UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Extracurricular educational activities: do they pay off?

van Ophem, J.C.M.; Chin, Joey

Publication date
2017

Link to publication

Citation for published version (APA):
van Ophem, J. C. M., & Chin, J. (2017). *Extracurricular educational activities: do they pay off?* (pp. 1-26).

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:27 Nov 2021

Extracurricular educational activities: do they pay off?

Joey Chin

Amsterdam School of Economics, University of Amsterdam

and

Hans van Ophem*

Amsterdam School of Economics, University of Amsterdam

and

Tinbergen Institute

October 2017

Abstract

In this paper we investigate the effect of two types of extracurricular activities during tertiary education on wages in the Netherlands. We distinguish activities directly related to the study: doing an honours programme, doing more than one educational programme, and earning a cum laude degree, and activities not directly related to the study: becoming a member of a student association. Depending on the estimation method used we find ambivalent results. Employing parametric estimation methods gives significant effects: positive for the non study related activities and usually negative for the study related activities. Semiparametric estimation reduces the significance of the estimation results considerably: in many all casea extracurricular activities do not appear to have any effect on wages, especially if we allow for heterogeneity across level and field of study.

*Corresponding author. Address: Amsterdam School of Economics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.Email: j.c.m.vanophem@uva.nl.

Keywords: extracurricular activities, wages, parametric and semiparametric estimation
JEL code: I23, I26 and J31.

1 Introduction

In recent years the higher education system in the Netherlands has faced many reforms, for example replacing student funding by a borrowing system and abolishing subsidies for tuition of a second bachelor's or master's degree. Most of these measures are taken to reduce the public costs of higher education by stimulating students to graduate more quickly. Students are asked to pay a greater part of their study costs themselves and are punished financially when they study longer than considered necessary. The intended effect of these measures is that students become more devoted to their study program, but the downside is that they have less time to get involved in other activities relevant for their personal development, such as extracurricular activities. Past research in the United States has shown that participation in such activities is profitable for students, see e.g. Massoni (2011). Students themselves are also aware that the possibilities for development besides studying get more and more limited. More recently, this awareness was one of the main drivers behind a proposal to let students pay per course instead of an annual tuition. This would provide students with more possibilities to compose their own study program with room for extracurricular activities. In the Netherlands, this possibility has been opened up recently to students to a very limited extent. Despite of this, students face increased pressure due to increased competition in the job market after graduation. The reason for this is that more and more people have access to higher education, which decreases the value of formal credentials, cf. Collins (2002). So students need to study more quickly and engage in extracurricular activities to distinguish themselves from the crowd. Students have to deal with this situation and consider their time and financial resources carefully.

The choices students have to make during their educational career can be highly influential on their future wages. Some choices have to be made on an early age, well before the actual studying at the university of college commences. This is particularly true in the Netherlands. At the age of 12, children change from their elementary school to a high school. They transfer to one of three levels of high schools: VMBO, HAVO and VWO. These high schools differ in academic content and duration. VWO prepares for university education and is the most academically challenging high school. HAVO prepares for college education (HBO) and on top of that prepares for specific professions as is revealed by the translation of abbreviation that HBO stands for: higher vocational education. VMBO prepares for middle level vocational education (MBO), and is from an academic point of view the least challenging. The duration of the high school education is 4 (VMBO), 5 (HAVO) or 6 (VWO) years. To be admitted to a Dutch university, students need to have completed VWO or HBO. Entrance to HBO requires a diploma on the level of HAVO (or VWO) or MBO.[1] Admittance to MBO requires having finished either one of the three levels of high schools.

In this investigation we concentrate on the effect of engaging in extra curricular activities while studying in one of the two highest levels of tertiary education and we allow for possible endogeneity of this choice. However during the entire educational career of an individual many choices need to be made. Some of these choices, like choosing the level of high school or the field of study, will influ-

[1]Quite often there are additional requirement on the subjects taken in high school or even grades earned.

ence wages as well. To keep our research tractable we will consider other educational choices to be exogenous. Furthermore, we believe that we have some good arguments to impose these restrictions.

The selection into high school level is based on two factors: the advice of the elementary school and one or more formal tests done in the last years of elementary schooling. Given that this choice has to be made on a very early age and that in fact the individual pupil or the parents do not have really an influence on the outcome, we believe that we can safely assume that the choice of the high school level is exogenous. However, the initial high school choice does not completely determine the final educational level reached by the individual. In the past, some students changed to a higher level high school after having finished their original high school. This usually took one additional year, i.e. doing VWO after having finished HAVO would require two years of additional high school at the VWO level. Nowadays, this route is no longer very popular and if students attempt to earn a higher level degree they usually do so by first finishing MBO or HBO before they transfer to HBO or university. But, clearly this choice on a follow up education is a true individual choice and this might introduce a selectivity problem. We decided to ignore this problem because we want to correct for the endogeneity of extracurricular activities and allowing for additional endogeneity will make the model intractable. To give an idea about the number of students taking the alternative route: 9.1% of the university graduates in our sample also have an HBO-diploma and 25.7% of the HBO-graduates also have an MBO-diploma.

Another choice that pupils have to make is on the subjects they want to study in high school. In the first two years of high school, pupils take all subjects offered by the high school (usually 10-15 subjects) and after that they have to choose a number of subjects at least 6 (VMBO, HAVO) or 7 (VWO) subjects and usually offered in a smaller number of packages of related subjects. The packages determine which specialisation the individual can choose at the MBO, HBO or university level. For example, to study economics at a university it is required to have done mathematics at the middle or highest level of the three possible levels offered at a VWO. Despite the fact that pupils have to make this choice at the early age of 14 or 15, it is more likely that they self have at least some saying in this choice for their preferred package of subjects. On the other hand, the high school gives advice and usually pupils simply choose subjects they like and are good at. As a consequence, we believe that the endogeneity problem introduced by this choice is rather limited and we will ignore it in our analysis.

In this paper we concentrate on the effect of extracurricular activities on wages early in the career of HBO and university graduates. We quantify the value of participating in extracurricular activities and thus giving a broader perspective on the consequence of the current reforms. We believe that pursuing extracurricular activities is much more a conscious choice. Not only is this choice made at a much more mature age but also during their final stage of education. As a consequence we will allow for endogeneity of the choices related to extracurricular activities. Since extracurricular activities are a broad concept, we will use the categorization of Tchibozo (2007) to distinguish two types of extracurricular activity: leisure related and study related. We emphasize that we focus on the effects on an individual level and thus not approach this question from a regulatory standpoint, at which we would measure the value for society when giving students more room to participate in extracurricular activities.

For our analysis, we will use data from *SEO Amsterdam Economics*, a Dutch institute for economic research, which distributes a survey under university and HBO graduates in the Netherlands every year. These repeated cross sections are combined to yield 56320 observations. In order to examine the effects of extracurricular choices on wages earned about 1.5 years after graduation, we apply a selection model that is estimated both parametrically and semiparametrically. As noted, we distinguish two different types of activities giving rise to the necessity to take two different endogenous individual choices into account. We apply extensions of the Heckman selectivity model (parametric case, cf. Heckman (1979)) and the semiparametric estimation methods of the Heckman selectivity model proposed by Cosslett (1991) and Newey (2009).

Our results show that the use of selection models are appropriate here, that is, the participation decisions are indeed endogenous. In both the parametric and the semiparametric models we find that the participation in extracurricular activities not directly related to study positively affects salary. The study related extracurricular usually have in some cases a negative effect and also a reduced significance. The significance of the parameters in the semiparametric estimation is reduced considerably.

The contribution of this paper is two-fold. To our knowledge this paper is the first to examine extracurricular activities with a selection model. All other researches are based on linear regression, logistic models or instrumental variables models. Secondly, this paper is the first to apply a selection model with two selection dummy variables using a semiparametric estimation technique.

The remainder of this paper is organized as follows. In section 2 we provide with some background information on the concept of extracurricular activities. We discuss the definition of this concept, the examined effects and the methods which were used in examination of these effects. In section 3 we give insights in selection models and we discuss the models used for this paper. Section 4 discusses the data set used in this paper. In section 5 we discuss the results of our model. The conclusion can be found in section 6.

2 Theoretical background

2.1 Definition of extracurricular activities

Extracurricular activities were introduced in the United States by the introduction of literacy clubs at Harvard University and Yale University in the 19th century. These clubs are the first known extracurricular activities, see Massoni (2011). After that, more and more types of activities were added. Several American schools added academic clubs, school clubs or athletic clubs. Today, schools and universities are surrounded by different types of activities for students with different interests.

All these activities can be seen as extracurricular activities, which Valentine et al. (2002) defined as activities which take place in the context of an educational institution and its environment. These distinguish themselves from regular curricular activities in three ways. First, these extracurricular activities are optional, i.e. students are not obliged to participate in them. Second, extracurricular activities are not graded and are thus not listed on any certificate received after graduation. Third, the activities take place outside regular study hours, cf. Mahoney and Cairns (1997).

Because of the great variety of activities, researchers have made an attempt to categorize extracurricular activities. Valentine et al. (2002) distinguished five categories: study related activities, sports, cultural clubs, social or political activities and unstructured activities. Another categorization is applied by Tchibozo (2007), who distinguished two types of activities: employment activities, and leisure or social activities. In this paper, we apply a modified form of the latter categorization. Like Tchibozo (2007), we distinguish two types of activities. We adopt the definition of leisure or social activities and abbreviate this to *leisure related activities*. However, we do not include employment activities due lack of information. Instead we define *study related activities*, which are extracurricular activities which have to do with study affairs, such as doing extra courses, doing a different study program as well or aiming at high grades. This will be further explained in section 4.

2.2 Effects of extracurricular activities

Different effects of extracurricular activities can be distinguished. We will discuss effects on learning, personal development and the job market. Extracurricular activities can also studied in the context of for example college decision making (DesJardins et al. (1999)) and generosity of alumni (Tucker (2004)). Since this does not bear direct relevance for the present research, these potential effects will not be discussed.

2.2.1 Effects on learning

Although extracurricular activities consume time, which could also be allocated to studying, participation in these activities is associated with multiple positive effects on education (see Massoni (2011)). All types of extracurricular activities can induce an increase in the engagement of students. Engagement is considered to be one of the most important aspects for students. Students develop a positive attitude towards learning and enlarge their interest in studying (cf. Marks (2000). Furthermore, they invest more time and energy in education which causes lower drop-out rates (cf. Mahoney and Cairns (1997)) and results in higher grades (cf. Marks (2000)). These enhancing effects on academic performance is estimated to be larger than the effects resulting from paid student jobs (see Cooper et al. (1999)).

2.2.2 Effects on personal development

Besides positive effects on learning, the participation in extracurricular activities also contributes to a better personal development. Participants in extracurricular activities are associated with a lower probability of showing risky behaviour. The probability of excessive drug or alcohol use is lower than for non-participants, cf. Eccles and Barber (1999). Furthermore, extracurricular activities provide room to develop better self-esteem, see Holland and Andre (1987). Students become more aware of their capacities and participation is also related with a development of individual and group responsibilities, see Finn and Voelkl (1993). This is also reflected in the positive effects on learning, which we discussed earlier. Furthermore, extracurricular activities offer the possibility to develop social norms (see Willems (1967)), which will also be useful in a job setting.

2.2.3 Effects on the job market

Compared to other potential effects of extracurricular activities, effects on the job market have generally been ignored. Due to the increased number of graduates who have participated in extracurricular activities, the number of researches on this topic has been increasing in recent years and this paper also contributes to this topic. The effects have been studied in various countries. Tchibozo (2007) has analysed transition to the job market in the United Kingdom. He found that extracurricular experience gives a better occupational status. He also noticed that these activities lengthened the period of unemployment between graduation and the first job. In the United States, Eide and Roman (2001) found that participation in sports activities contributes to a better wage. The higher wages may be a result of leadership experience. Kuhn and Weinberger (2005) found that leadership experience gives access to higher wages and that it also increases the chances of becoming a manager. Extracurricular activities may also result in improved ability to perform in job interviews. Ming Chia (2005) examined the process towards a first job. He investigated the effects of academic performance, extracurricular activities and emotional intelligence. He concentrated on job offers of the Big 5 public accounting firms. Results indicated that emotional intelligence enhances the graduate's performance at job interviews and extracurricular activities are stated to be an appropriate experience to develop emotional intelligence.

Students mostly participate in extracurricular activities to gain skills or expand their social network (LKvK (2012)). This personal development is also acknowledged by employers. Companies are not looking for students who only have a high educational level. Students need also to be developed on a social level. Indeed, big consultancy firms appear to attach more value to extracurricular activities than graduating cum laude (Ackerman (2008)).

The recognition of the value of extracurricular activities by employers causes more and more students to participate (ASVA (2010)). Moreover, students feel extra pressure due to a highly competitive job market after finishing their higher educational program. One of the main reasons for this is the rise of mass higher education (Brown et al. (2003)). More and more people have access to higher education and this results in a high number of graduates with a similar degree and experience entering the job market and competing for a limited number of jobs. Because of the increase in number of graduates, the value of formal credentials like university degrees has decreased (Collins (2002). Students themselves also have realized this and they know that degrees are not a guarantee of future employability anymore (Tomlinson (2007)). To improve their chances, students need to distinguish themselves from other students both during and besides studying. Participation in extracurricular activities is one way to accomplish this. The increased competition is also visible in the Netherlands, where the number of students has steadily grown in the last decades (CBS (2015)) and the prognosis for the next decade is an even further increase (VSNU (2012)).

2.3 Important differences between the United States and the Netherlands

Most of the aforementioned discussed literature on extracurricular activities relates to the United States. However, there are a number of important differences between the Netherlands and the United States concerning extracurricular activities. First, the choice of university is much more important in the United States than it is in the Netherlands. Extracurricular activities can increase chances to get access to a good university. Therefore, the majority of the studies on extracurricular activities is executed at high schools. In the Netherlands, grades usually do not determine whether one has access to a university. Only an appropriate high school certificate is sufficient. Moreover, the universities hardly differ in quality. As a consequence, extracurricular activities are less important at a high school level. Another difference between the countries is the focus on sports. In the United States, outstanding athletic abilities can be rewarded with an admission to a good university. Athletics is therefore more intensively exercised than in the Netherlands, where sports count as an important form of leisure. The last difference is the structure of the higher educational system. For one thing, tuition in the Netherlands is much lower since it is heavily subsidized by the government. As we mentioned before, another important difference is that students are selected into types of education, for the present research the relevant types are university and college education, at the early age of 12 and that different levels of high school exist and this restricts the available alternatives for tertiary education.

2.4 Student associations

We need to explain the concept of student associations in the Netherlands since this will be included in our model as an explanatory variable. There are typically two types of associations: study associations and student associations . Study associations are typically affiliated with one specific study or sometimes a group of similar studies. These associations unite students with a similar study background by organizing events. In addition, these associations can effectively organize study related events or setup study related structures like discount on book sale, study trips or tutoring classes. A great number of students is member of study associations without being actively involved. As a consequence, the relation between participation in student associations and wages may be heavily influenced by non-active members. In contrast, student associations are not study specific. Therefore, they encompass a much broader range of types of students. These associations put emphasis on a specific activity. Mostly the emphasis lays on the social aspect. Other examples are an emphasis on international relations, religion or philosophy. Student associations can be compared to fraternities and sororities in the United States. These associations are more interesting for our study, since members are more likely to be active at these associations and as a consequence they are more likely to develop skills that will affect their future wage. We therefore will concentrate on estimating the effect of participating in student associations.

3 Empirical model

We estimate a wage equation with two dummies representing extracurricular activities employed by the student and a large number of controls. We distinguish leisure related (ECAL) and study related extracurricular activities (ECAS). For individual i ($i = 1, ..., N$) we write:

$$ln(w_i) = \beta' X_i + \alpha_1 ECAL_i + \alpha_2 ECAS_i + \varepsilon_i \quad (1)$$

In order to get unbiased estimates if the wage equation is estimated with ordinary least squares we require that the error term ε_i does not correlate with the explanatory variables. Since engaging in extracurricular activities is an individual choice, this requirement is quite likely violated for the explanatory variables *ECAL* and *ECAS*. However, we do assume that it holds for the controls.[2] The usual way to model this endogeneity is by extending eq. 1 with two choice or selection equations:

$$log(w_i) = X_i' \beta + \alpha_1 ECAL_i + \alpha_2 ECAS_i + \varepsilon_i \quad (2)$$

$$ECAL_i = \ell(Z_{1i}' \gamma_L + v_i^L > 0)$$

$$ECAS_i = \ell(Z_{2i}' \gamma_S + v_i^S > 0)$$

where $\ell(.)$ is the indicator function and where we allow v_i^L and v_i^S to correlate with ε_i. To cope with the potential selection bias we will apply a parametric model inspired by Heckman (1979) and semiparametric estimation methods inspired by Cosslett (1991) and Newey (2009) in which we relax the distributional assumptions. Both original estimation methods deal with only one endogenous dummy explanatory variable. We extend the original model in a straightforward fashion and estimate the model of eq. 2 in two steps.

First stage estimation

In the first stage we estimate the selection equations. For the parametric model of Heckman, assuming a trivariate normal distribution of $(\varepsilon_i, v_i^L, v_i^S)$, we estimate this with the bivariate probit model (cf. Cameron and Trivedi, 2005, p. 521):

$$\Pr(ECAL_i = 1; ECAS_i = 1, |Z_{1i}, Z_{2i}) = \Phi(Z_{Li}' \gamma_L, Z_{Si}' \gamma_L; \rho_{LS}). \quad (3)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the bivariate standard normal distribution.

[2]In section 5 we will check whether this is indeed true for some important explanatory variables.

For the semiparametric model of Cosslett, we also relax the assumption of normal distributed errors and estimate the selection equations semiparametrically. For this we choose the semiparametric estimator proposed Gallant and Nychka (1987). They specify the density by a Hermite series and write it as a squared polynomial times a normal distribution with a polynomial expansion with Gaussian leading term in the following way:

$$f_K(\varepsilon) = \frac{1}{\tilde{\theta}} \left(\sum_{k=1}^{K} \theta_k v^k \right)^2 \phi(v), \tag{4}$$

$$\tilde{\theta} = \int_{-\infty}^{\infty} \left(\sum_{k=1}^{K} \gamma_k v^k \right)^2 \phi(v) dv$$

where $\phi(\cdot)$ the univariate normal density function. This general density function is taken to be the density of v_i^L and v_i^A and the resulting model is estimated with maximum likelihood. We will ignore the correlation between v_i^L and v_i^A in the semiparametric case. We have two arguments for doing this: (1) as indicated by Cameron and Trivedi (2005, p. 150) ignoring the correlation might give rise to a loss of efficiency but does not introduce a bias. Ignoring the correlation speeds up the estimations, and especially the necessary bootstrapping of the standard errors considerably. Applying the method of Gallant and Nychka (1987) is possible as shown in e.g. Van der Klauw and Koning (2003) but the order of the polynomials has to be limited to a very strong degree to get estimation results. (2) Applying the bivariate probit model, gives an estimate of the correlation of -0.026 (significant at 5%). Given this relatively small absolute magnitude the effect of the correlation is expected to be very modest.[3]

Second stage estimation

For two selection dummies the conditional expectation of the errors becomes more complex. For instance, when both dummies are equal to one, we get the following conditional expectation:

$$\begin{aligned} \mathrm{E}\left(log(w_i)|ECAL_i = 1, ECAS_i = 1\right) &= \mathrm{E}\left(X_i'\beta + \alpha_1 ECAL_i + \alpha_2 ECAS_i + \mathrm{E}(\varepsilon_i|ECAL_i = 1, ECAS_i = 1\right) \\ &= X_i'\beta + \alpha_1 ECAL_i + \alpha_2 ECAS_i + \mathrm{E}\left(\varepsilon_i|\alpha_2 ECAL_i = 1, \alpha_2 ECAS_i = 1\right), \\ &= X_i'\beta + \alpha_1 ECAL_i + \alpha_2 ECAS_i + \mathrm{E}\left(\varepsilon_i|v_i^L > -Z_{Li}'\gamma_L, i > -Z_{Si}'\gamma_S\right) \\ &= X_i'\beta + \alpha_1 ECAL_i + \alpha_2 ECAS_i + g(Z_{Li}'\gamma_L, Z_{Si}'\gamma_S) \end{aligned} \tag{5}$$

for $i = 1, \ldots, N$. Therefore, we need an estimate for the last term on the right-hand side of (5). For this, we will use three methods: estimating $g(.)$ under the assumption of normal distributed errors similar to the Heckman selectivity model, (2) approximating $g(.)$ using the method of Cosslett (1991)

[3]Ignoring the correlation in the parametric Heckman setup hardly has any effect on the final estimation results of the wage equation.

and (3) approximating $g(.)$ using the method of Newey (2009).

Under the assumption of normality, the expression of $g(.)$ is given in Pudney (1989). This conditional expectation can be seen as a truncation rule with two truncation boundaries. Corresponding to Pudney's notation for two truncation boundaties, we can describe this compound event as:

$$\Xi_i = \{\nu_i^L \geq c_{Li}, \nu_i^S \geq c_{Si}\}. \tag{6}$$

Define the thresholds as:

$$c_{Li} = \begin{cases} -Z'_{Li}\gamma_L & \text{if } ECAL_i = 1, \\ Z'_{Li}\gamma_L & \text{if } ECAL_i = 0, \end{cases} \qquad c_{Si} = \begin{cases} -Z'_{Si}\gamma_S & \text{if } ECAS_i = 1, \\ Z'_{Si}\gamma_S & \text{if } ECAS_i = 0, \end{cases}$$

for $i = 1, \ldots, N$. Ultimately, we get the following expectation:

$$g(Z'_{Li}\gamma_L, Z'_{Si}\gamma_S) = \pi_L P_{Li} + \pi_S P_{Si}, \tag{7}$$

$$P_{Li} = \frac{\phi(c_{Li})}{\Pr(\Xi_i)} \left\{1 - \Phi\left(\frac{c_{Si} - \rho_{LS}c_{Li}}{(1-\rho_{LS}^2)^{1/2}}\right)\right\} \tag{8}$$

$$P_{Si} = \frac{\phi(c_{Si}^*)}{\Pr(\Xi_i)} \left\{1 - \Phi\left(\frac{c_{Li} - \rho_{LS}c_{Si}}{(1-\rho_{LS}^2)^{1/2}}\right)\right\} \tag{9}$$

Estimation with ordinary least squares of

$$log(w_i|ECAL_i = 0,1, ECAS_i = 0,1) = X_i'\beta + \alpha_1 ECAL_i + \alpha_2 ECAS_i + \pi_L P_{Li} + \pi_S P_{Si} + \varepsilon_i^*, \tag{10}$$

for $i = 1, \ldots, N$ and $\varepsilon_i^* = \varepsilon_i - \pi_L P_{Li} - \pi_S P_{Si}$ will result in consistent estimates, provided that the unknown parameters in (8) and (9) are replaced by consistent estimates, because

$$\mathrm{E}(\varepsilon_i^*|ECAL_i = 0,1, ECAS_i = 0,1) = 0.$$

Of course, the standard errors of these OLS-estimators need to be corrected due to heteroskedasticity and plugging in estimates. We do this by bootstrapping the standard errors nonparametrically.

The second estimation method is the semiparametric method of Cosslett (1991). This method can be seen as the semiparametric analogue of Heckman's selection model. In this model the selection bias is estimated using selection correction dummies. These dummies are defined by cutting the value-ordered propensity scores $Z'_{Li}\hat{\gamma}_L$ and $Z'_{Si}\hat{\gamma}_S$ both into M intervals.[4] For two selection variables, we estimate the selection bias by approximating $g(.)$ using a large number of dummies. These are the interaction terms between the selection correction dummies per propensity score $Z'_{Li}\hat{\gamma}_L$ and $Z'_{Si}\hat{\gamma}_S$. This interaction term naturally incorporates the correlation between the two selection dummies. We

[4]In fact, it is not necessary to use the same number of intervals for both propensity scores. we have experimented with determining the number of intervals in different ways. The method advocated by Cosslett (1983) and Hussinger (2008), Ayer's algorithm (cf. Ayer et al. (1955), resulted in an unfeasible number of intervals. In the end we decided to use 20 intervals per propensity score, giving rise to 1599 dummies.

get:

$$y_i = X_i'\beta + \alpha_1 ECAL_i + \alpha_2 ECAS_i + \sum_{m_1=1}^{M}\sum_{m_2=1}^{M} \lambda_{m_1,m_2}^{ECAL_i,ECAS_i} C_{m_1,i}^{L}(Z_{Li}'\gamma_L)\cdot C_{m_2,i}^{S}(Z_{Si}'\gamma_L) + \xi_i, \qquad (11)$$

for $i = 1,\ldots N$. $\lambda_{m_1,m_2}^{ECAL_i,ECAS_i}$ are additional parameters to be estimated[5] and

$$\xi_i = \varepsilon_i - \sum_{m_1=1}^{M}\sum_{m_2=1}^{M} \lambda_{m_1,m_2}^{ECAL_i,ECASD_i} C_{m_1,i}^{L}(Z_{Li}'\gamma_L)\cdot C_{m_2,i}^{S}(Z_{Si}'\gamma_L). \qquad (12)$$

$C_{m_1,i}^{j}$ is an indicator function equal to one if the argument is in interval m_1. Cosslett's argument for consistency is that by using a dummy variable specification of order M any selectivity term can be approximated to any degree by choosing M high enough, To prove consistency, M should increase with N, but at a smaller rate. If that is the case it holds that:

$$E\left(\xi_i | ECAL_i = 0,1, ECAS_i = 0,1\right) \longrightarrow 0. \qquad (13)$$

Note that the number of extra parameters to be estimated is quite large. If we use M dummies per propensity score the number of additional parameters is $4M^2$. Again, the OLS-standard errors are incorrect so we will use bootstrapping to get the correct standard errors of the estimated parameters in the second step of the estimation.

In the method of Newey (2009) $g(.)$ is approximated by a polynomial. In our two dimensional case we need to extend the procedure to allow for two propensity scores. This extension is straightforward:

$$g(Z_{Li}'\gamma_L, Z_{Si}'\gamma_S) = \sum_{k=1}^{K}\sum_{j=1}^{K} \eta_{kj}\tau(Z_{Li}'\gamma_L)^{k-1}\cdot\tau(Z_{Si}'\gamma_S)^{j-1} \qquad (14)$$

where, following Newey (2009), $\tau(.) = (2\Phi(.)-1)$ and $\Phi(.)$ is the cumulative univariate standard normal distribution. The second step equation to be estimated is:

$$y_i = X_i'\beta + \alpha_1 ECAL_i + \alpha_2 ECAS_i + \sum_{k=1}^{K}\sum_{j=1}^{K} \eta_{kj}\tau(Z_{Li}'\gamma_L)^{k-1}\cdot\tau(Z_{Si}'\gamma_S)^{j-1} + \xi_i^*, \qquad (15)$$

where $\xi_i^* = \varepsilon_i - \sum_{k=1}^{K}\sum_{j=1}^{K}\eta_{kj}\tau(Z_{Li}'\gamma_L)^{k-1}\cdot\tau(Z_{Si}'\gamma_S)^{j-1}$. If we let K increase with sample size, consistency of the OLS-estimates of (15) follows as long as we use consistent estimates of γ_L and γ_S because in that case we have:

$$E\left(\xi_i^* | ECAL_i = 0,1, ECAS_i = 0,1\right) \rightarrow 0. \qquad (16)$$

Since we have to distinguish four different combination of $ECAL$ and $ECAS$, the number of additional

[5]Note that the number of additional parameters to be estimated is $4*M^2-1$. For M=20, this equals 1599.

parameters to be estimated is $4K^2$ and this number is usually considerably lower than in the Cosslett case. Again, the OLS-standard errors are incorrect so we will use bootstrapping to get the correct standard errors of the estimated parameters in the second step of the estimation.

4 Data

We use data from the database *Studie & Werk* (Study & Work). The data is collected by *SEO Amsterdam Economics* (SEO, for more info: www.seo.nl). Commissioned by Elsevier, which is a weekly Dutch news magazine, the labour market position of graduates in the Netherlands is examined with this data set. The data is collected from a written survey in which graduates from Dutch higher education institutions are questioned about their career in higher education and their career after higher education. The main objective of their research is to determine which studies give the best job market opportunities. The most important results are published yearly in a special edition of the Elsevier magazine. This research started in 1997 and since then about 120000 graduates are questioned. We used the version of Studie & Werk which was published in the period 2003-2015.[6] It contains graduates questioned until the first two months of 2015. This means that the most recent respondents in this data set graduated in the academic year 2012/2013. In total our estimation sample consists of 56320 observations.

Dependent and selection variables

We are interested in analysing the effect of extracurricular activities on (log) wages. As a dependent variable we use the log of the net hourly wage of the current job. Here, current means at the time when the survey was conducted, i.e. on average 15 months after graduation. Respondents were not obligated to share their salary which leads to 16% missing values for this variable. Furthermore, it should be noted that the salary variable has been extensively checked for correctness by SEO. The reason for this is that this variable is the most important variable in their annual research on studies and jobs.

The selection variables are the variables which incorporate the treatment decision. Here, these variables capture the participation decision on the extracurricular activities. We distinguish two types of extracurricular activities: leisure related activities (ECAL) and study related activities (ECAS). Both types of activities are incorporated in our model as a dummy variable. For leisure related activities, we use the engagement in a student association as a proxy. Student associations connect students from different fields of study and let them interact in social activities. These associations help students to expand their network and to develop themselves on a social level, rather than on an academic level. By lack of other suitable variables in the data set, this is only variable to measure engagement in leisure related activities. For study related activities, we capture whether the individual excels on

[6] Some relevant variables are only available from the 2003 survey, so we had to delete the earlier surveys from our investigation.

an academic level. For this, we use data on whether an individual has achieved cum laude, i.e. having earned an grade point average of 8.0 or higher[7] whether an individual has participated in a honours program or whether an individual has finished two study programs. If either one of these is true, than we set the dummy variable for study related activities to one and zero otherwise. For 23.7% of the respondents ECAL = 1, whereas only 5.5% of the students score on extracurricular study activities, i.e. ECAS =1. 1.4% of the respondents score on both variables.

Exogenous variables

Explanatory variables in the selection equations

To explain the decision on participating in either one of the two types of extracurricular activities, we use explanatory variables which are selected to be consistent with literature on extracurricular activities. First, we include two variables about high school achievements. We include the grade point average at high school (GPA high school) [8] and four count variables on the number of subjects chosen in the categories alpha (i.e. languages, culture, variable name #alpha), beta (i.e. science, mathematics, variable name #beta), gamma (i.e. economics, business etc., variable name #gamma) and arts (variable name #arts). Students of Dutch high schools have to choose at least six subjects. This is not a completely free choice: Dutch is compulsory, students have to choose at least one other language and do at least one mathematics course. Next, we include some individual characteristics. A dummy 'male' which is one for males and zero for females. We hypothesize that females are significantly more inclined to participate in extracurricular activities than males, cf. LKvK (2012). A dummy 'living with parents' for whether the respondent lives at his/her parent's house is included as well. We expect that these students are less inclined to participate in extracurricular activities. Accordingly, these students are less independent and live further away from university which causes their lower probability of participation. Starting age at tertiary education is also used as an explanatory variable. Younger individuals are less likely to engage in extracurricular activities since they are usually less independent and have less live experience so that they might attach less value to extracurricular activities. Since individuals with a lower socioeconomic background have a lower probability of participating in extracurricular activities, see e.g Eide and Roman (2001), we include variables to control for this. We capture the individual's background characteristics by including a dummy on whether the respondent has Dutch origins and a variable on the level of education of the mother. The nonnative-variable is one if the respondent is does not have Dutch origins, that is, whether either the mother or the father is not born in the Netherlands. The possible levels of education of the mother are (from high to low): mother uni ed (mother has completed university), mother HBO ed (mother has completed HBO), mother middle ed (mother has completed middle level of education) and mother LBO ed (mother has completed lower level vocational education (LBO)). The reference group is

[7] In the Netherlands grades in education are almost always on a scale from 1 (lowest grade) to 10 (highest grade).

[8] In 4.5% of the cases this can not be calculated. We added a dummy variable representing these observations to the specification to correct for this.

mothers that have completed only elementary schooling. Finally we include three dummy variable on the region of where the respondent has studied (East, North, South) with region West as a benchmark. The western part of the Netherlands is likely to offer more and better employment than the other parts of the Netherlands, since it is the industrial and administrative center of the Netherlands. Descriptive statistics of all variables are given in Table 1.

Explanatory variables in the wage equation

In the outcome equation we include the age at graduation, which is found to have a positive effect on wage, cf. SEO (2015). We also include a dummy variable which is one for males and zero for females. SEO (2015) finds that being a male has a positive effect of 2% on salary. A dummy which is one if the respondent lives at his/her parent's house is also included, which, according to SEO, has a negative effect of 7% on future salary. A dummy which is one for university education and zero for HBO education is included with the hypothesis that university graduates earn more because the required cognitive level is higher for university education. Using the CROHO-register we classified the studies of the respondents into nine categories: agriculture, economics, humanities, health, education, language and culture, engineering, science and law. A considerable number of studies could not be classified in one of these categories. We introduce nine dummy variables for these fields of study with those that can not be classified as the reference category. Finally, we also include dummy variables on graduation year (reference category: 2003) and region (reference category: West).

Endogeneity revisited

As in almost all economic empirical studies we have to make a choice on which variables to treat as exogenous. Our emphasis lies on the analysis of the effect of extracurricular activities of students during their studies on wages. These extracurricular activities, represented by the dummy variables ECAL and ECAS are treated as endogenous. In the selection equation representing the choice to engage in such activities we use explanatory variables that relate to characteristics of the individual (gender, education mother, nonnative), characteristics related to high school (GPA high school, numbers of specific classes) and some other variables: starting age, regions and living with parents. The first group is purely exogenous. The second group we consider to be exogenous since the characteristics relate to choices made starting at a the very young age of 11 or 12. The remaining group of variables are clearly individual choices made at the beginning of tertiary education or even later in the case of region). However, we believe that the nature of these choices are such that we can safely assume that they are unrelated to the error term in the selectivity equations.

With respect to the wage equation there are again one obvious exogenous variables (male), some variables we will assume are exogenous (region, graduation age and graduation year) and some variables that are hard to justify to being exogenous: HBO and fields of study dummies. Given the complexity of the model we want to estimate it is intangible to take these choices also into account.

To investigate this issue, we will estimate the model also for only the HBO and university graduates and we will also differentiate across fields of study, although in somewhat broader categories. We distinguish: α-studies (language and cultural studies, 3835 observations), β-studies (agricultural, nature, science and engineering studies, 10503 observations), γ-studies (economics, business, law and social studies, 18387 observations), health-studies (6592 observations) and a remaining category of studies that are hard to classify in only one of the preceding categories (e.g. education, 17003 observations).

Table 1: Descriptive statistics (N = 56320)

	Average	Stand. Dev.	Minimum	Maximum	Used in
Logwages	2.754	0.286	1.391	4.591	W
ECAL	0.237	0.425	0	1	W, S
ECAS	0.055	0.227	0	1	W, S
Starting age	19.535	2.341	16	39	S
GPA high school*	6.674	1.567	0	10	S
Male	0.453	0.498	0	1	W, S
Living with parents	0.146	0.354	0	1	W, S
Nonnative	0.062	0.240	0	1	S
North	0.090	0.286	0	1	W, S
East	0.187	0.390	0	1	W, S
South	0.197	0.397	0	1	W, S
Mother low ed	0.423	0.468	0	1	S
Mother middle ed	0.296	0.457	0	1	S
Mother HBO ed	0.254	0.435	0	1	S
Mother uni ed	0.079	0.269	0	1	S
#alpha	2.717	1.068	0	7	S
#beta	2.,229	1.449	0	6	S
#gamma	1.724	1.374	0	5	S
#art	0.450	0.742	0	6	S
Graduation age	24.761	2.467	19	41	W
HBO	0.465	0.499	0	1	W

W = used in wage equation; S = used in selection equations. * = dummy added to the specification if not observed (4.5% of the cases). Graduation year dummies are adde to the wage equation as well.

5 Empirical Results

Table 2 presents the estimation results of the selection equation both for the parametric (bivariate probit) and semiparametric (Gallant & Nychka) estimation. The polynomial parameters are significant up to a polynomial order of 3, so we present the corresponding estimates here.[9] Starting with the parametric estimation method we see that basically all estimated coefficients relating to the choice for extracurricular activities on leisure are significant at a 1% level, apart from nonnative (not significant) and the lowest level of the education of the mother (not significant). For extracurricular activities relating directly to the study, a lower significance is found but still the majority of explanatory variables has a significant effect. Younger students and female students are less likely to engage in non-study related extracurricular activities, whereas older students and males are less likely to spend their time on extracurricular activities directly related to their studies. The grades earned in high school have a positive effect on both extracurricular activities but the effect is considerably larger of activities related directly to the study. Still living with parents has a negative effect on the probability to engage in extracurricular activities. The explanation for this is straightforward: travelling home takes time and this can not be invested in other activities. Not having studies in the Western part of the Netherlands also reduces the probability to participate in extracurricular activities. The higher the education of the mother the more likely the student will invest time in extracurricular activities. Especially students who attended relatively many alpha and beta classes in high school are more interested in spending time on extracurricular activities not related to the study, whereas students with a gamma or arts interest appear to concentrate more on extracurricular activities directly related to the study. Note that the correlation between the error terms of both selection equation is very small but significant at a 5% level. Putting the correlation to zero and estimating two independent probits, gives results that are hardly distinguishable from the results presented in Table 2.

If we now turn to the semiparametric estimation of the same selection equations most conclusions listed above remain valid, although the significance has diminished somewhat. The estimated coefficients are somewhat larger than in the parametric case but that is due to the variance of the error term. It is estimated to be about 2.9 for ECAL and a little more than 3.5 for ECAS. The most prominent changes appear to be in the effect of living with parents on ECAL and the effect of the education on the mother on ECAS. In the latter case there is even a sign reversal even though the significance remains high. We do not have an explanation for this but the reference group (education mother unknown) is quite likely to be very heterogeneous. The overall tendency is that the effect of higher educated mothers brings about a higher probability of ECAS = 1.

The main object of the analysis is the wage equation. Estimation results for the Heckman selectivity model and the estimation results of the selectivity model using Cosslett and Newey semiparametric estimation methods are presented in Table 3. The sample used is the full sample consisting of 56320 observations. The differences between the estimation results of the three estimation methods employed appears to be rather marginal. Starting relatively late with working increases the wage by about 1.5% per additional year. Males earn 2-3% more than comparable females. Former students that lived with their parents have such unobserved characteristics that they earn significantly less: depending on the estimation method used minus 1.5% to minus 3.9%. Graduates from an HBO earn about 12-13% less than university students, corresponding to the notion that university schooling is

[9] Higher orders (4 and 5) were estimated but the estimation results on the wage equation were very similar.

schooling on a higher level than HBO education. Finding a job in the western part of the Netherlands, which is most likely the case if the person has studied in the western part of the Netherlands, is, from a perspective of wages, the most advantageous region. The south is second best in this respect and living in the north or east diminishes wages significantly. The graduation year dummies display the expected pattern, although perhaps with some delay: after 2011 the wages 1.5 years since the individual graduated are significantly lower than before. The most rewarding fields of study are health studies and education. From the Dutch perspective this is somewhat surprising since the present general consensus is that the substantial groups of nurses (always HBO level) and teachers (mostly HBO level) are underpaid compared to other professions. For health studies the positive effect can be explained by many health jobs requiring a university education (physicians, specialists) which are known to pay much better than health jobs at the HBO level. For the education field our explanation is that the positive result is due to the early stage we observe the wage. The wage relates to the time of 1.5 years after graduation and the growing potential in educational jobs is apparently not very good. Significant negative effects are estimated for the more technical science and agriculture fields. The least rewarding field is languages and culture.

Table 2: Estimation results ECAL and ECAS

	Bivariate probit				Gallant & Nychka			
Variable	ECAL		ECAS		ECAL		ECAS	
	Estimate	Std dev	Estimate	Std dev	Estimate	Std dev	Estimate	Std dev
Constant	-1.516	(0.108)**	-6.847	(0.169)**				
Starting age	-0.014	(0.003)**	0.045	(0.004)**	-0.016	(0.003)**	0.005	(0.004)
GPA high school	0.074	(0.011)**	0.465	(0.016)**	0.093	(0.011)**	0.276	(0.014)**
Male	0.131	(0.012)**	-0.026	(0.020)	0.150	(0.016)**	-0.030	(0.024)
Living with parents	-0.054	(0.021)**	-0.111	(0.032)**	-0.880	(0.092)**	-0.216	(0.035)**
Nonnative	-0.024	(0.025)	0.126	(0.037)**	-0.064	(0.035)	0.145	(0.044)**
North	-0.254	(0.023)**	0.040	(0.034)	-0.298	(0.031)**	-0.002	(0.042)
East	-0.286	(0.019)**	-0.017	(0.024)	-0.358	(0.029)**	-0.065	(0.031)*
South	-0.285	(0.016)**	-0.017	(0.026)	-0.368	(0.031)**	-0.046	(0.032)
Mother low ed	0.058	(0.030)	0.187	(0.061)**	0.095	(0.045)*	-4.263	(0.047)**
Mother middle ed	0.180	(0.030)**	0.236	(0.060)**	0.249	(0.045)**	-4.191	(0.048)**
Mother high ed	0.346	(0.031)**	0.405	(0.061)**	0.432	(0.047)**	-3.971	(0.049)**
Mother uni ed	0.655	(0.035)**	0.495	(0.065)**	0.739	(0.054)**	-3.780	(0.063)**
#alpha	0.086	(0.006)**	0.080	(0.012)**	0.108	(0.010)**	0.031	(0.012)*
#beta	0.087	(0.006)**	0.084	(0.009)**	0.103	(0.008)**	0.062	(0.011)**
#gamma	0.018	(0.006)**	0.100	(0.008)**	0.028	(0.007)**	0.083	(0.010)**
#art	-0.047	(0.009)**	0.249	(0.012)**	-0.058	(0.010)**	0.329	(0.016)**
Correlation	-0.026	(0.013)*						

Bootstrapped standard errors (r=250) in parentheses. */** = significant at 5%/1%.

Table 3: Estimation results logwages

Variable	OLS - Heckman		OLS - Cosslett		OLS - Newey	
	Estimate	Std dev	Estimate	Std dev	Estimate	Std dev
Constant	2.439	(0.015)**	2.397	(0.017)**	2.405	(0.096)**
Graduation age	0.014	(0.001)**	0.014	(0.001)**	0.015	(0.001)**
Male	0.030	(0.003)**	0.023	(0.002)**	0.021	(0.003)**
Living with parents	-0.039	(0.003)**	-0.021	(0.004)**	-0.015	(0.008)
HBO	-0.126	(0.003)**	-0.124	(0.003)**	-0.123	(0.003)**
North	-0.028	(0.004)**	-0.018	(0.004)**	-0.016	(0.005)**
East	-0.023	(0.003)**	-0.011	(0.003)**	-0.010	(0.004)*
South	-0.015	(0.003)**	-0.002	0.003	-0.001	(0.004)
Grad year 2004	-0.003	(0.005)	-0.004	0.005	-0.003	(0.004)
Grad year 2005	0.027	(0.005)**	0.027	(0.005)**	0.027	(0.005)**
Grad year 2006	0.053	(0.005)**	0.055	(0.005)**	0.055	(0.005)**
Grad year 2007	0.059	(0.005)**	0.062	(0.005)**	0.062	(0.005)**
Grad year 2008	0.062	(0.005)**	0.065	(0.005)**	0.065	(0.005)**
Grad year 2009	0.029	(0.005)**	0.034	(0.005)**	0.034	(0.006)**
Grad year 2010	0.003	(0.006)	0.009	(0.005)	0.008	(0.005)
Grad year 2011	-0.020	(0.005)**	-0.015	(0.005)**	-0.014	(0.005)**
Grad year 2012	-0.037	(0.005)**	-0.033	(0.006)**	-0.032	(0.006)**
Grad year 2013	-0.032	(0.006)**	-0.028	(0.006)**	-0.028	(0.006)**
Sector agri	-0.053	(0.007)**	-0.058	(0.009)**	-0.056	(0.007)**
Sector eco	0.003	(0.004)	0.003	(0.004)	0.004	(0.004)
Sector hum	-0.005	(0.004)	-0.004	(0.004)	-0.004	(0.004)
Sector health	0.123	(0.005)**	0.118	(0.005)**	0.118	(0.005)**
Sector educ	0.100	(0.005)**	0.099	(0.006)**	0.099	(0.005)**
Sector lang&cult	-0.113	(0.006)**	-0.112	(0.006)**	-0.110	(0.006)**
Sector engineering	0.007	(0.004)	0.004	(0.005)	0.004	(0.004)
Sector science	-0.043	(0.006)**	-0.047	(0.007)**	-0.047	(0.007)**
Sector law	0.025	(0.007) **	0.023	(0.008)**	0.026	(0.007)**
ECAL	0.039	(0.003)**	cf. Table 4		cf. Table 4	
ECAS	-0.028	(0.005)**	cf. Table 4		cf. Table 4	
Inv Mills ECAL	-0.026	(0.002)**				
Inv Mills ECAS	0.014	(0.002)**				
Adjusted R^2	0.128		0.132		0.129	

Bootstrapped standaard errors (r=250) in parentheses. */** = significant at 5%/1%.

For the parametric estimation we find very modest correlations of -0.026 (ECAL) and 0.014 (ECAS), so selection bias does not appear to be a real issue, despite that the correlations are estimated to be significant. The effect of engaging in extracurricular activities is positive (3.9%) for leisure related activities and negative (-2.6%) for study related activities. Both estimates being significant at an 1% level. For the semiparametric estimations we also find significant selectivity correction terms: the corresponding test statistics of these terms in the Cosslett and Newey estimation methods are F(1371,54920)=115.2 (critical value (1%): 1.089) and F(80,56211)=11.7 (critical value (1%): 1.280). Due to identification issues, we need to employ the methods of Heckman (1990) or Andrews and Schafgans (1998) to estimate the effects of participating in extracurricular activities in tertiary education.[10] In the case of the Heckman (1990) method we assume that the highest 5% or 10% of the estimated systematic part of the selection equations results in an almost certain engagement in extracurricular activities. The estimated effects are presented in Table 4. With respect to leisure-activities we find a significant effect of about 5%, somewhat higher than the estimated effect in the parametric case. In the case of study related extracurricular activities we find an effect of about 2%, significant at a 5% for one estimate, and close to being significant at 5% for the other two estimates. These positive effects of extracurricular activities are in line with theory.

One can argue that ignoring the possible endogeneity of the tertiary schooling level choice or the field of study might partly drive the results of Tables 3 and 4. It might well be that HBO and university or certain education fields attract real different students, so that the choice for extracurricular activities across these groups might be different even if the individual characteristics we observe are the same. To investigate this we will first split up the sample in HBO and university graduates. After that we will also distinguish fields of study. Since the estimation results using either the Cosslett or Newey methods are quite similar, we will only concentrate on the parametric and semiparametric Cosslett estimation results. Furthermore we only report the estimates on the effects of extracurricular activities on wage. Table 5 presents the results for HBO and university graduates separately. Starting with the parametric case, for HBO-graduates we find significant effects for both variables: ECAL +3.1% and ECAS: -3.6%. For university students the effects are only significant for leisure related activities: +3.7%. In the semiparametric case all significance is lost although the estimates point in the same direction as in the parametric case. From this we conclude that indeed the HBO and university graduates appear to behave differently, although if anything, in both cases leisure related extracurricular activities appear to pay off better.

Table 6 presents the results for different categories of studies. The parametric estimation results show considerable heterogeneity across the different fields of study. The effect of leisure related extracurricular activities is mostly positive and significant. For study related activities the evidence is mixed: we find both negative (β and γ, both HBO and university) and positive effects (α and health, but only for university education). The effects are sometimes estimated to be quite substantial:

[10]The problem is that both the dummy approximation (Cosslett) and the polynomial approximation (Newey) of the selectivity correction terms do identify the constant part of the approximation, resulting in nonidentification of the effect of ECAL and ECAS. The constant estimated in the wage equation combines the real constant plus the constant of the approximation. See see Hussinger (2008) for further explanation.

study related extracurricular activities have a -21.8% effect on the wage for university graduates in the γ-field, whereas it is +12.4% for university graduates in health studies. If we now turn to the semiparametric estimation, we have to conclude that again basically all significance is lost. The only significant effect is found for the heterogeneous 'remaining studies' category. A significant effect at a 5% level is found for leisure related extracurricular activities. All other estimated effects are not significant. In the ECAL case the tendency appears to be positive effects whereas in the ECAS case the effects are predominantly negative.

Table 4: Estimation effects of ECAL and ECAS

	Cosslett				Newey			
	ECAL		ECAS		ECAL		ECAS	
	Estimate	Std dev	Estimate	Std dev	Estimate	Std dev	Estimate	Std dev
Heckman 5%	0.057	(0.011)**	0.022	(0.011)*	0.047	(0.094)	0.009	(0.094)
Heckman 10%	0.053	(0.011)**	0.021	(0.011)	0.042	(0.094)	0.008	(0.095)
Andrews & Schafgans	0.049	(0.012)**	0.021	(0.011)	0.037	(0.094)	0.008	(0.095)

Bootstrapped standard errors (r=250) in parentheses. */** = significant at 5%/1%.

6 Conclusion

In this investigation we have estimated the effect of two different types of extracurricular activities on wages close after graduation. Both parametric and semiparametric estimation techniques were employed. The parametric results indicate considerable heterogeneity in the estimated effects across type of extracurricular activity, level of schooling and field of study. Extracurricular activities not relating to the study directly, predominantly have a modest positive effect on wages, whereas study related activities have no or even a negative effect. The estimation results in the case of the semiparametric estimation methods, are remarkably similar both in the direction of the effects and the significance. However, for the parameters of interest, the wage effect of extracurricular activities, virtually all significance is lost if we distinguish level and field of study. The parameter estimates also show a lot of heterogeneity across fields of study but due to the insignificance it is impossible to draw real conclusions. Also note that although most parameter estimates of the parametric and semiparametric estimations are quite similar, this does not hold for the estimates of the wage effect of extracurricular activities. Again, this conclusion is not substantiated by the significance levels. The reduced significance of the semiparametric estimation results, in particular with respect to extracurricular activities, is not likely to be due a too small number of observations. We employ about 56000 observations but this is either a too small number of find significance in estimation the selectivity model or there is simply no effect. Given that there are numerous studies employing significant estimation methods using fewer observations but still finding strong significance of the parameters of interest, for the time being we conclude that there is no strong wage effect of extracurricular activities employed by students at the higher level of tertiary education. The significance of these effects in the parametric case is probably the consequence of imposing too restrictive distributional assumptions.

As in many empirical economic investigations strong assumptions on the absence of endogeneity had to be made. We allow for the endogeneity of the variables representing the main objective of the analysis, but the choice for educational level and field of study were assumed to be exogenous. Separate models for different levels of education and fields of study were estimated. For the parametric case many effects remained significant, although sign changes occurred in some instances. Although separate estimation solves the endogeneity to some extent, it is not a complete solution especially if we want to estimate homogeneous effects. The evidence we do have suggest considerable heterogeneity, so imposition of homogeneity might not a good way to proceed even if we can overcome the estimation problems caused by allowing for the endogeneity of level of schooling of field of study. Adding this complexity to the model is a daunting task, however.

Table 5: Estimation effects of ECAL and ECAS differentiated for HBO and University

	HBO				University			
	ECAL		ECAS		ECAL		ECAS	
Parametric	0.031	(0.005)**	-0.036	(0.012)**	0.037	(0.004)**	-0.011	(0.006)
Semiparametric: Heckman 5%	0.019	(0.057)	0.002	(0.036)	0.057	(0.036)	-0.003	(0.037)
Semiparametric: Heckman 10%	0.019	(0.056)	-0.011	(0.036)	0.042	(0.036)	0.005	(0.037)
Semiparametric: Andrews & Schafgans	0.016	(0.016)	-0.007	(0.037)	0.043	(0.036)	0.006	(0.036)

Semiparametric estimation: osslett estimation method. Bootstrapped standaard errors (r=250) in parentheses. */** = significant at 5%/1%.

Table 6: Estimation effects of ECAL and ECAS differentiated for field of study

	Parametric estimation											
	Total				HBO				University			
	ECAL		ECAS		ECAL		ECAS		ECAL		ECAS	
α	0.018	(0.012)	0.033	(0.031)	0.032	(0.047)	-0.083	(0.073)	0.006	(0.011)	0.094	(0.028)**
β	0.045	(0.005)**	-0.049	(0.011)**	0.104	(0.010)**	-0.194	(0.027)**	0.035	(0.006)**	-0.020	(0.013)
γ	0.037	(0.004)**	-0.091	(0.013)**	0.046	(0.007)**	-0.218	(0.021)**	0.040	(0.005)**	-0.033	(0.015)*
Health	0.021	(0.010)*	0.028	(0.018)	0.016	(0.012)	0.068	(0.042)	0.013	(0.014)	0.124	(0.026)**
Remaining	0.041	(0.005)**	-0.025	(0.008)**	-0.062	(0.009)**	0.110	(0.018)**	0.041	(0.007)**	-0.015	(0.009)
	Semiparametric (Cosslett) estimation											
	Total				HBO				University			
	ECAL		ECAS		ECAL		ECAS		ECAL		ECAS	
α	0.207	(0.114)	0.170	(0.114)	0.058	(0.185)	-0.024	(0.191)	0.034	(0.103)	-0.004	(0.102)
β	0.047	(0.029)	0.021	(0.031)	0.030	(0.054)	0.040	(0.055)	-0.050	(0.072)	-0.068	(0.071)
γ	0.011	(0.023)	0.005	(0.024)	-0.081	(-0.082)	-0.084	(0.083)	-0.054	(0.056)	-0.054	(0.057)
Health	-0.019	(0.064)	-0.027	(0.065)	-0.035	(0.174)	-0.035	(0.174)	-0.251	(0.199)	-0.321	(0.198)
Remaining	0.058	(0.024)*	0.023	(0.025)	0.033	(0.236)	0.014	(0.237)	-0.012	(0.080)	-0.054	(0.079)

Semiparametric results based on Andrews & Schafgans-method. Bootstrapped standard errors (r=250) in parentheses. */** = significant at 5%/1%.

References

Ackerman, E., 2008. Hoe belangrijk zijn extracurriculaire activiteiten?, url: www.intermediair..nl, accessed: October 19th 2015.

Andrews, A., Schafgans, M., 1998. Semiparametric estimation of the intercept of a sample selection model. Review of Economic Studies 65, 497–517.

ASVA, 2010. De actieve student: extra-curriculaire activiteiten van studenten. Tech. rep.

Ayer, M., Brunk, H., Ewing, G., Reid, W., Silverman, E., 1955. An empirical distribution function for sampling with incomplete information. The Annals of Mathematical Statistics 26(4), 641–647.

Brown, P., Hesketh, A., WILIAMS, S., 2003. Employability in a knowledge-driven economy. Journal of Education and Work 16 (2), 107–126.

Cameron, A., Trivedi, P., 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

CBS, 2015. Leerlingen, deelnemers en studenten; onderwijssoort, vanaf 1900, url: statline.cbs.nl, accessed: October 21st 2015.

Collins, R., 2002. The future of the city of intelellect: the changing American university. Stanford University Press, Ch. Chapter 1: Credential inflation and the future of universities, pp. 23–46.

Cooper, H., Valentine, J. C., Nye, B., Lindsay, J. J., 1999. Relationships between five after-school activities and academic achievement. Journal of Educational Psychology 91 (2), 369.

Cosslett, S., 1983. Distribution-free maximum likelihood estimator of the binary choice model. Econometrica 51, 765–782.

Cosslett, S., 1991. Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, Ch. 7: Semiparametric estimation of a regression model with sample selectivity, pp. 175–197.

DesJardins, S. L., Dundar, H., Hendel, D. D., 1999. Modelling the college application decision process in a land-grant university. Economics of Education Review 18 (1), 117–132.

Eccles, J. S., Barber, B. L., 1999. Student council, volunteering, basketball, or marching band what kind of extracurricular involvement matters? Journal of Adolescent Research 14 (1), 10–43.

Eide, E., Roman, N., 2001. Is participation in high school athletics an investment or a consumption good? Economics of Education Review 20, 431–442.

Finn, J. D., Voelkl, K. E., 1993. School characteristics related to student engagement. Journal of Negro Education, 249–268.

Gallant, A., Nychka, D., 1987. Semi-nonparametric maximum likelihood estimation. Econometrica 55, 363–390.

Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Heckman, J., 1990. Varieties of selection bias. The American Economic Review 80, 313–318.

Holland, A., Andre, T., 1987. Participation in extracurricular activities in secondary school: What is known, what needs to be known? Review of Educational Research 57 (4), 437–466.

Hussinger, K., 2008. R&D and subsidies at the firm level: an application of parametric and semiparametric two-step selection models. Journal of Applied Econometrics 23, 729–747.

Kuhn, P., Weinberger, C., 2005. Leadership skills and wages. Journal of Labor Economics 23 (3), 395–436.

LKvK, 2012. Wat is de daadwerkelijke meerwaarde van het doen van extracurriculaire activiteiten? Tech. rep.

Mahoney, J. L., Cairns, R. B., 1997. Do extracurricular activities protect against early school dropout? Developmental Psychology 33 (2), 241.

Marks, H. M., 2000. Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. American Educational Research Journal 37 (1), 153–184.

Massoni, E., 2011. Positive effects of extra curricular activities on students. ESSAI 9 (1), 27.

Ming Chia, Y., 2005. Job offers of multi-national accounting firms: The effects of emotional intelligence, extra-curricular activities, and academic performance. Accounting Education 14 (1), 75–93.

Newey, W., 2009. Two-step series estimation of sample selection models. The Econometrics Journal 12, S217–S229.

Pudney, S., 1989. Modelling Individual Choice: the Econometrics of Corners, Kinks and Holes. Blackwell.

SEO, 2015. Studie & Werk, Research Report SEO.

Tchibozo, G., 2007. Extra-curricular activity and the transition from higher education to work: a survey of graduates in the United Kingdom. Higher Education Quarterly 61 (1), 37–56.

Tomlinson, M., 2007. Graduate employability and student attitudes and orientations to the labour market. Journal of Education and Work 20 (4), 285–304.

Tucker, I. B., 2004. A reexamination of the effect of big-time football and basketball success on graduation rates and alumni giving rates. Economics of Education Review 23 (6), 655–661.

Valentine, J. C., Cooper, H., Bettencourt, B. A., DuBois, D. L., 2002. Out-of-school activities and academic achievement: The mediating role of self-beliefs. Educational Psychologist 37 (4), 245–256.

Van der Klauw, B., Koning, R., 2003. Testing the normality assumption in the sample selection model with an application to travel demand. Journal of Business & Economic Statistics 21, 31–42.

VSNU, 2012. Prestaties in perspectief, www.vsnu.nl, accessed: October 21st 2015.

Willems, E. P., 1967. Sense of obligation to high school activities as related to school size and marginality of student. Child Development, 1247–1260.