UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2011

Estimating the Return to College in Britain Using Regression and Propensity Score Matching

Wen Fan, University College Dublin

WP11/19

September 2011

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Estimating the Return to College in Britain Using Regression and Propensity Score Matching*

Wen Fan

School of Economics and Geary Institute UCD

wen.fan@ucd.ie

Abstract: College graduates tend to earn more than non-graduates but it is difficult to ascertain how much of this empirical association between wages and college degree is due to the causal effect of a college degree and how much is due to unobserved factors that influence both wages and education (e.g. ability). In this paper, I use the 1970 British Cohort Study to examine the college premium for people who have a similar ability level by using a restricted sample of people who are all college eligible but some never attend. Compared to using the full sample, restricting the sample to college-eligible reduces the return to college significantly using both regression and propensity score matching (PSM) estimates. The finding suggests the importance of comparing individuals of similar ability levels when estimating the return to college.

JEL Classification: I21; J31

Keywords: return to college, regression, propensity score matching

* I acknowledge the Economic and Social Data Service (ESDS) for access to the 1970 British Cohort Study (BCS70). I also gratefully acknowledge financial support from the Irish Research Council for the Humanities and Social Sciences. I am grateful to Paul Devereux for constant advice. I also thank Olivier Bargain and Kevin Denny for helpful comments. The usual disclaimer applies.

1. Introduction

Economists have long been interested in estimating the return to college and students expect to obtain higher pay in their future career as a result of attending college. However, many researchers argue that the return to schooling is likely to be overstated because innate ability is positively correlated with educational attainment. Some papers illustrate that this “big” coefficient is largely due to returns to unobservable ability instead of education *per se* (Blackburn and Neumark, 1992, 1993; Juhn, Murphy and Pierce 1993; Grogger and Eide, 1995; Taber, 2001; Caponi and Plesca, 2009). If college really does increase skills, policy makers might consider policies aimed at subsidizing college especially if credit constraints are important or there are positive spillover effects. However, if college graduates earn more because they have higher ability, policies that focus on skill acquisition earlier in life may be more appropriate. In this paper, I attempt to identify the return to education from the return to other unobservable ability.

In practice, researchers have used three strategies to deal with unobserved ability: analysis based on twins or siblings data, instrumental variable (IV) methods, and adding controls for observed proxies for ability. Given the first method is usually limited by small sample size and potential endogenity of schooling differences between twins, and the fact that it is not easy to find a valid instrument in practice, I intend to adopt the third approach and point out several improvements in my study compared to previous work.

2. Relationship to the Literature

I. Education and Ability Test Scores

In empirical studies, IQ test scores are widely used to proxy ability and added into the wage equation, but if these scores are affected by individuals' education levels, the estimated ability parameter reflects the combined effect of ability and education rather than the pure ability effect. It is perhaps unsurprising that different studies have come to different conclusions. Murnane *et al.* (1995) find a large impact of cognitive skills on wages using US data. Hansen *et al.* (2004) argue that the estimated effect of latent cognitive ability on attending school has been overstated in many studies since researchers do not correct for reverse causality between schooling and test scores. Therefore, it is very difficult to distinguish ability effects from education effects by simply adding test scores, so it is important to use measures that are clearly determined at the time of starting college and so are definitely not results of college attendance.

In the UK, obtaining 2 A-levels is the minimum requirement for college entrance (Chevalier *et al.* 2004; Walker and Zhu 2008). As people with 2 or more A-levels are basically defined as college eligible, I am able to probe to which extent a college degree/diploma would benefit college attendees compared to their counterparts, i.e. people with similar ability (academic ability at least) who never attend a college. By virtue of conditioning on a list of covariates (say family background and ability components) which are measured before college attendance, I can obtain a more precise result by matching college attendees with other non-college (but also college eligible) as one might expect any difference between outcomes (for example, wages) from such matching is largely attributed to college attendance and degree achievement. Given my restricted sample strategy avoids the problem caused by potential bilateral causality between schooling and test scores; it should lead to a more reliable result relative to the previous literature.[1]

II. Wide Range of Control Variables

Most previous studies only use a few ability and other factors in their specifications; I utilise a large list of relevant control variables. These variables cover a large range of concerns from family incomes to subject grades, which have been proved to have strong explanatory power for people's education attainments in empirical studies. Belley and Lochner (2007) have documented that family income affects college attendance. Dearden (1999) estimates the effects of families and ability on men's education and earnings in Britain; Blundell *et al.* (2000), Sloane and O'Leary (2005), Walker and Zhu (2008) examine the return to a university education in Britain; Caponi and Plesca (2009) use a rich dataset to control for ability selection into higher education (HE) so as to investigate post-secondary education in Canada. These five papers are the most relevant studies to mine. But my work differs a lot in terms of data, methodology, and motivation.

In Dearden's (1999) paper, she uses NCDS data, which is quite similar to my BCS 70 with respect to collection and management methods. The control variables she chooses are school type in 1974, teachers' assessment, father's social class,

[1] In a similar spirit, Chevalier *et al.* (2004) employ test scores measured at age 7 in order to isolate the inherent ability (i.e. ability not affected by acquisition of schooling) from the education effect and they find little evidence that unobserved ability plays an important role in biasing the return to education.

parental education, along with some family backgrounds such as financial difficulty and number of siblings. The ability measure she uses is test scores at age 7. However, as it is 10 or more years from the age of 7 to the year students enter college, test scores at such a young age may be poor predictors of the decision to attend college. Unlike me, she does not use either O-levels or A-levels which should be more useful for measuring student’s ability before going to college. Additionally, she only focuses on male respondents whereas I consider the full population. Moreover, while she simply adds a variety of family and ability components to the regression, my study further creates the restricted sample and uses propensity score matching as well as regression to double check the relationship between education and earnings.

Blundell *et al.* (2000) use NCDS data and also use the matching method in their paper. Besides variables similar to those used in Dearden (1999), they use A-levels and restrict the sample to people with A-levels. But they do not include O-levels in their specifications. In addition, their matching approach is quite different from what I am doing, and their study only looks at people with A-levels so they cannot explore the exact extent to which the return college differs across the whole population and the population who have the prospect of undertaking college. This comparison would be important for people to make decisions on college attendance. My analysis generally provides a complementary job here.

Sloane and O’Leary (2005) explore how the college premium varies by degree and by gender. Walker and Zhu (2008) report estimates of the college wage premium using a college eligible sample and find no significant fall in the return to college over time (other than a little fall for men with low ability). However, both of these papers above use Labour Force Survey (LFS) and the paucity of individual and family controls means that they cannot attempt matching estimators to compare to OLS.

Although Caponi and Plesca (2009) invoke a variety of personal controls (e.g. education and working history, composition of current family and education of parents) to correct ability selection bias and find the returns to each level of education decrease as expected once accounting for such variables, they differ from my study by using Canadian data and focusing on explaining the wage differentials between community college and university. My analysis based on more standard college degrees or diplomas will tend to provide general results which could be more comparable to other studies since most datasets do not record whether qualifications are achieved from community college or university.

III. Using Matching to Estimate the Return to Education

Matching has become more widely used by economists in recent years but, while much used by researchers assessing the effect of school quality (Berg Dale and Krueger 2002; Brand and Halaby 2003; Galindo-Rueda and Vignoles, 2004), has not been widely used to estimate the return to education. Blundell *et al.* (2000) use a matching approach to estimate the return to higher education in their paper but they simply employ a regression-based linear matching, which differs a bit from the propensity score matching I use here. Using nearest neighbour matching, I find that the return to college is cut almost 40% once the sample is restricted to the college eligible. Other matching methods also produce a smaller coefficient relative to using the full sample.

IV. Testing Performance of Regression and Matching

Does propensity score matching outperform regression? Many studies say yes. For example, Caponi and Plesca (2009) found propensity score matching (PSM) to be the most reliable estimator and claimed that OLS would overstate the return to education. In this paper, I assess the performance of regression compared to that of matching using a conventional full sample as well as my unique restricted sample of college-eligible persons. I find a similar return to college using OLS and propensity score matching, suggesting that regression performs similarly to PSM if the covariates are well controlled. This finding is consistent with Angrist and Pischke (2009), where they document that once covariates are entered flexibly, regression can be seen as a type of propensity score weighting.

The remainder of the paper is organized as follows: Section 3 describes my methodology in detail; Section 4 presents the data, Section 5 illustrates empirical results, and the conclusions are in Section 6.

3. Methodology

I. Empirical Models

Regression Models

The following two OLS models are used to identify the causal effect of education on earnings. Eq. (1) presents $Y_i = f\ (D_{i,}\ X_i)$; Eq. (2) replaces X_i with the propensity score $p(X_i)$.

(1) $Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \varepsilon_i$

(2) $Y_i = \beta_0 + \beta_1 D_i + \beta_2 p(X_i) + \varepsilon_i$

where Y_i represents log weekly earnings; D_i denotes a higher education attendee dummy variable which equals 1 if individuals achieve a HE diploma, first degree, postgraduate certificate or postgraduate degree (including masters and Ph. D degrees), 0 otherwise. X_i indicates covariates. The propensity score $p(X_i)$ will be estimated first by modelling the probability of being a higher education attendee as a function of a set of conditioning variables. When stratification on the propensity score is adopted, regressions can be run within individual strata or the strata act as a factor included in regression, and this method is highly regarded (Winship and Ware, 1992). In some circumstances, controlling for the propensity score may do a better job than using all Xs (see Dehejia and Wahba (1999)). From a statistical perspective, asymptotically, it is always best (minimum variance) to control for all the Xs rather than p(x); but, in small samples it may be more efficient to reduce the number of covariates using the propensity score. In my case, there is little difference (in the estimates) between using p(x) and using X. Appendix Table C provides coefficient estimates from regressions that control for the propensity score and the standard errors account for the fact that the propensity scores are estimated.

Propensity Score Matching

My main interest is to examine the effect of attending college; Eq. (3) shows the difference between the actual average earnings of college graduates (Y_{i1}) and what they would have earned if they had not attended college (Y_{i0}).[2]

(3) $\Delta|_{D=1} = E(Y_{i1} \mid D_i = 1) - E(Y_{i0} \mid D_i = 1)$

However, researchers usually can only estimate Eq. (4):

(4) $E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] = E[Y_{1i} - Y_{0i} \mid D_i = 1] + \{E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]\}$

The last term on the right side is selection bias, which could be solved if we assume the statistical independence of D_i and Y_{i1}, Y_{i0}, i.e. $Y_{i1,}Y_{i0} \perp D_i$.

In reality, as college attendance could hardly be random, people usually adopt a Conditional Independence Assumption (CIA) which states the researchers can observe some variables X_i correlated with both ability and educational attainment, so that conditional on these variables assignment to college is random, say $Y_{i1,}Y_{i0} \perp D_i \mid X_i$. The counterfactual can now be easily obtained from the observed outcomes of the non-college individuals given $E(Y_{i0} \mid X_i, D_i = 0) = E(Y_{i0} \mid X_{i,} D_i = 1)$. Within the class of selection on observables X_i, I use the matching model which finds, for each treated individual who attends and completes college $D_i = 1$, a very similar (conditioning on X_i) control individual $D_i = 0$ who does not attend college, and then compare the earnings of these two very similar individuals. Therefore, after doing such matching, people might expect any difference in earnings would be due to the college effect and the closer these two groups match, the more precise estimates I could obtain. Averaging these effects across individuals who actually attend and complete college gives the treatment on the treated parameter:

[2] Although only Average Treatment effect on Treated (ATT) equations are shown here, the Average Treatment Effect (ATE) is also of interest to examine; because while ATT tells us how much the typical higher education attendee gained or lost as a consequence of going to college, ATE tells us a more general picture on how much the full population would gain or lose. Therefore, I consider both ATT and ATE as being useful treatment effects to report. In practice the estimates are generally similar in my case.

$$(5)\Delta|_{D=1} = E(Y_{i1} \mid X_i, D_i = 1) - E(Y_{i0} \mid X_i, D_i = 1) = E(Y_{i1} \mid X_i, D_i = 1) - E(Y_{i0} \mid X_i, D_i = 0) = E(Y_{1i} - Y_{0i} \mid X_i)$$

Propensity score matching methods are growing in popularity in empirical work. The rational behind this model is that if X_i satisfies the CIA, then rather than matching on the multi-dimensional vector X, matching can be performed instead on a scalar index $p(X_i)$, which is just the estimated propensity score in Eq. (2). A further requirement besides independence is common support. It rules out the phenomenon of perfect predictability of D given X, and ensures that persons with the same X values have a positive probability of being both participants and non participants. Success in passing balancing property tests in my study guarantees that every single participant can find a matched non-participant based on propensity scores. So my final matching specification in practice turns out:

$$(6)\Delta|_{D=1} = E(Y_{i1} \mid p(X_i), D_i = 1) - E(Y_{i0} \mid p(X_i), D_i = 1) = E(Y_{i1} \mid p(X_i), D_i = 1) - E(Y_{i0} \mid p(X_i), D_i = 0)$$

II. Restricted Sample Strategy

Given I know the exact number of A-levels obtained by individuals, restricting the sample to persons who have the required number of A-levels to attend college could enable me to look at people with a similar ability level. In other words, I can pick up college-eligible people and compare people who consequently attend college with those who fail to attend college. While they clearly don’t cover all dimensions of ability, A-levels are fairly adequate to cover students’ academic ability due to two facts. First, there is a well-developed applications system for matching students to courses and this ought to ensure that most students with two or more A-levels passes can find a place on some course at some institution (Walker and Zhu, 2008). Secondly, A-levels are still the primary route into higher education. Normally, if they wish to go to college, students are expected to try their best to obtain good grades in A-levels. One big advantage of A-level results is that they are obtained very shortly before university entry and, all else equal, more up-to-date information on ability is preferable. Therefore, A-levels are expected to be a more informative measure of people's ability than are tests that took place long time ago (e.g. ability tests at age 7).

With this restricted sample, in addition to running the OLS model, I employ PSM in which I try to match people's characteristics as closely as possible across the treatment group (people who actually attend and complete college) and the control group (people who do not attend college) and thus any difference of outcomes could be attributed to college attendance and completion.

III. Variables Selection

The literature suggests that two factors have a large impact on college attendance and completion: 1) different tastes for education affect individuals' decisions to attend college when they are eligible; 2) financial support for higher education may impact attendance of college, or completion of college. I select sex, ethnic group, students' O-levels grade dummy for three subjects (mathematics, English and English literature) which most students choose to study[3], family background information including parents' age left full time education, social class, ethnic groups, working hours per week, family income and family size (how many people sharing food) in my matching specification to estimate the propensity score[4].

[3] I value this variable 1 if students achieved grade A for any of these three main subjects, 0 otherwise.
[4] A-levels are also used to predict propensity score in the college eligible sample.

These variables are also used as covariates in regression and PSM estimates and further help me account for more factors relevant to college attendance other than A-levels.

The basic identification assumption of PSM is that, conditional on selected covariates, whether people choose to go to college should be random. Given so, it is useful to be clear about what particular variables are used here. Apart from A-levels which I have discussed before, I use O-level grades which are from exams typically taken at age 16. These grades illustrate students educational achievement history and should be correlated with individual's long-run taste for education as well as non-cogitative abilities (patience, responsibility and so on) which may also affect her/his academic progress. Many empirical papers have shown that parental and family background characteristics play important roles in children's education attainment via multiple channels. For example, from the intergenerational transmission perspective, parental characteristics are strong indictors not only of their children's cognitive ability but of some types of non-cognitive ability. Thus, even though I may fail to fully account for ability, the included parental controls may do a complementary job in helping to predict educational choices.

IV. Two Sets of Estimations

There are two sets of estimations applied in my paper based on different samples. I employ regressions and PSM methods in each set. In the first set, I include the full sample and exclude the number of A-levels variable from both regression and PSM. This approach is the regular method people use to estimate the return to college and has the problem that individuals included in my sample differ substantially in multiple dimensions (e.g. ability level).

So what would happen to the return to college if I am able to look at people basically on a similar ability level? Following this line, I restrict the sample to persons who are college-eligible (2+ A-levels) and also use the A-levels information in my regression and PSM estimates. I expect this strategy to provide more compelling estimates as the ability levels of college graduates and non-graduates are more similar in the restricted sample.

4. Data

My dataset is the 1970 British Cohort Study (BCS70). The BCS70 began when data were collected about the births and families of babies born in the UK in one particular week of 1970. The survey took place when respondents were aged 5 in 1975, aged 10 in 1980, aged 16 in 1986, aged 26 in 1996, aged 30 in 1999-2000, aged 34 in 2004-2005, and aged 38 in 2008-2009. With each successive attempt, the scope of BCS70 has broadened from a strictly medical focus at birth, to encompass physical and educational development at the age of 5, physical, educational and social development at the ages of 10 and 16, and physical, educational, social and economic development at 26 years and beyond. The aged 16 and 26 waves provide all information used in my paper. Participants from Northern Ireland (NI), who had been included in the first wave, were dropped from the study in all subsequent sweeps, which only included respondents from Great Britain. I also drop Scottish observations as their educational system is different to that in England and Wales. The earnings are on a weekly basis. Higher Education Attendees represent people who report they have ever obtained either a higher education diploma, or first degree, or other postgraduate certificates (including degrees). Parents' social class is measured by interval, where higher figure indicates lower class. Parents' ethnic group values 1 to be British, the smaller number means people are more likely to be natives. Family incomes are reported by interval, the higher value indicates more income. Table 1 reports the descriptive statistics for the full sample and college eligible sample. Table 2 presents the descriptive statistics for the identical sample as Table 1 by treatment and control groups, and also presents standardized percentage differences, defined as the mean difference between treatment and control groups as a percentage of the standard deviation $\left\{[100(\bar{x}_t - \bar{x}_c)]/\sqrt{[(s_t^2 + s_c^2)/2]}\right\}$, where $\bar{x}_t$ and $\bar{x}_c$ are the sample means in the treatment and control groups; s_t^2 and s_c^2 are the corresponding sample variances. Also presented are the variance ratios s_t^2 / s_c^2. In the full sample, while the standardized differences in student characteristics and some family background characteristics (parental working hours, family size, family income, etc.) between groups are fairly small (almost all below 10%), some parental variables (e.g. education, social class) and student's O-level achievements are relatively large (above 60%). In the college eligible subsample where people are basically at a similar ability level, the standardized differences are generally reduced for most cases except the number of A-levels (88%). As Rosenbaum and Rubin (1985) show, one virtue of matched sampling is that non-technical audiences often find that matching, when successful, is a persuasive method of adjusting for imbalances in observed covariates. In my case, given the control group in my college subsample is relatively small; it is desirable to figure out to what extent the level of standardized differences is acceptable. To this purpose, I first present the estimation of the propensity score (i.e. probit model) for each sample in Appendix Table A and then report the comparison of covariate imbalance for variables with substantial initial bias (standardized absolute bias greater than 20%[5]) before and after Nearest Neighbour matching[6] in Appendix Table B. It is clear to see that matching does a very good job here --- all standardized differences vanish and all two-sample t statistics turn out insignificant after matching for the full sample. For the college eligible sample, all the standardized differences are reduced to less than 20%, which suggests that the treatment and control group characteristics are similar enough here.

5. Empirical Results

I. Propensity Score Estimation

Table 1 lists the variables used to estimate the propensity score (except, of course, earnings and HEA are not used in estimating the propensity score). The difference between estimating the propensity score for the full sample and college eligible sample is I include number of A-levels obtained for the latter sample. The trimming rule used in my paper is based on the regions of common support which are [.05021118, .99999981] and [.30688258, .99999992] for full sample and college

[5] Rosenbaum and Rubin (1985) consider the initial standard difference beyond 20% could be disturbingly large; in the meanwhile, the reduction in bias appears to be quite unstable when the initial bias is small.

[6] Standardized differences in covariates vary quite slightly across matching methods in my case.

eligible sample, respectively. Any case off common support, that is, any treatment observations whose propensity score is higher than the maximum or less than the minimum propensity score of the controls are dropped from the sample.[7] Figures 1 and 2 depict the distribution of estimated propensity scores on common support for full and college eligible samples.

Given these pre-treatment variables are fully controlled; under the assumptions, one would expect any difference in earnings between treatment and control groups is due to the effect of college attendance and degree/diploma achievement.

The details of estimation are offered in Table 3. The blocks indicate the unit in which the average propensity score between treated and controls are equal, and the average for each explanatory variable are also equal. The average treatment effect on treated (ATT) of interest is obtained as an average of the ATT of each block with weights given by the distribution of treated units across blocks.

<u>II. Parameters of Interest</u>

Tables 4 and 5 offer the result of regressions and PSM for the full and college eligible sample, respectively. Matching methods include nearest neighbour, kernel and stratification. There is a different rationale behind each of the three matching estimators. In nearest-neighbour, all treated units find a match based on the closest propensity score; in kernel matching (bandwidth is .06[8]), all treated are matched with a weighted average of all controls with weights that are inversely proportional to the

[7] There are 7 and 0 cases off common support for full and college eligible samples. In order to form fully comparison between regression and matching, the exactly same observations are used in empirical study.
[8] I have considered other bandwidth values other than .06; my estimates are not sensitive to the choice.

distance between the propensity scores of treated and controls; in stratification matching, a set of interval (or strata) are used to divide the common support of the propensity score, then treated and control cases are matched within each interval/strata. It is helpful to look at the returns to college with different matching methods. The coefficients appear to be quite similar across methods, suggesting ability controls should be more relevant than method choice in this case.

Abadie and Imbens (2008) find that the bootstrap is not valid for nearest neighbour matching with replacement. Thus, I employ the subsampling method (also suggested by Abadie and Imbens (2008)) instead of bootstrapping for nearest neighbour matching. This involves taking subsamples that are much smaller than the original sample size; therefore I randomly draw a 20% subsample respectively from the full sample and the college eligible subsample without replacement and perform 800 replications. In Table 5 we can see the result obtained from subsampling is not significantly different from those of other two matching estimators with conventional bootstrapping.

It is clear that the return to college falls considerably as the sample is restricted to college eligible individuals regardless of whether regression or PSM is implemented. The coefficient of interest reduces by almost 1/3 for the college eligible sample relative to the full sample in naive OLS regressions – the return to college falls from 64% in the full sample to 46% in the restricted college-eligible sample. The PSM results support this point as the return to college shrinks in the college-eligible sample across different matching methods. This finding implies that if one only looks at the full population, the return to college will be overestimated no matter which method is applied.

As both ATT and Average Treatment Effect (ATE) are interesting to look at, I report ATT and ATE for nearest neighbour and Kernel, and ATT for stratification matching in Table 5. As shown in Table 5, it is clear to see that both nearest neighbour and Kernel matching produce quite similar ATT and ATE in each sample. The results suggest that the ATT and ATE tend to look very similar when considerable ability noise has been removed.

Theoretically, in order to make regression more comparable to matching, one wants to fully saturate in X and include interactions between X and the treatment D. Angrist and Pischke (2009) argue that, in practice, OLS and matching may provide similar estimates even with a much more parsimonious OLS specification. The estimates from my study are consistent with this argument. With rich controls on individual and family background, OLS regressions in which covariates are included fairly flexibly (even without saturating in X) generate similar results to those from PSM.

While the smaller coefficient found in the college-eligible sample suggests the return to college is over estimated in the full sample, people might argue that this decrease could be driven by parameter heterogeneity as the distributions of covariates are different in the two samples. Unfortunately, by definition of the samples, it is impossible to make the two samples comparable; however the similarity of the mean value of covariates in both, to some extent, rules out the possibility that heterogeneity is a main driver of differences in the estimates. For example, fractions of female and British are exactly the same in the two samples; the family income index is 5.93 for the full sample and 5.92 for the college-eligible sample; family sizes are 4.02 for the full and 4.11 for eligible samples and so on; clearly, the distribution of covariates does not differ a lot across two samples, which corroborates my conclusion that the returns to college in the full sample are over-estimated due to the lack of ability controls.

6. Conclusions

From my empirical study, the return to college is remarkable but not as high as people usually expect. The coefficient of interest appears to be smaller once one restricts the sample to college-eligible individuals. I find the return to college would be over-estimated by about 40% in regressions and using nearest-neighbour matching if the full sample is used.[9] The similar results from regressions and PSM also suggest that regression models can be seen as a type of propensity-score weighting when covariates are included fairly flexibly, so the model is close to saturated. Overall, there does not seem to be a great advantage in using PSM rather than regression when estimating the return to college.

[9] Interestingly, these findings are quite consistent with those of Devereux and Fan (2011) who estimate the return to education using the 1990s expansion in higher education as an instrument.

Table 1 Descriptive Statistics

	Full Sample Mean (SD)		College Eligible Subsample Mean(SD)	
Log (weekly wage)	6.51	(1.33)	6.95	(1.35)
Higher Education Attendee (HEA)	0.34	(0.47)	0.79	(0.40)
Number of A-levels	0.91	(1.41)	3.01	(0.77)
Female	0.55	(0.50)	0.55	(0.50)
British	0.98	(0.13)	0.98	(0.15)
Father's social class	3.11	(1.23)	2.58	(1.23)
Mother's social class	3.59	(1.41)	3.19	(1.36)
Father's age left full time education	16.35	(2.76)	17.60	(3.56)
Mother's age left full time education	16.18	(2.16)	17.02	(2.75)
Father's hours worked per week	45.00	(11.31)	44.11	(10.81)
Mother's hours worked per week	27.77	(15.25)	28.71	(15.60)
Father's ethnic group	1.10	(0.68)	1.16	(0.91)
Mother's ethnic group	1.09	(0.62)	1.13	(0.74)
Grade A dummy of Any Main Subjects of O levels	0.19	(0.40)	0.47	(0.50)
Family income	5.93	(3.58)	5.92	(3.64)
Family size (number of people sharing the food)	4.02	(1.52)	4.11	(1.53)
Sample size	1,443		413	

Notes: HEA is as defined in Methodology section. Parents' social class is measured by interval, where the higher figure means lower class. Parents' ethnic group values 1 be British, the smaller number means people are more likely to be native. Family incomes are reported by interval, the higher value indicates more income.

Table 2 Descriptive Statistics (Treatment and Control Groups)

Panel A FULL SAMPLE

	Treatment Mean (SD)		Control Mean(SD)		Initial Standardized Difference (%)	Variance Ratio
Female	0.53	(0.50)	0.56	(0.50)	-6	1.0
British	0.98	(0.15)	0.99	(0.11)	-9	1.9
Father's social class	2.58	(1.23)	3.36	(1.11)	-67	1.2
Mother's social class	3.19	(1.35)	3.77	(1.38)	-42	1.0
Father's age left full time education	17.63	(3.81)	15.69	(1.66)	66	5.3
Mother's age left full time education	17.06	(2.85)	15.72	(1.52)	58	3.5
Father's hours worked per week	44.67	(10.75)	45.14	(11.60)	-4	0.9
Mother's hours worked per week	27.87	(14.21)	27.44	(15.29)	3	0.9
Father's ethnic group	1.14	(0.82)	1.08	(0.61)	8	1.8
Mother's ethnic group	1.14	(0.77)	1.06	(0.53)	12	2.1
Grade A dummy of Any Main Subjects of O levels	0.42	(0.49)	0.08	(0.27)	87	3.3
Family income	5.68	(3.42)	6.04	(3.64)	-10	0.9
Family size (number of people sharing the food)	4.06	(1.55)	4	(1.51)	4	1.1
Sample size	493		943			

Panel B COLLEGE ELIGIBLE SUBSAMPLE

	Treatment Mean (SD)		Control Mean(SD)		Initial Standardized Difference (%)	Variance Ratio
Number of A-levels	3.14	(0.74)	2.51	(0.68)	88	1.2
Female	0.55	(0.50)	0.59	(0.50)	-9	1.0
British	0.97	(0.16)	0.99	(0.11)	-11	2.1
Father's social class	2.47	(1.22)	3.01	(1.16)	-46	1.1
Mother's social class	3.15	(1.36)	3.38	(1.38)	-17	1.0
Father's age left full time education	18	(3.79)	16.05	(1.77)	66	4.6
Mother's age left full time education	17.24	(2.92)	16.18	(1.75)	44	2.8
Father's hours worked per week	44.29	(10.56)	43.41	(11.78)	8	0.8
Mother's hours worked per week	28.19	(15.04)	30.75	(17.54)	-16	0.7
Father's ethnic group	1.18	(0.92)	1.11	(0.88)	7	1.1
Mother's ethnic group	1.16	(0.82)	1.03	(0.24)	20	11.7
Grade A dummy of Any Main Subjects of O levels	0.52	(0.50)	0.26	(0.44)	56	1.3
Family income	5.84	(3.55)	6.19	(3.98)	-9	0.8
Family size (number of people sharing the food)	4.10	(1.54)	4.13	(1.52)	-2	1.0
Sample size	328		85			

Notes: Parents' social class is measured by interval, where the higher figure means lower class. Parents' ethnic group values 1 be British, the smaller number means people are more likely to be native. Family incomes are reported by interval, the higher value indicates more income. The initial standardized difference is computed as $\left\{[100(\bar{x}_t - \bar{x}_c)]/\sqrt{[(s_t^2 + s_c^2)/2]}\right\}$, where $\bar{x}_t$ and $\bar{x}_c$ are the sample means in the treatment and control groups; s_t^2 and s_c^2 are the corresponding sample variances. Variance ratios is calculated as s_t^2 / s_c^2.

Figure 1

Distribution of Estimated Propensity Scores (Full Sample)

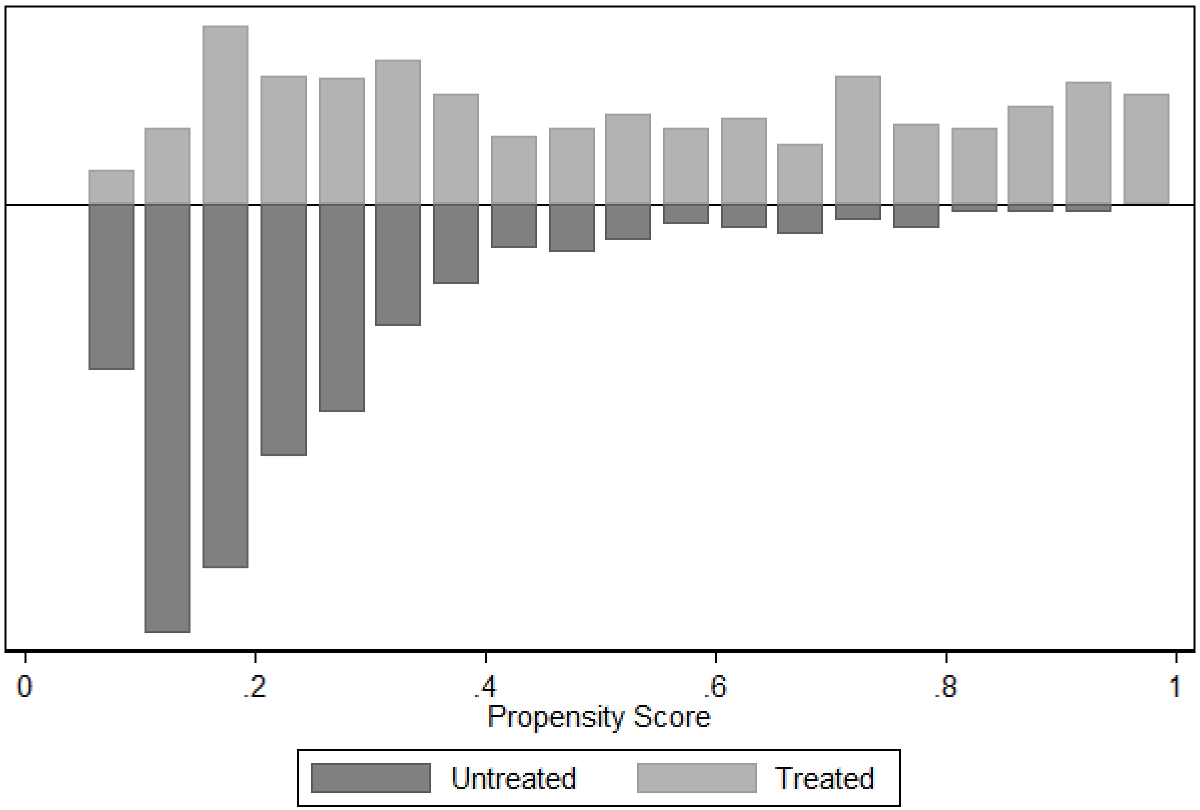

Figure 2

Distribution of Estimated Propensity Scores (College Eligible Subsample)

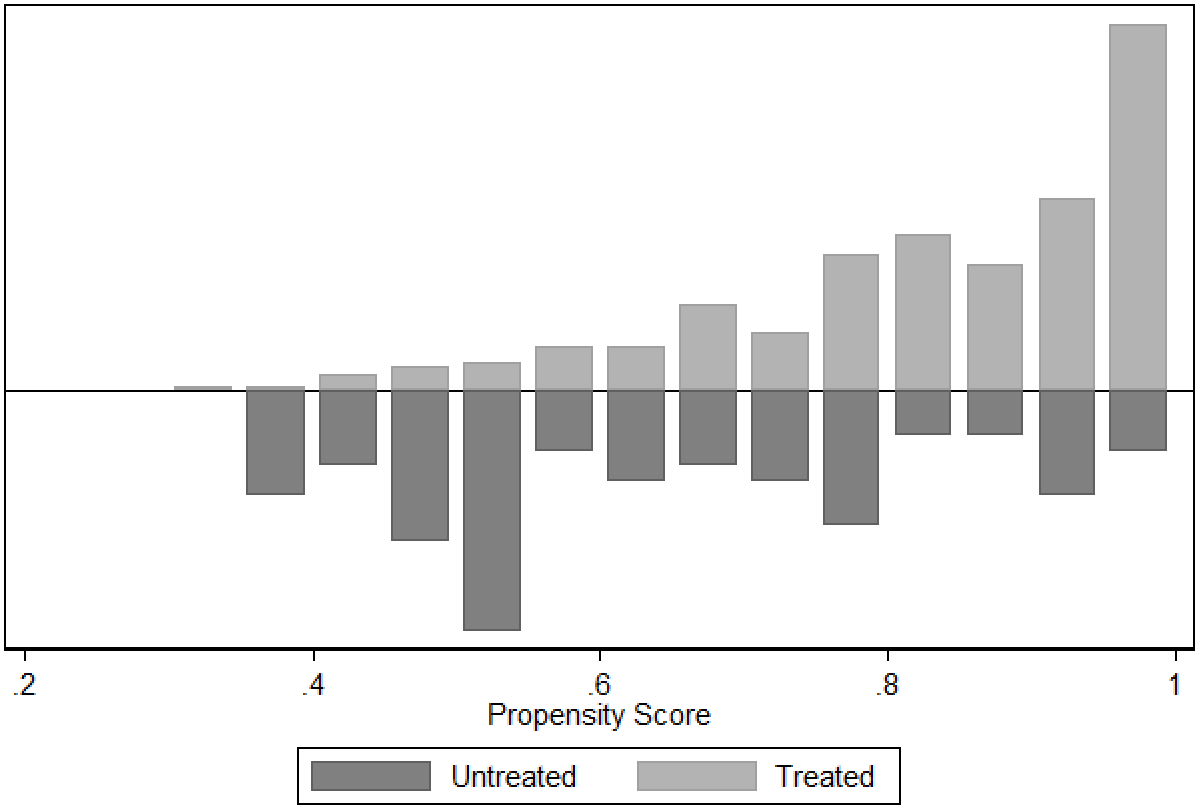

Table 3

Description of the Estimated Propensity Score

	Full Sample		College Eligible Subsample	
Means	0.34		0.79	
Standard Deviation	0.25		0.17	
Number of Blocks	8		5	
Observations per Block	**Treated**	**Control**	**Treated**	**Control**
	9	83	0	0
	20	214	2	7
	47	182	28	35
	67	230	83	26
	67	101	215	17
	82	74		
	94	47		
	107	12		
Number of Observations	493	943	328	85

Notes: sex, ethnic group, parents' age left full time education, parents' social class, parents' ethnic group, parents' working hours per week, students' O-level grade dummy for any main subjects, family income and family size are used to estimate propensity score; for college eligible sample, number of A-levels is included as well.

Table 4

Regression Estimates of the Return to College

Y=Log (weekly wage)	[1] Full Sample	[2] College Eligible Subsample
OLS (Y = *f*(D, X))	0.643***	0.458***
	(0.082)	(0.175)
	[N=1,436]	[N=413]

Notes: Standard Errors in Parentheses. All specifications include sex, ethnic group, parents' age left full time education, parents' social class, parents' ethnic group, parents' working hours per week, students' O-level grade dummy for any main subjects, family income and family size. Column 2 also includes number of A-levels in specification. Standard errors are reported in round brackets whereas sample sizes are reported in squared brackets.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5 Propensity Score Matching Estimates on Return to College

Full Sample Y=Log (weekly wage)			
Matching Method	**N**	**ATT**	**ATE[b]**
Nearest Neighbour		0.778*** (0.148)[a]	0.788*** (0.113)
Kernel	1436	0.670*** (0.116)	0.678*** (0.097)
Stratification	1436	0.665*** (0.103)	---

College Eligible Subsample Y=Log (weekly wage)			
Matching Method	**N**	**ATT**	**ATE[b]**
Nearest Neighbour		0.529** (0.266)[a]	0.407** (0.186)
Kernel	413	0.551*** (0.158)	0.514*** (0.144)
Stratification	413	0.561*** (0.143)	---

a: using subsampling method with 20% draws from the original sample.
b: Sample size of ATE is 1,436 for full sample and 413 for the college-eligible subsample.
Notes: The treatment is Higher Education Attendee (HEA) by definition. Sex, ethnic group, parents' age left full time education, parents' social class, parents' ethnic group, parents' working hours per week, students' O-level grade dummy for any main subjects, family income and family size are used to estimate propensity score; for college eligible subsample, number of A-levels is included as well. All specifications are estimated on common support. While Nearest Neighbour matching reports subsampling standard errors, Kernel and Stratification matching report conventional bootstrapping standard errors in parentheses; subsampling and bootstrapping standard errors are calculated based on 800 replications. Kernel bandwidth is .06.

*** p<0.01, ** p<0.05, * p<0.1

Reference

Abadie, A. and Imbens, G. W. (2008), "On the Failure of the Bootstrap for Matching Estimators", *Econometrica, Vol.76, Issue 6, pp. 1537-1557, November, 2008.*

Angrist, J. D. and Pischke, Jörn-Steffen (2009), "Mostly Harmless Econometrics: An Empiricist's Companion", *Princeton University Press.*

Belley, P. and Lochner, L. (2007), "The Changing Role of Family Income and Ability in Determining Education Achievement", *NBER Working Paper No.13527.*

Berg Dale, S. and Krueger, A. B. (2002), "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables", *The Quarterly Journal of Economics, Vol. 117, No. 4, pp. 1491-1527.*

Blackburn M. L. and Neumark, D. (1992), "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials", *The Quarterly Journal of Economics, Vol. 107, No. 4, pp. 1421-1436.*

Blackburn, M. L. and Neumark, D. (1993), "Omitted-Ability Bias and the Increase in the Return to Schooling", *Journal of Labour Economics, Vol. 11, No. 3, pp. 521-544.*

Blundell, R., Dearden, L., Goodman, A., and Reed, H. (2000), "The Return to Higher Education in Britain: Evidence from a British Cohort", *The Economic Journal, 110 (Feb), F82-F99.*

Brand, J. E. and Halaby C. N. (2003), "Regression and Matching Estimates of the Effect of Elite College Attendance and Career Outcomes", *Conference Paper for 2003 Annual Meeting of the American Sociological Association, Atlanta.*

Caponi, V. and Plesca, M. (2009), "Post-Secondary Education in Canada: Can Ability Bias Explain the Earnings Gap Between College and University Graduates?" *Canadian Journal of Economics, 42(3), pp. 1100-1131.*

Chevalier, A., Harmon, C., Walker, I., and Zhu, Y. (2004), "Does education raise productivity, or just reflect it?", *Economic Journal, 114(499).*

Dearden, L. (1999), "The Effects of Families and Ability on Men's Education and Earnings in Britain", *Labour Economics 6, pp. 551-567.*

Dehejia, R. and Wahba, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association 94(448, .pp.1053-1062.*

Devereux, P. and Fan, W. (2011), "Earnings Returns to the British Education Expansion", *Economics of Education Review, forthcoming, 2011.*

Galindo-Rueda, F. and Vignoles, A. (2004), "The Heterogeneous Effect of Selection in Secondary Schools: Understanding the Changing Role of Ability", *IZA Discussion Paper No.1245.*

Grogger J. and Eide, E. (1995), "Changes in College Skills and the Rise in the College Wage Premium", *The Journal of Human Resources, Vol. 30, No. 2, pp. 280-310.*

Hansen, K. T., Heckman, J. J. and Mullen, K. J. (2004), "The Effect of Schooling and Ability on Achievement Test Scores", *Journal of Econometrics, 121(1-2), pp. 39-98.*

Juhn, C., Murphy, K. M. and Pierce, B. (1993), "Wage Inequality and the Rise in Return to Skill", *The Journal of Political Economy, Vol. 101, No. 3, pp. 410-442.*

Murnane, R. J., Willett, J. B. and Levy, F. (1995), "The Growing Importance of Cognitive Skills in Wage Determination", *Review of Economics and Statistics, Vol. 77, pp. 251-266.*

Rosenbaum, P. R. and Rubin, D. B. (1985), “Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score”, *The American Statistician, Vol. 39., No. 1 (Feb., 1985), pp. 33-38.*

Sloane, P. J. and O’Leary, N. C. (2005), “The Return to a University Education in Great Britain”, *National Institute Economic Review, 193(1), pp. 75-89.*

Taber, C. R. (2001), “The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?”, *The Review of Economic Studies, Vol. 68, No. 3, pp.665-691.*

Walker, I. and Zhu, Y. (2008), “The College Wage Premium and the Expansion of Higher Education in the UK”, *The Scandinavian Journal of Economics 110(4), pp. 695-709.*

Winship, C. and Mare R. D. (1992) “Models for Sample Selection Bias”, *Annual Review of Sociology, 1992, 18: pp. 327-350.*

Appendix Table A
Estimation of Propensity Score

Dependent Var: **HEA**	**Full Sample Coef. (S.E.)**		**College Eligible Subsample Coef (S.E.)**	
Number of A-levels	---		0.60***	(0.12)
Female	-0.10	(0.08)	-0.03	(0.16)
British	-0.32	(0.35)	-0.26	(1.46)
Father's social class	-0.17***	(0.04)	-0.09	(0.07)
Mother's social class	-0.10***	(0.03)	0.01	(0.07)
Father's age left full time education	0.09***	(0.03)	0.11**	(0.04)
Mother's age left full time education	0.05**	(0.02)	0.03	(0.04)
Father's hours worked per week	-0.00	(0.00)	0.01	(0.01)
Mother's hours worked per week	-0.00	(0.00)	-0.01	(0.01)
Father's ethnic group	-0.01	(0.07)	-0.06	(0.18)
Mother's ethnic group	0.10	(0.07)	0.28	(0.31)
Grade A dummy of Main Subjects of O levels	1.16***	(0.10)	0.29*	(0.17)
Family income	-0.02**	(0.01)	-0.01	(0.02)
Family size (number of people sharing the food)	0.01	(0.02)	0.02	(0.05)
Sample size	1443		413	

Notes: Probit model is used to estimate propensity score. The dependent variable is Higher Education Attendee (HEA) by definition. All specifications also include Sex, ethnic group, parents' age left full time education, parents' social class, parents' ethnic group, parents' working hours per week, students' O-level grade dummy for any main subjects, family income and family size; for college eligible subsample, number of A-levels is included as well.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix Table B
Covariate Imbalance for Variable with Substantial Initial Bias
(Standardized Absolute Bias > 20%)

Panel A FULL SAMPLE

	Before Matching		After Matching	
	Two-sample T statistics	**Standardized Difference %**	**Two-sample T statistics**	**Standardized Difference %**
Father's social class	-12.19	-67	0.70	5
Mother's social class	-7.54	-42	-0.46	-3
Father's age left full time education	13.36	66	0.28	2
Mother's age left full time education	11.56	59	-0.81	-7
Grade A dummy of any Main Subjects of O levels	17.14	87	0.26	2
Sample size (treated/control)		493/943		

Panel B COLLEGE ELIGIBLE SUBSAMPLE

	Before Matching		After Matching	
	Two-sample T statistics	**Standardized Difference %**	**Two-sample T statistics**	**Standardized Difference %**
Number of A-levels	7.09	88	0.36	3
Father's social class	-3.69	-46	-0.31	-6
Father's age left full time education	4.62	66	1.85	15
Mother's age left full time education	3.23	44	0.95	8
Grade A dummy of any Main Subjects of O levels	4.41	56	1.56	13
Sample size (treated/control)		328/85		

Appendix Table C

Propensity Score Regressions Estimates on Return to College
(Standard Errors in Parentheses)

Y=Log (weekly wage)	**[1]** **Full Sample**	**[2]** **College Eligible Subsample**
OLS (Y = *f*(D, p(X))	0.638***	0.452**
	(0.083)	(0.183)
	[N=1,436]	[N=413]

Notes: All specifications include sex, ethnic group, parents' age left full time education, parents' social class, parents' ethnic group, parents' working hours per week, students' O-level grade dummy for any main subjects, family income and family size. Column 2 also includes number of A-levels in specification. Standard errors are reported in round brackets whereas sample sizes are reported in squared brackets.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP10/34 Ivan Pastine and Tuvana Pastine: "Political Campaign Spending Limits" October 2010
WP10/35 Brendan Walsh and Dermot Walsh: "Suicide in Ireland: The Influence of Alcohol and Unemployment" October 2010
WP10/36 Kevin Denny: "Height and well-being amongst older Europeans" October 2010
WP10/37 Alan Fernihough: "Malthusian Dynamics in a Diverging Europe: Northern Italy 1650-1881" November 2010
WP10/38 Cormac Ó Gráda: "The Last Major Irish Bank Failure: Lessons for Today?" November 2010
WP10/39 Kevin Denny and Veruska Oppedisano: "Class Size Effects: Evidence Using a New Estimation Technique" December 2010
WP10/40 Robert Gillanders and Karl Whelan: "Open For Business? Institutions, Business Environment and Economic Development" December 2010
WP10/41 Karl Whelan: "EU Economic Governance: Less Might Work Better Than More" December 2010
WP11/01 Svetlana Batrakova: 'Flip Side of the Pollution Haven: Do Export Destinations Matter?' January 2011
WP11/02 Olivier Bargain, Mathias Dolls, Dirk Neumann, Andreas Peichl and Sebastian Siegloch: 'Tax-Benefit Systems in Europe and the US: Between Equity and Efficiency' January 2011
WP11/03 Cormac Ó Gráda: 'Great Leap into Famine' January 2011
WP11/04 Alpaslan Akay, Olivier Bargain, and Klaus F Zimmermann: 'Relative Concerns of Rural-to-Urban Migrants in China' January 2011
WP11/05 Matthew T Cole: 'Distorted Trade Barriers' February 2011
WP11/06 Michael Breen and Robert Gillanders: 'Corruption, Institutions and Regulation' March 2011
WP11/07 Olivier Bargain and Olivier Donni: 'Optimal Commodity Taxation and Redistribution within Households' March 2011
WP11/08 Kevin Denny: 'Civic Returns to Education: its Effect on Homophobia' April 2011
WP11/09 Karl Whelan: 'Ireland's Sovereign Debt Crisis' May 2011
WP11/10 Morgan Kelly and Cormac Ó Gráda: 'The Preventive Check in Medieval and Pre-industrial England' May 2011
WP11/11 Paul J Devereux and Wen Fan: 'Earnings Returns to the British Education Expansion' June 2011
WP11/12 Cormac Ó Gráda: 'Five Crises' June 2011
WP11/13 Alan Fernihough: 'Human Capital and the Quantity-Quality Trade-Off during the Demographic Transition: New Evidence from Ireland' July 2011
WP11/14 Olivier Bargain, Kristian Orsini and Andreas Peichl: 'Labor Supply Elasticities in Europe and the US' July 2011
WP11/15 Christian Bauer, Ronald B Davies and Andreas Haufler: 'Economic Integration and the Optimal Corporate Tax Structure with Heterogeneous Firms' August 2011
WP11/16 Robert Gillanders: 'The Effects of Foreign Aid in Sub-Saharan Africa' August 2011
WP11/17 Morgan Kelly: 'A Note on the Size Distribution of Irish Mortgages' August 2011
WP11/18 Vincent Hogan, Patrick Massey and Shane Massey: 'Late Conversion: The Impact of Professionalism on European Rugby Union' September 2011