerated household members. The individual instrument covers a range of topics including demographic characteristics, economic activity, and finances, and includes some recall items on family background, education and employment history. The original sample of approximately 5000 households (comprising around 12000 individuals) was recruited in 1991. This study uses data from 18 annual waves available between 1991 and 2008.¹⁴

¹³ The far north of Scotland is excluded because of the prohibitive sampling costs. The original survey excludes Northern Ireland. Booster samples for Wales and Scotland recruited in 1999 and a sample for Northern Ireland recruited in 2001 are excluded from the analysis.

¹⁴ Sample attrition rates in the BHPS are generally low and certainly comparable to those achieved in other similar household panels. As is typical with household panels the highest attrition rate of individuals was between Waves 1 and 2 (12%). Attrition between Waves 2 and 3 was 7% of the original individuals and subsequently averaged 2.4% of the original sample between waves. In common with nearly all previously published research using this data source, attrition is assumed to be a random event. From 2009 onwards the BHPS sample has been merged into a much larger new longitudinal household study with further widening of scope, including biosocial analysis. However, some reductions in questionnaire detail yields the larger dataset unsuitable for the present analysis.

The BHPS survey design has evolved to incorporate a number of regional boosts, however the sample used in the analysis is restricted to the originally enumerated sample across Great Britain (i.e. excluding Northern Ireland), and to those individuals who are either in paid employment or self-employed. Self-employed is defined here as those who self-identify as self-employed business owners. This is checked by the interviewer against their UK tax status, under which those who declare themselves to be self-employed are responsible for own income tax declarations and payments, rather than directly through employer-made deductions. Freelancers and subcontractors who may be self-employed for tax purposes but are not business owners are excluded from the definition and the analysis, drawing on information in a questionnaire item about the nature of the self-employment. This leaves approximately 80% of the self-employed who are business owners.

3.2. Defining and measuring intrinsic optimism

In establishing the relationship between optimism and earnings, the first step is to construct a measure of optimism defined as an excessive belief in the probability of good financial realizations. The measure of optimism is forecast error, the challenge being to separate systematic bias (intrinsic optimism) from random error. Positive (negative) errors may just reflect bad (good) luck. A further issue is that bias may depend on what it is that is being forecast. Optimism is greater when individuals believe events are under their control (e.g. Harris, 1996)¹⁵ and when the task is difficult (Lichtenstein and Fischhoff, 1977).¹⁶ Ideally, the forecasting task should be reasonably uniform across individuals and similar to that in the setting of study.

The optimism measure is constructed from two questionnaire items on financial expectations and realizations, asked of all individuals in each year. The first is:

¹⁵ The “illusion of control” (Langer, 1975) is the excessive belief that an individual can influence events.

¹⁶ Starting a business normally involves both characteristics.

“Looking ahead, how do you think you yourself will be financially a year from now; better than you are now, worse than you are now, or about the same?”

Individuals who gave a valid response at year t are then matched with their self-reported financial realization at year $t+1$, obtained from the second question:

“Would you say that you yourself are better off, worse off or about the same financially than you were a year ago?”

The survey instrument asks for responses to both questions on three-point scales. So the empirical approach set out in section 2 must be adapted to the categorical nature of these forecast and realization data. To measure optimism from data of this type we follow Das and van Soest (1997), Arabsheibani et al. (2000), and Souleles (2004) in constructing a five-point measure of forecast error, defined as the difference between the financial forecast (of $t+1$) at t , minus the financial realization at $t+1$.¹⁷ As our optimism measure is based on financial forecast error, to determine the influence of optimism on earnings there must be no overlap in the time periods covered by these variables. For instance, in a cross-sectional approach, random negative shocks occurring after the forecast is made mechanically imply optimism and earnings are negatively correlated. To eliminate this concern, optimism is computed for two groups of individuals who will be referred to throughout the analysis as *futures* and *nevers*. *Futures* are those currently in paid-employment who become self-employed later in the panel. For this group, the optimism measure is computed over their period of paid employment prior to entry into self-employment. Data for the year prior to transition into

¹⁷ This procedure involves cardinalization of the forecast error. For example, forecasting *better* and achieving *same* is treated as equivalent to forecasting *same* and achieving *worse*. Although the five-point scale is commonly used, there is no fundamental defence of the procedure beyond saying it represents a convenient mapping from continuous but unobserved underlying forecasts and realizations. If the specification is wrong, it will make it harder to find optimism effects.

self-employment is excluded as forecasts may be associated with unusually low financial outcomes if the switch to self-employment was occasioned by involuntary severance from paid employment, and therefore not anticipated. *Nevers* are those who remain as employees over the full period covered by the dataset. For this group, optimism is computed over the first half of available years in paid-employment (specifically the next highest integer to the midpoint number of years).¹⁸

The categorical nature of the data is a drawback, but the longitudinal feature is an advantage since it allows more precision in identifying intrinsic optimism. Averaged over a number of periods, the noise in the optimism measure will be diminished though not completely eliminated. To take advantage of this property, a linear fixed-effect regression is estimated for all those in the sample as follows:

$$M_{it} = X'_{it}B + \hat{\delta}_i + \varepsilon_{it} \quad (6)$$

where M_{it} is the forecast error by individual i at time t . X'_{it} is a vector of time-varying demographic and other person-specific characteristics of individual i , as well as region and year dummy variables. For *futures* the observations are for the period in paid employment, and for *nevers* it is the first half of the period for which they are observed. The individual fixed effects in this regression, $\hat{\delta}_i$, are extracted to provide estimates of intrinsic optimism net of any environmental influences from location and time and any changes to individual circumstance. These fixed effects are used as regressors in the second-phase earnings equation.

Table A1 in the Appendix provides summary statistics for the *nevers* and *futures* in their first-phase when everyone is an employee. In total, there are 31,968 observations from

¹⁸ A transition into self-employed business ownership is defined to have occurred if an individual's full-time or main economic status changes to that state. A small number of transitions into part-time self-employment alongside full-time or part-time paid employment are excluded from the self-employed. Only the first spells of paid and self-employment are included in our sample. Few individuals start in self-employment and they are excluded.

7,985 individuals. Of these, there are 3,138 *futures* observations from 618 individuals. For *nevers* there are 28,830 observations from 7,367 individuals. Intrinsic optimism is therefore constructed from an average of 5.1 observations per individual for *futures* and 3.9 observations per individual for *nevers*. The average financial forecasts of *futures* exceed those of *nevers*, but average realizations are only marginally lower for *futures*. The forecast error is in the optimistic direction for both groups but *futures* are more optimistic than *nevers*.¹⁹

The full estimate of equation (6), which provides the optimism estimates, is in Column 1 of Table A2. It includes a range of demographic status, education and housing tenure status variables that might *a priori* be reasonably expected to influence financial forecasting or realizations. Although the controls are jointly significant, few are individually significant. The fixed effects from this equation are our estimates of intrinsic optimism though it should be noted that using simple averages of forecast error yield similar second-stage results as to whether optimism is correlated with subsequent earnings in self-employment for *futures* and paid-employment for *nevers*.

3.3 Earnings definition and measurement

Gross monthly self-employment earnings are computed as follows. Approximately 82% of self-employed business owners prepare annual accounts and so provide estimates of their share of profits (Table A4). Earnings data for the remainder are taken as the response to a supplementary question about pre-tax monthly self-employment earnings. The BHPS public release data file merges these into a single derived variable, measuring estimated monthly gross self-employment earnings. Annual loss data is available for the self-employed, but only for respondents who prepare annual accounts. For this reason, the single derived self-

¹⁹ Some *nevers* may enter self-employment later, in which case the tendency is to under record the extent of the optimism difference with *futures*.

employment earnings variable sets earnings to zero for those whose annual reported profits are negative.²⁰ Because this treats the earnings distribution as left censored, a Tobit estimator is used in the earnings specification described in the next section. To allow for any systematic difference in measurement error between the two response types, the self-employed earnings regression includes dummy variables to control for data reporting method. For *nevers*, earnings are defined as gross monthly salary from main paid employment job. Table A4 reports the mean and percentiles for gross monthly earnings by employment status. These reveal that the paid-employed have a relative advantage at the lower end of the distribution, but that above the 90th percentile the self-employed have a relative advantage. This earnings pattern will therefore appeal to optimists.

3.4 Earnings equation specification

The relationship between prior optimism and earnings is estimated by means of two earnings regressions. One is for the self-employed business owners who were previously *futures* and the other for *nevers* estimated over the second half of their employment period. Following equations (4) and (5), these take the form:

$$E_{it} = \text{constant} + \alpha \hat{\delta}_i + \beta \text{pastearnings}_i + \gamma \text{pastrealizations}_i + \sum_j \delta_j z_{jit} + \varepsilon_{it} \quad (7)$$

where E_{it} is gross monthly earnings and z_{it} observable characteristics including age and hours of work and $\hat{\delta}_i$ is the standardized fixed effect from the first-stage optimism equation.²¹

To emphasize, there is no overlap in the periods over which the first-stage variables and E_{it} are measured. The primary interest is in the sign and significance of α .

²⁰ Incorporating the available negative earnings data does not materially affect any of the subsequent results.

²¹ Instead of (7), a fixed-effect equation for the earnings of *futures* can be estimated with the optimism measure (and other controls) interacted with a self-employment dummy. This formulation yields a negative differential effect of optimism which is significant at the 5% level. The effect of optimism on those remaining in paid employment cannot be captured by this procedure, only the differential effect.

One potential criticism of the method is that, by construction, the optimism measure will tend to be negatively correlated with contemporaneous realizations and, to the extent that shocks are permanent, with future realizations and income. If expectations are not rational, $\hat{\theta}_i$ will act as a proxy for low underlying earning power and, therefore, be directly correlated with low earnings in the future. A negative association between $\hat{\theta}_i$ and E_{it} in equation (7) may simply reflect this effect rather than the influence of optimism on business start-up. To eliminate this possibility, when estimating the effect of optimism on earnings, two controls are included. The first, *pastrealizations_i*, is the fixed-effect extracted from a linear realizations equation estimated for the same sample and period as for the optimism fixed effect, $\hat{\theta}_i$. The estimated equation, which provide the optimism regressor, is reported in column 2 of Table A2 in the Appendix. This procedure eliminates the impact of the optimism effect on earnings simply arising due to extrapolation from past performance.²² Moreover, the categorical realization variable is not the only measure of past earnings power available. There is also the self-reported wage. In principle this ought to be a better measure of past individual labour market performance as it measures labour income rather than the individual's perception of their overall financial situation. So a second control, *pastearnings_i*, is also used to eliminate any spurious optimism-earnings association. This is the individual fixed effect extracted from an hourly earnings regression estimated over the same period as $\hat{\theta}_i$ and *pastrealizations_i*. These fixed effects serve as a proxy measure of intrinsic earnings ability. Table A3 reports the earnings equation which provides this regressor. The estimated coefficients on key education, occupation and other labour market and employment characteristics conform to those in the huge body of past work on the determinants of earnings.

²² The greater the extent to which past realizations are due to permanent shocks, the lower might expectations be. Thus lower optimism implies worse performance, contrary to the self-employment finding.

The effect of past optimism on future earnings is thus measured and compared for individuals controlling for past earnings history, closing off the poor performance channel as the explanation of optimism effects. That this procedure succeeds in removing these mechanical effects is indicated by the fact that the relationship between optimism and earnings, as discussed in the next section, is found to be positive for *nevers*. This is consistent with recorded optimism sometimes reflecting bad luck. Individuals may make rational forecasts but, by chance, realizations are low, and so they appear in our data as optimists. Because of mean reversion these individuals should do better in the future. For *futures* this effect may still be present but now the effect of intrinsic optimism on entry more than offsets the rational expectations effect.

Table A4 summarises the second-phase data used to model earnings. The self-employed earn significantly less than employees, although from Table A1, when still in paid employment *futures* earn significantly more than *nevers*. The self-employed are much more likely to be male, reflecting the lower proportion of women amongst the self-employed in the UK. The self-employed are less likely to hold university/college degrees than *nevers* but are more likely to have dependent children, to be home owners and married. Just over 18% of self-employed respondents report leaving compulsory schooling with no formal qualifications compared to 16% of the employed. Home ownership and wealth has also been found to be correlated with self-employment activity consistent with a “collateral channel” (Black et al., 1996; Adelino et al., 2013; Kerr et al., 2015; Jensen et al., 2015, Schmalz et al., 2017 amongst others). As noted, high levels of home ownership are reported by the self-employed, with approximately 69% reporting a mortgage debt on their property. Labour market experience is captured through the inclusion of an employment tenure variable. On average the self-employed have nearly 4 years of employment tenure and *nevers* 5.7 years. For the self-employed, prior experience may, however, be provided indirectly through parental role

models (Fairlie and Robb, 2007; Colombier and Masclet, 2008) as well as genetic factors (Lindquist et al., 2015). Parental business ownership experience is included as a control. Over three-quarters did not have a self-employed parent, with 22% reporting that one or both parents were self-employed. The self-employed also work just over nine hours longer per week than *nevers*.

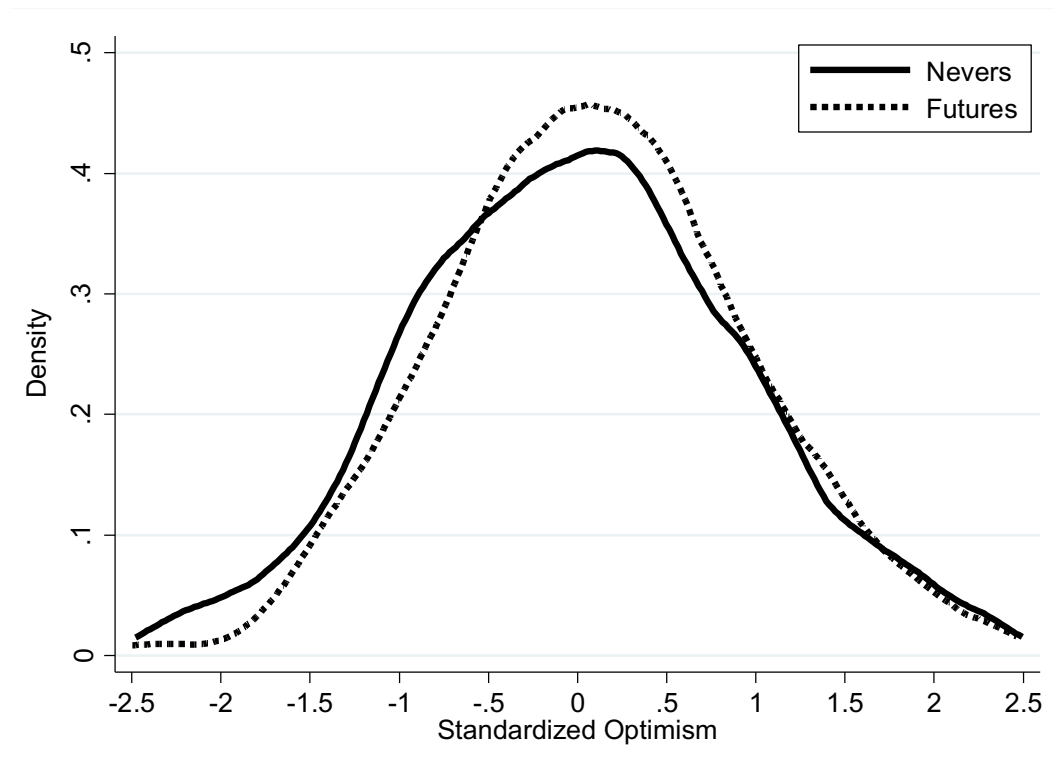
4. Results

4.1 Selection into self-employment

The first question is whether there is selection on optimism. Evidence in favour is that the respective mean standardized optimism scores for *futures* and *nevers* are 0.115 and -0.010, both measured whilst in paid employment, with the difference in means statistically significant at the 1% level.²³ Figure 2 displays the distributions of our measure of standardized optimism, \hat{o}_i , for *futures* and *nevers*. It shows that *futures* are significantly more optimistic at all points on the lower three-quarters of the distribution.

²³ The test of difference in means is bootstrapped to account for the fact that optimism (\hat{o}_i), is generated, rather than observed.

Figure 2: Distributions of standardized intrinsic optimism



Note: There is a single optimism score per individual yielding a sample of 618 individuals for *futures* and 7367 individuals for *nevers*.

A probit selection into self-employment equation for the combined sample of *nevers* and *futures*, is also estimated, using one observation per individual. This equation is reported in Table 1, where the dichotomous dependent variable takes on the value of one for individuals observed as *futures* and zero for individuals observed as *nevers*. Marginal effects are reported where characteristics are held constant at their respective mean values. Optimism is highly significant, consistent with previous research, notably Dawson et al. (2014). A one-point increase in the optimism measure is associated with a 1.3 percentage point increase in the likelihood of future self-employment. This implies a 21% increase in the probability of

future self-employment when evaluated as a one-point increase in optimism from the mean level of optimism in the sample.²⁴

Table 1: Optimism and selection into self-employment

Dependent Variable	1 if Future, 0 if Never
Estimator	Probit
Sample	Cross section
Variable	Marginal Effect (std. err.)
Standardized Optimism ($\hat{\theta}_i$)	0.013*** (0.004)
Past Earnings ($pastearnings_i$)	0.030*** (0.004)
Past Realizations ($pastrealizations_i$)	0.013** (0.004)
Year of Birth	0.00001 (0.0004)
Male	0.037*** (0.002)
Both parents self-employed	0.068*** (0.006)
Father self-employed	0.026*** (0.003)
Mother self-employed	0.022** (0.008)
Observations	7,609
Mean of Dependent variable	0.076

Note: The estimated standard errors in parenthesis are bootstrapped to account for the fact that $\hat{\theta}_i$, $pastearnings_i$ and $pastrealizations_i$ are generated, rather than observed. A bootstrap procedure is used involving 500 repetitions which draws bootstrap samples (random samples with replacement) and puts them through the multiple stage-procedure. * indicates significance level (p-value) below 0.10, ** below 0.05 and *** below 0.01.

²⁴ For robustness we also run the probit regression presented in Table 1 using all the available individual observations for our sample of *futures* and *nevers*. This equation includes the time invariant controls reported in Table 1 and a further barrage of time-varying controls. This procedure yields a marginal effect of approximately 2 percentage points on our optimism variable. This equates with a 24% increase in the probability of future self-employment when evaluated as a one-point increase in optimism from the mean level of optimism in the sample. This effect is statistically significant at the 5% level, where the standard errors are clustered and bootstrapped to account for the panel nature of the data and the generated regressors. In addition, using OLS also does not alter the conclusions drawn from Table 1.

4.2 Earnings and optimism

Table 2 reports the key findings, namely those for the earnings model set out in equation (7). The first column of Table 2 reports results for the self-employed (those previously *futures*) using the Tobit estimator because of the earnings left-censoring issue noted earlier. In the second column the earnings model for *nevers* in the second half of their panel presence is reported. This is estimated by OLS with earnings in levels. Paid employment earnings functions are usually estimated in semi-log form to allow for distributional skewness. However, the purpose here is to provide meaningful comparison with the self-employed. In the third column the differential effect of optimism on self-employment and paid-employment earnings is formally investigated. Specifically, the results from a Tobit estimator that pools the self-employed and employee sub-samples are reported, imposing the restriction of a common coefficient on each of the control variables across the two sub-groups.

The effect of optimism on earnings is significantly positive for employees, negative for the self-employed and the difference between the groups is highly significant statistically. Importantly, the inclusion of the prior earnings control (the individual fixed-effect from the prior paid-employment earnings equation, in Table A3) means that the coefficients on time-invariant intrinsic optimism in Table 2 measure differential effects of the variable in paid and self-employment, or, in the case of *nevers*, early versus later career effects. Specifically, if optimism affects earnings as an employee, this effect is captured by the inclusion of first-stage earnings fixed effects. The self-employment optimism coefficient in Table 2 therefore measures how optimism as an employee boosts or limits self-employment earnings. The effects are not small. For *nevers*, a one-point increase in standardized optimism is associated with an increase in monthly earnings of £48 (An increase of 2.8% when evaluated at mean employee earnings). For *futures*, a one-point increase in optimism is associated with reduced

monthly self-employed earnings of £283 (A reduction of 20.5% when evaluated at mean self-employed earnings). Moreover, the self-employment earnings of those with optimism below the mean exceed those with above average optimism by 32%, controlling for other characteristics. The results from the pooled Tobit regression are also in line with these conclusions. A one-point increase in the standardized optimism score lowers (increases) earnings by £239.37 (£37.84) for the self-employed (employees), with the difference being statistically significant at the 1% level.²⁵

Table 2: Optimism and earnings

Dependent Variable	Gross Monthly Earnings (in £'s)		
	(1) Tobit	(2) OLS	(3) Tobit
Estimator	Self-Employed	Employee	Pooled
Sample	Coef.	Coef.	Coef.
Variable	(std. err.)	(std. err.)	(std. err.)
Standardized Optimism ($\hat{\theta}_i$)	-282.50** (143.40)	48.10** (22.56)	-239.37*** (88.67)
Employee			743.80*** (80.82)
Employee*Optimism			277.20*** (89.03)
Observations	1,964	25,537	27,501
Mean of Dependent Variable	£1381.5	£1733.9	£1708.8

Note: The estimated standard errors in parenthesis are clustered and bootstrapped to account for the panel nature of the data and the fact that $\hat{\theta}_i$, $pastearnings_i$ and $pastrealizations_i$ are generated, rather than observed. A bootstrap procedure is used involving 500 repetitions which draws bootstrap samples from the clusters (random samples with replacement) and puts them through the multiple stage-procedure. * indicates significance level (p-value) below 0.10, ** below 0.05 and *** below 0.01. The regressions also include additional controls for whether the respondent draws up profit and loss accounts (column 1 only), prior performance controls, age in quadratic form, gender, marital status and household composition, health, educational attainment, housing tenure, parental background in self-employment, hours worked per week, employment tenure and whether the respondent is holding a second job. The regressions all include a series of one-digit industry dummy variables, and a set of year and region of residence dummy variables. Columns (1) and (3) have 137 left censored observations at gross monthly earnings ≤ 0 . From column (3) the

²⁵ For robustness, the analysis in Table 2 is redone with the generated variables, standardized optimism, past earnings and past realizations (i.e. the generated regressors) replaced by the respective raw individual time-averaged financial forecast error, log hourly real wage and financial realization over the relevant period. Results are similar to those reported in Table 2.

effect of optimism on earnings for employees is £37.84, which is statistically significant at approximately the 10% level. Full estimates are reported in Table A5 of the Appendix.

These findings are consistent with expectations involving both rational and psychological elements. To the extent that the optimism measure captures temporary negative income shocks, it will be associated with improved subsequent performance.²⁶ This is likely to be the main effect in paid employment. To the extent that measured optimism reflects systematic psychological bias, entry errors arise, imparting a negative relationship between optimism and self-employment earnings.²⁷ The self-employment finding might also reflect optimists being relatively less successful at running businesses compared to experience as employees. As optimists potentially take their operating decisions based on false information, this could lead to lower earnings. If this is the case, then realists should do best. Earnings would then not be monotonic in optimism. Unreported results, which estimate self-employment earnings with the inclusion of quadratic and cubic optimism effects, fail to find significant higher order terms in optimism. This suggests that optimism does not affect operating performance.

As Table A5 looks at second-phase earnings controlling for first-phase earnings, time invariant characteristics are only significant for employees to the extent that their impact changes through time or first-period temporary shocks are important. In the case of the self-employed, effects are also possible because different characteristics matter in the two employment modes. It is notable that relative to initial earnings, men earn more than women but the difference is greatest for the self-employed. Taken in conjunction with the Table 1 result that men are more likely to be self-employed, the gender gap arguably suggests that

²⁶ Possibly the optimists work harder, overestimating the earnings effect

²⁷ Including graduate/optimism interaction does not yield significant coefficients suggesting optimism effects are not restricted to the unsophisticated.

men may have a comparative advantage in self-employment. The association between earnings and university/college education, are much stronger for employees than for the self-employed. This is similar to previous research (see Le, 1999; Parker, 2009). Although some previous research has observed parental background effects, there is no particularly significant association in these data between self-employment earnings and family background in self-employment. Table 1 does find that parents self-employed increases the probability of self-employment so this may reflect inheritance rather than comparative advantage. Holding a second job reduces earnings significantly for both groups, and in quantitative terms particularly for the self-employed. Business start-up tends to be a time-consuming activity, and although holding a second job will provide some degree of insurance against failure it will also reduce earnings capacity.

5. Optimism, Divorce and Smoking

If the financial optimism measure captures an innate psychological trait, then it should be correlated with outcomes beyond the narrowly economic. As a test of validity, results are provided for a context involving rather similar issues - the relationship between optimism marriage and divorce. Viewed from the perspective of search theory, marriage has something in common with entry into self-employment (Shimer and Smith, 2000). The issue is to decide when a sufficiently good prospect has arrived. The optimism perspective is captured by the adage “marry in haste, repent at leisure”. Optimists may overestimate match quality, eventually realise that the marriage is a mistake, and are therefore more likely to divorce.²⁸ To test this, we compare the optimism - measured as the five-point difference between forecast and realization - of those who are currently married and who never divorce within the period covered by the dataset (*non-divorcees*) with those who are currently married but

²⁸ Optimists might overrate their own attractiveness and therefore wait longer to get married. Nevertheless, matches based on one party overestimating their worth are also likely to be unsatisfactory and therefore more likely to terminate.

divorce later in the panel (*future-divorcees*). We also include a series of dummy variables to capture the year of divorce, divorced status, year of remarriage (should this occur) and for remarried status. The results of this optimism equation are reported in the first column of Table A6. Married individuals who will divorce in the future are more optimistic than the married who never divorce. It could be argued that the difference is due to unlucky negative income shocks triggering divorce rather than intrinsic optimism. To counter this, two further equations are estimated. One examines whether these two groups differ in their financial forecasts and a parallel equation examines whether they differ in terms of their financial realizations. The respective results are presented in columns 2 and 3 of Table A6. The two-equation procedure makes it possible to reject the negative shock interpretation. Specifically, *future-divorcees* have significantly higher expectations, so an optimism effect cannot just be the result of income collapse. The final element is that *future-divorcees* have slightly higher financial realizations than *non-divorcees*, but the difference is not statistically significant. So it can be concluded that prior intrinsic optimism is associated with divorce.

As a further validity check, these equations also reveal that smokers have very significantly higher financial optimism. Although increased smoking is associated with lower financial realizations, at a marginal level of significance, optimism is not just the result of low income. Heavy smokers also have significantly higher financial expectations than those who do not smoke, given the same observables. Financial optimists tend to assume the worst will not happen. This psychology appears to transfer to the consequences of smoking suggesting that the optimism measure does at least partially capture a psychological trait.

6. Conclusion

This paper tests whether the equilibrium relationships implied by selection into self-employment on optimism hold in the earnings data, in particular, that intrinsic optimism leads

to mistaken entry into self-employment. Higher intrinsic optimism is therefore associated with lower self-employment earnings. Measured optimism is a compound of psychological bias and bad luck. In the latter case, optimism in the past implies higher future income. A negative relation between self-employment earnings and optimism therefore indicates the presence of a selection effect. For those continuing in paid employment, intrinsic optimism should not enter the second phase earnings function if past earnings are controlled for. The reason is that intrinsic optimism should have similar effects in all periods. This just leaves the luck effect, which imparts a positive optimism effect. The difference in the optimism coefficients between the two earnings regressions is good evidence that selection on optimism influences self-employment earnings.

The negative correlation between optimism and self-employment earnings could be a treatment rather than a selection effect. Optimism may directly affect performance in ways that are absent in salaried employment. Perhaps the self-employed have more discretion than employees and therefore it is more important that their decisions are based on a realistic appraisal of alternatives. An implication of optimism lowering productivity is that realists would do best, but self-employment earnings are monotonically decreasing in optimism. Even if optimism has a treatment effect, optimists can hardly recognise this, so excessive entry is still implied.

As always, the patterns found could be the result of omitted variable bias. For example, apart from optimism, there are no other psychological controls in our earnings equation. It is possible that our optimism measure is acting as a proxy for something else. Two papers that experimentally examine the correlation between optimism/overconfidence and other psychological traits find little connection (Dean and Orteleva, 2016; Stango, Yoong, and Zinman, 2016). In principle, there could be some variable other than intrinsic

optimism that is correlated with forecast earnings and has opposite effects on second-phase paid and self-employment earnings but no candidate comes to mind.

Recent debate has been critical of entrepreneurship policy, arguing that it can be poorly designed and confused with active labour market policy (Shane, 2009). To the extent that optimism leads to entry by those whose comparative advantage is not in entrepreneurship, this paper complements these concerns.

References

- Adelino, M., A. Schoar and F. Severino. (2015) “House Prices, collateral and self-employment” *Journal of Financial Economics* 117 (2), 288–306.
- Adomdza, Gordon K. and Åstebro, Thomas B. and Yong, Kevyn, (2016) “Decision biases and entrepreneurial finance” *Small Business Economics* 47 (4), 819–834.
- Arabsheibani, G., de Meza, D., Maloney, J. and Pearson, B. (2000). “And a vision appeared unto them of a great profit: Evidence of self-deception among the self-employed.” *Economics Letters* 67 (1), 35-41.
- Åstebro, T., P. Braunerhjelm and A. Broström, (2013) “The returns to academic entrepreneurship,” *Industrial and Corporate Change*, 22 (1), 281-311.
- Åstebro, T, and J Chen (2014) “The entrepreneurial earnings puzzle: Mismeasurement or real?” *Journal of Business Venturing* 29 (1), 88–105.
- Åstebro, T. (2003) “The return to independent invention: Evidence of unrealistic optimism, risk seeking or skewness loving?” *Economic Journal* 113 (484), 226-239.
- Åstebro, T., S.A. Jeffrey and G.K. Adomdza (2007). “Inventor perseverance after being told to quit: The role of cognitive biases.” *Journal of Behavioral Decision Making* 20 (3), 253–272.
- Åstebro, T, Mata, J. and Santos-Pinto, L. (2015) “Skewness seeking: risk loving, optimism or overweighting of small probabilities? *Theory and Decision* 78 (2), 189-208.
- Bengtsson, O and D. Ekeblom (2014). “The Bright but Right View? A New Type of Evidence on Entrepreneurial Optimism”. IFN Working Paper No. 1008. Available at SSRN: <http://ssrn.com/abstract=2398954>
- British Household Panel Survey. 1991-2009. UK Data Service. <http://discover.ukdataservice.ac.uk/catalogue?sn=5151> (accessed 17/01/2015).
- Black, J.M., D. de Meza, D., and D. Jeffreys (1996). “House prices, the supply of collateral and the enterprise economy.” *Economic Journal* 106 (434), 60-75.
- Brunnermeier, M. and J. Parker (2005), “Optimal expectations”, *American Economic Review*, 95 (4), 1092-1118.
- Camerer, C. and D. Lovallo (1999). “Overconfidence and excess entry: an experimental approach.” *American Economic Review* 89 (1), 306-318.
- Cassar, G. (2010). “Are individuals entering self-employment overly-optimistic? An empirical test of plans and projections on nascent entrepreneur expectations.” *Strategic Management Journal* 31 (8), 822-840.
- Columbier, N. and D. Masclet (2008). “Intergenerational correlation in self-employment: some further evidence from French ECHP data.” *Small Business Economics* 30 (4), 423-437.

- Compte, O., and A. Postlewaite. (2004). "Confidence-enhanced performance". *American Economic Review* 94 (5), 1536-1557.
- Cooper, A.C., C.Y. Woo and W.C. Dunkelberg (1988). "Entrepreneurs' perceived chances for success." *Journal of Business Venturing* 3 (2), 97-108.
- Dawson, C., D. de Meza, A. Henley & R. Arabsheibani (2014). "Entrepreneurship: cause and consequence of financial optimism." *Journal of Economics and Management Strategy* 23 (4), 717-742.
- Das, J.W.M., and A. Van Soest (1997). "Expected and realized income changes: Evidence from the Dutch Socio-Economic Panel." *Journal of Economic Behavior and Organization* 32 (1), 137-154.
- Dean, Mark, and Pietro Ortoleva. 2015. "Is It All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena." mimeo
- de Meza, D. and C. Southey, (1996). "The borrower's curse: Optimism, finance and entrepreneurship." *The Economic Journal* 106 (435), 375-386.
- Fairlie, R.W. and A.M. Robb (2007). "Families, human capital and small business: evidence from the Characteristics of Business Owners Survey." *Industrial and Labor Relations Review* 60 (2), 225-24.
- Fang, H. and G. Moscarini G. (2005), "Morale hazard." *Journal of Monetary Economics*, 52 (4), 749-778.
- Gervais, S. and T. Odean (2001). "Learning to be overconfident." *The Review of Financial Studies* 89(1), 1-27.
- Hamilton, B.H. (2000). "Does entrepreneurship pay: An empirical analysis of the returns of self-employment." *The Journal of Political Economy* 108 (3), 604-631.
- Hamilton, B.H., N.W. Papageorge, and N. Pande (2014). "The Right Stuff? Personality and Entrepreneurship" mimeo, available at SSRN: <http://ssrn.com/abstract=2438944>
- Hurst, E., G.Li and B.W.Pugsley (2014). "Are household surveys like tax forms: evidence from income underreporting of the self-employed" *Review of Economics and Statistics*, 96 (1), 19-33
- Hurst, E. and A-M. Lusardi (2004). "Liquidity constraints, household wealth and entrepreneurship." *Journal of Political Economy* 112 (2), 319-347.
- Hurst, E. and B.W. Pugsley (2011). "What do small businesses do?" *Brookings Papers on Economic Activity* 2011 (2), 73-118.
- Hurst, E. and B.W. Pugsley (2011). "Wealth, tastes, and entrepreneurial choice" in *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges* (eds. John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar)
- Hvide, H.K. (2002). "Pragmatic beliefs and overconfidence." *Journal of Economic Behavior and Organization* 48 (1), 15-29.

- Hvide, Hans K., and Georgios A. Panos (2014). "Risk tolerance and entrepreneurship." *Journal of Financial Economics* 111 (1), 200-23.
- Jensen, T.L., S. Leth-Petersen and R. Nanda (2014). "Home equity finance and entrepreneurial performance—evidence from a mortgage reform." *Harvard Business School Working Paper*, No. 15-020, October 2014. (Revised September 2015.)
- Jovanovic, B. (1982). "Selection and the evolution of industry." *Econometrica*, 50 (3), 649-670.
- Kerr, W.R., R. Nanda and M. Rhodes-Kropf (2014). "Entrepreneurship as experimentation." *Journal of Economic Perspectives*, 28 (3), 25-48.
- Kerr, S.P., W.R. Kerr and R. Nanda (2015). "House money and entrepreneurship." *Harvard Business School Working Paper*, No. 15-069, February 2015 (Revised July 2015.)
- Landier, Augustin and David Thesmar (2009). "Financial contracting with optimistic entrepreneurs." *Review of Financial Studies* 22 (1), 117–150.
- Langer, E. J., and J. Roth (1975). "Heads I win tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task." *Journal of Personality and Social Psychology*, 32, 951-955.
- Le, A.T. (1999) "Empirical studies of self-employment." *Journal of Economic Surveys*, 13 (4) 381-416.
- Levine, R., and Y. Rubinstein (2017) "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?," *Quarterly Journal of Economics*, 132 (2), 963–1018
- Lindquist, M., J. Sol and C.M. van Praag (2015) "Why do entrepreneurial parents have entrepreneurial children?" *Journal of Labor Economics*, 33 (2) 269-296.
- Malmendier, U. and G. Tate (2008) "Who makes acquisitions? CEO overconfidence and the market's reaction." *Journal of Financial Economics* 89 (1), 20-43.
- Manove, M. (2000) "Entrepreneurs, optimism, and the competitive edge," Working Paper, Boston University
- Moskowitz, T. J. and A. Vissing-Jorgensen (2002) "The returns to entrepreneurial investment: A private equity premium puzzle?" *American Economic Review* 92 (4), 745-778.
- Parker, S. (2009) "The Economics of Entrepreneurship" *Cambridge University Press*..
- Puri, M. and D.T. Robinson (2007). "Optimism and economic choice." *Journal of Financial Economics* 86 (1) 71-99.
- Puri, M. and D.T. Robinson (2013). "The economic psychology of entrepreneurship and family business." *Journal of Economics and Management Strategy* 22 (2), 207–444.
- Schmalz, M. C., D. A. Sraer and D. Thesmar (2017) "Housing collateral and entrepreneurship." *Journal of Finance* . LXXII (1) 99-132.♦

Schwardmann, P. and J. van der Weele (2016) “Deception and self-deception” (2016). Tinbergen Institute Discussion Paper, TI 2016-012/I. Available at SSRN: <http://ssrn.com/abstract=2734736> or <http://dx.doi.org/10.2139/ssrn.2734736>

Shane, S. (2009) “Why encouraging more people to become entrepreneurs is bad public policy.” *Small Business Economics* 33 (2) 141-149.

Shimer, R. and L. Smith (2000) “Assortative matching and search.” *Econometrica* 68 (2) 343-369.

Souleles, N.S. (2004). “Expectations, heterogeneous forecast errors and consumption: micro evidence from the Michigan consumer sentiment surveys.” *Journal of Money, Credit and Banking* 36 (1), 39-72.

Stango, V., J. Yoong and J. Zinman (2017) “The Quest for Parsimony in Behavioral Economics: New Methods and Evidence on Three Fronts.” *NBER Working Paper No. 23057*

Trivers, R. L. (2000) “The elements of a scientific theory of self-deception.” *Annals of New York Academy of Sciences* 907, 114–192.

Van den Steen, E. (2004). “Rational overoptimism (and other biases).” *American Economic Review* 94 (4) 1141–51.

Weinstein N. (1980). “Unrealistic optimism about future life events.” *Journal of Personality and Social Psychology* 36 (5), 806–820.

APPENDIX:

Table A1: First-phase summary statistics by second-phase employment status

Variable	<i>Futures</i>		<i>Nevers</i>	
	Mean/ Frequency	<i>Std.</i> <i>Dev.</i>	Mean/ Frequenc y	<i>Std. Dev.</i>
<i>Financial Forecasting and Earnings:</i>				
Financial Forecast (t): (3-point scale. "Worse off" to "Better off". -1/0/+1)	0.324***	0.633	0.287	0.633
Financial Realization (t+1): (3-point scale. "Worse off" to "Better off". -1/0/+1)	0.155	0.764	0.160	0.757
Financial Realization (t): (3-point scale. "Worse off" to "Better off". -1/0/+1)	0.181	0.775	0.182	0.766
Forecast Error: (5-point scale. <i>Forecast (t) minus Realization (t+1)</i>)	0.169***	0.871	0.127	0.849
Gross monthly earnings (in £'s)	1552.79***	1209.42	1235.09	852.38
<i>Demographics:</i>				
Age (years)	35.52***	10.02	34.84	10.53
Male	0.634***		0.492	
<i>Marital Status and Household Composition:</i>				
Single, never married	0.206***		0.229	
Widowed/divorced/ separated	0.055*		0.063	
Married/cohabiting partner employed	0.607		0.611	
Married/cohabiting partner not employed	0.131***		0.097	
Number of dependent children in household	0.707	0.983	0.708	0.968
<i>Highest Educational Attainment:</i>				
University/college degree	0.175***		0.157	
HND/HNC - vocational college qualifications	0.088**		0.076	
A-level	0.268***		0.210	
O-levels/GCSEs	0.315***		0.382	
No qualifications	0.154***		0.175	
<i>Housing Tenure:</i>				
Outright owner	0.100		0.102	
Own with mortgage	0.736***		0.695	
Private sector rental	0.089		0.092	
Social sector rental	0.075***		0.111	
Observations	3138 (618 individuals)		28830 (7367 individuals)	

** Appendix **

Note: Not all variables are recorded for all individuals. Only individuals with at least some observations for all variables are included. The difference between the means/frequencies of *futures* and *nevers* is tested with * indicating significance level below 0.10, ** below 0.05 and *** below 0.01. Educational attainment is measured through a series of dummy variables indicating the highest level of attainment. These are: university or college degree level at undergraduate or postgraduate level; HND (Higher National Diplomas) and HNC (Higher National Certificates) which are work-related, or vocational, higher education qualifications; A-levels or equivalent (post-compulsory examinations taken at 18 as qualifying exams for college or university entrance); GCSE or O-levels (age 16 schooling attainment qualifications); and no formal qualifications.

Table A2: OLS fixed-effect optimism and realization equations

Dependent Variable Variable	Forecast Error		Realization $t+1$	
	Coef.	Std. err	Coef.	Std. err
<i>Demographics:</i>				
Age	-0.032	0.027	-0.005	0.022
Age ² /100	-0.002	0.015	0.056***	0.013
<i>Marital Status and Household Composition (Reference: Single, never married)</i>				
Widowed/divorced/separated	-0.058	0.055	0.055	0.047
Married/cohabiting-partner employed	0.001	0.034	0.010	0.030
Married/cohabiting-partner not employed	-0.021	0.042	0.064*	0.037
Number of dependent children in household	0.037***	0.014	0.009	0.012
<i>Highest Educational Attainment (Reference: No qualifications)</i>				
University/college degree	0.079	0.146	0.036	0.125
HND/HNC - vocational college qualification	-0.027	0.140	0.061	0.117
A-level	-0.091	0.112	0.073	0.101
O-levels/GCSEs	-0.027	0.114	0.058	0.104
<i>Housing Tenure (Reference: Social sector rental)</i>				
Outright owner	0.057	0.054	-0.131***	0.046
Own with mortgage	0.088*	0.043	-0.089**	0.037
Private sector rental	0.093*	0.048	-0.058	0.041
<i>Financial Realizations (t) (Reference: 'Worse off')</i>				
'Better off'	0.136***	0.017	-0.117***	0.014
'Same'	-0.001	0.016	-0.065***	0.013
Region Controls	Yes		Yes	
Year (survey wave) Controls	Yes		Yes	
Observations		31968		31968
		(7985 Individuals)		(7985 Individuals)
F Test		7.07***		6.28***

Note: All regressions are clustered by individual and include year and region of residence dummy variables (coefficients not reported). * indicates significance level (p-value) below 0.10, ** below 0.05 and *** below 0.01.

Table A3: OLS fixed-effect log hourly real wage equation

Dependent Variable	Log Hourly Real Wage	
Variable	Coef.	Std. err
<i>Demographics:</i>		
Age	0.071***	0.008
Age ² /100	-0.104***	0.005
<i>Marital Status and Household Composition</i> <i>(Reference: Single, never married)</i>		
Widowed/divorced/separated	0.046***	0.017
Married/cohabiting-partner employed	0.055***	0.011
Married/cohabiting-partner not employed	0.069***	0.013
Number of dependent children in household	-0.020***	0.005
<i>Health (Reference: Health-other)</i>		
Health-excellent	0.000	0.006
Health-good	0.002	0.005
<i>Highest Educational Attainment (Reference: No qualifications)</i>		
University/college degree	0.173***	0.050
HND/HNC - vocational college qualification	0.098**	0.047
A-level	0.071*	0.041
O-levels/GCSEs	-0.008	0.037
<i>Labour Market Characteristics:</i>		
Union covered, member	0.066***	0.010
Union covered, non-member	0.015*	0.008
Holding a second job	-0.022**	0.009
Job tenure	0.000	0.001
Job tenure ²	0.004	0.005
Manager / supervisor	0.042***	0.006
Promotion opportunities available	0.004	0.005
Pay includes bonus / profit share	0.031***	0.005
Employer provided pension available	0.070***	0.008
Pay includes annual rises	0.015***	0.005
Shift worker	0.017*	0.010
Seasonal/Agency Temping/Casual contract	-0.023	0.017
Fixed-term contact	-0.009	0.016
<i>Flexibility in Job Location (Reference: Work at employers' premises)</i>		
Work from home	0.129***	0.048
Other work location	0.012	0.010
Work needs travelling	0.022**	0.010
<i>Occupation (Reference: Other)</i>		
Managers & Administrators	0.116***	0.017
Professional	0.132***	0.018
Associate Professional & Technical	0.088***	0.017
Clerical & Secretarial	0.048***	0.017
Craft & Related	0.044**	0.017
Personal & Protective Service	-0.018	0.018
Sales	-0.004	0.019
Plant & Machine Operatives	0.036**	0.017

** Appendix **

<i>Employing Sector (Reference: Private Firm)</i>		
Civil Service	-0.010	0.020
Local Government	0.031*	0.017
Other Public	-0.006	0.015
Non-Profit	-0.002	0.024
<i>One-digit Level Industry (Reference: Agriculture & Fishing)</i>		
Mining & Quarrying	0.136***	0.049
Manufacturing	0.060*	0.031
Electricity, Gas & Water	0.088*	0.047
Construction	0.046	0.034
Wholesale & Retail Trade	-0.017	0.032
Hotels & Restaurants	-0.066*	0.034
Transport, Storage & Communication	0.016	0.034
Financial Intermediation	0.049	0.036
Real Estate & Business Activities	0.053*	0.032
Public Administration & Defence	0.059*	0.032
Education	0.036	0.036
Health & Social Work	-0.022	0.033
Social & Personal Services	-0.013	0.034
Private Households & Extra-Territorial Organizations	0.060	0.038
<i>Firm Size -Number of Co-workers (Reference: Over 500)</i>		
1-9	-0.070***	0.010
10-24	-0.048***	0.010
25-49	-0.039***	0.010
50-99	-0.021**	0.010
100-199	-0.016*	0.009
200-499	-0.005	0.007
Region Controls	Yes	
Year (survey wave) Controls	Yes	
Observations		36391
		(9718 Individuals)
F Test		47.87***

Note: The regression is clustered by individual and includes year and region of residence dummy variables (coefficients not reported). * indicates significance level (p-value) below 0.10, ** below 0.05 and *** below 0.01.

Table A4: Second-phase summary statistics by employment status.

Variable	Self-Employed		Employee	
	Mean/Median/ Frequency	Std. Dev.	Mean/Median/ Frequency	Std. Dev.
<i>Gross monthly earnings (in £'s):</i>				
Mean	1381.48***	2389.04	1733.92	1241.79
25 th percentile	291.67***		958.75	
50 th percentile	833.33***		1499.15	
75 th percentile	1583.33***		2208.33	
90 th percentile	3031.00		3097.38	
95 th percentile	4583.33***		3788.92	
<i>Whether Draws up Profit/Loss Accounts:</i>				
Draws up accounts	0.823		-	
Does not draw up accounts	0.089		-	
Not yet but will be	0.089		-	
<i>Health:</i>				
Health-excellent	0.267*		0.250	
Health-good	0.496		0.509	
Health-other	0.237		0.241	
<i>Demographics:</i>				
Age (years)	42.48	10.01	42.36	10.58
Male	0.686***		0.482	
<i>Marital Status and Household Composition:</i>				
Single, never married	0.100***		0.135	
Widowed/divorced/separated	0.060***		0.090	
Married/cohabiting, partner employed	0.690***		0.661	
Married/cohabiting, partner not employed	0.150***		0.113	
Number of dependent children in household	0.824***	1.063	0.622	0.916
<i>Highest Educational Attainment:</i>				
University/college degree	0.157***		0.181	
HND/HNC - vocational college qualification	0.087		0.078	
A-levels	0.261***		0.216	
O-levels/GCSEs	0.314***		0.363	
No qualifications	0.181**		0.163	
<i>Housing Tenure:</i>				
Outright owner	0.183		0.173	
Own with mortgage	0.693		0.678	
Private sector rental	0.075		0.068	
Social sector rental	0.048***		0.081	

** Appendix **

Parental Background at Age

14:

Both parents self-employed	0.034***		0.012
Father self-employed	0.159***		0.103
Mother self-employed	0.029**		0.021
Neither parent self-employed	0.778***		0.864

Labour Market

Characteristics:

Job tenure (years)	3.98***	4.64	5.68	6.56
Usual hours worked per week	43.73***	16.41	34.56	9.59
Holding a second job	0.095***		0.072	
Observations	1964		25537	
	(559 individuals)		(6057 individuals)	

Note: The table comprises individuals included in Table A1 with at least some observations for all characteristics shown. Differences in the medians/means/frequencies between the self-employed and employees are tested with * indicating significance level below 0.10, ** below 0.05 and *** below 0.01. Educational attainment is measured through a series of dummy variables indicating the highest level of attainment. These are: university or college degree level at undergraduate or postgraduate level; HND (Higher National Diplomas) and HNC (Higher National Certificates) which are work-related, or vocational, higher education qualifications; A-levels or equivalent (post-compulsory examinations taken at 18 as qualifying exams for college or university entrance); GCSE or O-levels (age 16 schooling attainment qualifications); and no formal qualifications. Self-employment earnings are computed from a monthly self-employment profit variable for self-employed respondents who draw up profit and loss accounts, and a monthly self-employed gross pay variable if a self-employed respondent does not draw up profit and loss accounts. Dummy variables are shown which indicate whether or not the respondent draws up profit and loss accounts, to allow for systematic differences in measurement error between the two response types in the earnings model. Specifically, respondents are asked: “*In the business are annual business accounts prepared for the Inland Revenue for tax purposes?*”, with the possible responses being “*Yes*”, “*No*” and “*Not yet but will be.*”

Table A5: Estimation of optimism on gross monthly self-employed and employee earnings

Dependent Variable	Gross Monthly Earnings (in £'s)		
	(1) Tobit	(2) OLS	(3) Tobit
Estimator			
Sample	Self-Employed	Employee	Pooled
Variable	Coef. (std. err.)	Coef. (std. err.)	Coef. (std. err.)
Standardized Optimism ($\hat{\delta}_i$)	-282.50** (143.40)	48.10** (22.56)	-239.37*** (88.67)
Employee			743.80*** (80.82)
Employee*Optimism			277.20*** (89.03)
Draws up accounts	-26.23 (164.57)		
Not yet but will be	416.20** (204.55)		
Past Earnings (<i>pastearnings_i</i>)	980.60*** (351.76)	1550*** (64.99)	1467*** (67.16)
Past Realizations (<i>pastrealizations_i</i>)	90.36 (173.40)	146.8*** (31.77)	137.0*** (34.95)
Health-excellent	56.31 (183.55)	76.68*** (23.48)	70.53*** (25.76)
Health-good	58.55 (114.25)	25.78* (15.10)	26.61 (16.42)
Age	85.28 (64.28)	115.6*** (14.66)	112.1*** (13.75)
Age ²	-1.276* (0.71)	-1.521*** (0.11)	-1.479*** (0.12)
Male	375.0** (177.89)	144.0*** (24.28)	198.3*** (25.09)
Widowed/divorced/separated	-572.30* (335.58)	96.29** (44.04)	68.24 (46.28)
Married/cohabiting-partner employed	-104.50 (305.48)	74.54*** (28.25)	52.85 (32.36)
Married/cohabiting-partner not employed	-16.76 (336.25)	233.6*** (48.06)	209.5*** (51.42)
Number of dependent children in household	-97.84 (95.23)	-12.96 (12.92)	-17.29 (15.43)
University/college degree	-134.30 (355.48)	617.5*** (83.97)	600.6*** (85.28)
HND/HNC - vocational college qualification	-92.88 (290.42)	245.2*** (85.55)	245.9*** (86.33)
A-Level	-464.1* (251.98)	93.20 (67.85)	71.41 (68.79)
O-levels/GCSEs	-211.4	-80.16	-80.67

** Appendix **

	(235.49)	(60.63)	(63.03)
Outright owner	298.2	-6.427	8.619
	(284.37)	(32.97)	(36.68)
Own with mortgage	127.1	7.972	29.56
	(229.28)	(24.59)	(25.83)
Private sector rental	-185.0	-71.74**	-68.34*
	(262.70)	(36.72)	(36.39)
Both parents self-employed	14.30	-25.62	60.34
	(421.86)	(122.44)	(121.01)
Father self-employed	55.81	22.33	22.01
	(198.57)	(41.85)	(49.19)
Mother self-employed	-593.1*	32.18	-26.05
	(398.80)	(88.81)	(83.03)
Usual hours worked per week	9.638**	37.64***	32.27***
	(4.36)	(1.13)	(1.25)
Job tenure	42.03**	-0.788	-0.0849
	(18.03)	(1.33)	(1.43)
Holds a second job	-315.3**	-89.83***	-124.4***
	(136.15)	(27.76)	(28.10)
One-digit Industry Controls	Yes	Yes	Yes
Region Controls	Yes	Yes	Yes
Year (survey wave) Controls	Yes	Yes	Yes
Observations	1,964	25,537	27,501
Wald Chi ²	644.57***	15886.52***	19981.08***
R ²		0.5953	

Note: The estimated standard errors in parenthesis are clustered and bootstrapped to account for the panel nature of the data and the fact that $\hat{\delta}_i$, $pastearnings_i$ and $pastrealizations_i$ are generated, rather than observed. A bootstrap procedure is used involving 500 repetitions which draws bootstrap samples from the clusters (random samples with replacement) and puts them through the multiple stage-procedure. * indicates significance level (p-value) below 0.10, ** below 0.05 and *** below 0.01. Columns (1) and (3) have 137 left censored observations at gross monthly earnings ≤ 0 . From column (3) the effect of optimism on earnings for employees is £37.84, which is statistically significant at approximately the 10% level.

Table A6: Ordered probit regression for forecast error, expectations and realizations

Dependent Variable	Forecast Error		Expectation t		Realization $t+1$	
Column	(1)		(2)		(3)	
Variable	Coef.	Std. err	Coef.	Std. err	Coef.	Std. err
<i>Divorce Status (Reference: Non-divorcees)</i>						
Future Divorcee	0.103***	0.033	0.158***	0.050	0.001	0.038
Year of divorce	0.141	0.155	-0.192	0.148	-0.341**	0.166
Divorced	0.087***	0.025	0.099***	0.037	-0.026	0.027
Year of Remarriage	-0.056	0.130	0.298*	0.157	0.325**	0.156
Re-Married	0.063	0.051	0.062	0.054	-0.031	0.047
Smoker (Number of Cigarettes)	0.005***	0.001	0.006***	0.001	-0.001*	0.001
<i>Demographics:</i>						
Age	-0.007	0.006	-0.042***	0.007	-0.028***	0.006
Age ² /100	0.000	0.000	0.000***	0.000	0.000***	0.000
Male	0.034**	0.013	0.040**	0.018	-0.009	0.014
<i>Employment Status (Reference: Employee)</i>						
Self-Employed	0.126***	0.022	0.207***	0.031	0.012	0.023
<i>Household Composition</i>						
Number of dependent children in household	0.022***	0.007	0.011	0.009	-0.019***	0.007
<i>Highest Educational Attainment (Reference: No qualifications)</i>						
	-					
University/college degree	0.125***	0.023	-0.040	0.032	0.120***	0.025
HND/HNC - vocational college qualification	-0.043	0.028	0.044	0.039	0.085***	0.030
A-level	-0.018	0.020	-0.008	0.029	0.012	0.022
O-levels/GCSEs	0.001	0.018	0.018	0.025	0.011	0.019
<i>Housing Tenure (Reference: Social sector rental)</i>						
	-					
Outright owner	0.086***	0.029	-0.207***	0.038	-0.058**	0.030
Own with mortgage	-0.035	0.025	-0.052	0.032	0.001	0.026
Private sector rental	-0.015	0.037	-0.018	0.048	0.004	0.039
<i>Financial Realizations (t) (Reference: 'Worse off')</i>						
	-					
'Better off'	0.337***	0.017	0.464***	0.023	0.790***	0.020
	-					
'Same'	0.301***	0.015	-0.009	0.020	0.365***	0.017
Region Controls	Yes		Yes		Yes	
Year (survey wave) Controls	Yes		Yes		Yes	

** Appendix **

Observations		41289 (5676 Individuals)	
Log Likelihood	-48832.9	-35064.0	-41274.2
chi ²	1099.4***	1937.5***	2823.8***
Pseudo R ²	0.014	0.044	0.046

Note: All regressions are clustered by individual and include year and region of residence dummy variables (coefficients not reported). * indicates significance level (p-value) below 0.10, ** below 0.05 and *** below 0.01.