Breaking the Cycle?
The Effect of Education on Welfare Receipt Among Children of Welfare Recipients

Michael B. Coelli*, David A. Green** and William P. Warburton***

October 2004

We examine the impact of high school graduation on the probability individuals from welfare backgrounds use welfare themselves. Our data consists of administrative educational records for grade 12 students in a Canadian province linked with their own and their parents' welfare records. We address potential endogeneity problems by: 1) controlling for ability using past test scores; 2) using an instrument for graduation based on school principal fixed effects; and 3) using a Heckman-Singer type unobserved heterogeneity estimator. Graduation would reduce welfare receipt of drop-outs by ½ to 3/4. Effects are larger for individuals from troubled family backgrounds and low income neighbourhoods.

Corresponding author: David Green, Department of Economics, University of British Columbia, #997-1873 East Mall, Vancouver, BC, Canada V6T 1Z1, green@econ.ubc.ca

JEL codes: I38, I21

Key words: Welfare, Education, Treatment Effects

* Department of Economics, University of British Columbia
** Department of Economics, University of British Columbia, and Research Associate, Institute for Fiscal Studies, UCL
*** Ministry of Human Resources, British Columbia
We are grateful for comment from participants at seminars at the 2003 Canadian Economics Association meetings, Western Research Network on Education and Training, the University of British Columbia, and the Institute for Fiscal Studies, UCL.

Breaking the Cycle?
The Effect of Education on Welfare Receipt Among Children of Welfare Recipients

Michael B.Coelli, David A. Green and William P. Warburton

Welfare systems in a number of developed countries have undergone reforms in recent years, with some of the reforms being extensive. A key focus in many reforms has been the desire to break “cycles of dependency” in which welfare receipt at one point in time is believed to generate higher probabilities of receipt in the future. Of greatest concern are inter-generational welfare traps where welfare receipt by parents may lead to increased welfare dependency among children once they grow up. In several countries, empirical work has shown a strong positive correlation between welfare rates of successive generations within families. Whether this correlation reflects a causal impact of parental welfare use on children’s behavioural patterns or simply correlations in unobserved characteristics common to the parents and children has been the focus of some debate. Several, though not all, investigations of this issue for the US point to the conclusion that a substantial part of the correlation is causal. This result has led some to advocate a policy of reducing access to welfare for current adults in part in order to reduce dependency of future generations. We are interested in investigating an alternative policy response: breaking the cycle by improving the educational outcomes of children of welfare recipients. Thus, we examine the effects of increasing the education levels of children from welfare households on their future welfare use[1].

We study the impact of high school graduation on the probability individuals from welfare backgrounds use welfare themselves at each age between 19 and 24. Our data is a unique linking of educational records from all individuals entering grade 12 in the Canadian province of British Columbia (BC) between 1991 and 1996 with both their own and their parents’ welfare records. The education data includes information on whether the individual graduated from high school (i.e., completed grade 12), whether he or she had failed any earlier grades, test scores from exams in grade 11, and characteristics of the school. We identify children from welfare backgrounds as students from households which received welfare income at any time before the child turned 17. For the children themselves, we know whether they received welfare at each age from when they first become eligible for welfare (age 19) to the end of our data period in 1999.

The main difficulty in establishing the impact of education on welfare use, as in studies of educational impacts on other outcomes, is determining the extent to which observed correlations reflect causal impacts. In this case, it is reasonable to assume that there exist individual characteristics that lead to both better educational outcomes and lower use of welfare. For example, a more responsible person could both have better educational outcomes and a better ability to keep a job, which in turn would be associated with less time on welfare. In this case, we could observe individuals with higher education having less time on welfare even if improving educational outcomes would not reduce future welfare use for any individual.

We address the problem of establishing the causal impact of high school graduation in three ways. First, we include an extensive set of controls in our analysis, including whether the province had to intervene in the household to protect the child, whether the child failed any earlier grades, and his or her grade 11 exam results. To the extent that these controls capture all the individual characteristics associated with both education and welfare outcomes, any remaining estimated correlation between completing high school and welfare outcomes reflects a causal impact of education in reducing future welfare use. Second, we use an estimator described in Aakvik, Heckman and Vytlacil (2000) in which we allow for an unobservable factor which may affect both high school graduation and welfare outcomes. This estimator, in principle, generates consistent estimates on its own because it directly incorporates a common unobservable factor, but we implement this estimator using exclusion restrictions that form our third (instrumental variables) approach to the identification problem. To form our instrument, we use the fact that we know the name of the principal at the high school the child attends in grade 12. We show that there is considerable variation in graduation rates across different principals, even after controlling for the average, over-time graduation rate at a school. Further, high school principals are constantly shuffled in the BC system, staying at a given school for no more than a few years. Because we observe high school graduations over a 6 year period, we have variation in principal effects over and above cross-school variation in graduation rates. We use these principal effects to instrument for high school graduation, under the assumption that a good principal can alter high school drop out rates but would not directly affect welfare outcomes.

We find that there are substantial differences in welfare (called Income Assistance (IA) in BC) use between high school graduates and drop outs from IA families. For females, drop outs have IA usage rates of over .2 while graduates have rates under .1. Our estimates indicate that there is substantial heterogeneity in the actual causal impact of graduation on IA use and, as a result, we obtain different average treatment effects for different groups. For females, among drop-outs, graduation would cause a substantial reduction in IA use (on the order of between .1 and .15), with effects rising with age and the magnitude of the impact depending on how we control for heterogeneity. For males, impacts are of the same order of magnitude and also rise with age. Causal effects for graduates appear to be much larger, indicating that it is those who get the best return from graduating who actually do so. The most substantial impacts of graduation are for individuals from low income neighbourhoods and from families in which the province intervened at some point for child protection reasons. Indeed, one of the clearest results from this exercise is that children in these "at risk" households have particularly poor educational and IA outcomes and are particularly likely to benefit from efforts to improve their education levels. Finally, differences in unobserved heterogeneity between drop-outs and graduates play a substantial role in observed raw differences in IA use between drop-outs and graduates for females while our observed heterogeneity controls play a larger role for males.

The paper proceeds in 7 sections. In section 1), we provide a brief description of the welfare and educational systems in BC. In section 2), we discuss the previous literature related to our investigation. In section 3), we describe the data and provide a characterization of basic patterns. In section 4), we set out our estimation framework. Section 5) contains results from our estimation, section 6) defines and presents relevant treatment effects, and section 7) concludes.

1) Income Assistance and Education in British Columbia

Income Assistance (IA), BC's form of welfare, is part of a system of social security programmes.[2] Unemployed individuals who meet employment related eligibility requirements can collect benefits from a federally administered unemployment insurance system for defined durations. In contrast, IA is provincially run and covers all individuals based on need for as long as the needs based requirements are met. Applicants for IA are first sorted into employable and non-employable groups[3], then their income levels are compared to a cut-off level specified for their employability status. Benefits are granted if the applicant's resources are below the cut-off. The benefit amounts are determined by employability status and family size, and are paid monthly. Benefits are deemed to be paid to a whole family, thus there is no need to distinguish which parent received IA when we establish whether a child comes from an IA background.

IA is very different from welfare systems in US states. In particular, while most US welfare systems focus on lone parent families, IA is open to all individuals and families. Thus, in 1992, single, employable adults and childless couples made up 68% of beneficiaries in BC (Barrett and Cragg (1998)). In fact, IA beneficiaries made up 9.7% of BC's population in December 1995. On the other hand, IA resembles basic assistance systems in several European countries which also have universal, needs based social assistance systems backed by a separate unemployment insurance system. Thus, while our results are difficult to apply directly to the US experience, they may be useful in considering outcomes in other developed economies.

BC's schooling system is largely publicly funded, with only 7.5% of grade 12 students attending private schools in 2002. Graduating from high school means completing grade 12, and the legal school leaving age is 16, which for most students occurs in grade 11. There is an equivalent to the US GED in BC but it is very rarely used. Thus, records of graduation from high schools provide quite a complete record of high school educational outcomes for residents of BC.

It is worth emphasizing that BC is still a resource dominated economy where many students leave high school before graduating in order to take up well paying jobs in the resource sector. Thus, impacts of graduating might be expected to be small, particularly at young ages where drop outs may well be high ability individuals in good jobs. If the drop-outs lose those jobs, however, the lack of education may have a greater impact on them and thus we expect educational impacts on welfare use to grow with age.

2) Previous Literature

Our study is at least indirectly related to a number of literatures studying welfare and educational effects. Our motivation for studying this topic stems from the literature on the inter-generational transmission of welfare use. Several studies have established a strong positive correlation in welfare use across generations within a family for the US (e.g., Antel (1992), An et. al. (1993), Gottschalk (1990, 1996), and Levine and Zimmerman (1996)). The issue of whether that correlation reflects a causal transmission across generations or simply persistence in unobserved traits within families is well recognized and the focus of attention in several of these papers (e.g. Levine and Zimmerman (1996) and Gottschalk (1996)).

Beaulieu et al (2001) is the only paper of which we are aware that examines these issues for Canada, using administrative data for Quebec spanning the period from 1979 to 1995. As in Gottschalk (1996), they obtain identification of causal effects based on the assumption that only parental welfare use that occurs before the child achieves adulthood can have a causal impact on the welfare use of their children. As in the US data, they observe strong positive correlations between parental and child welfare use. Their results indicate that this correlation reflects both common unobserved heterogeneity and a causal dependence effect.

If some, possibly large, component of the cross-generational correlation in welfare use is causal, how is it transmitted? One possible channel is through impacts on educational outcomes. Boggess (1998) and Ku and Plotnick (2000) both find significant negative effects of the family being on welfare on the educational attainment of the children for some groups. Impacts on education may also be related to pregnancy effects. MaCurdy (2000) finds that higher welfare benefits cause a delay in high school completion once an out-of-wedlock birth has occurred.

Of course, the impact of parental welfare receipt on educational outcomes is only half the story. For this to be a route of transmission of welfare receipt between generations, it must also be the case that poorer educational outcomes induce higher recipiency rates. Several studies include educational variables among a set of controls in examining spells of welfare receipt. Blank (1989), for example, using US data, finds that an extra year of education increases the exit rate from welfare for female household heads by between 5 and 8%. Similarly O'Neil et. al. (1987) finds an extra year of education increases the exit rate by 12%. Barrett (2000) examines this issue using data from the Canadian province of New Brunswick and finds that males with less than a high school education had exit rates from a welfare spell that were over 6% lower than male high school graduates. However, males who had some high school education but did not graduate had exit rates that were not different from graduates. For females, having less than a high school education decreased exit rates by over 18% and being a high school drop-out decreased exit rates by 9% relative to a high school graduate. Thus, females had much stronger education effects.

There are two main difficulties with earlier attempts to estimate educational impacts on welfare receipt. First, the specified studies do not directly address endogeneity issues related to the education variables. In particular, it seems quite possible that individual, unobserved characteristics that increase success in education will also be associated with a reduced likelihood of using welfare. If that is true then one could observe the type of significant negative coefficients on education variables just listed even if no individual would change his or her welfare related behaviour in response to extra years of schooling. In principle, studies such as Blank (1989) and Barrett (2000) address the issue through their use of methods to address unobserved heterogeneity. As described in Aakvik et. al. (2000), one can view these estimators as matching estimators in which a key matching variable is unobserved and therefore has to be integrated out. However, identification of the distribution of the unobserved factor is essentially achieved through functional form restrictions, which is not very satisfactory. While we use this approach, we also use a complete set of controls and a key exclusion restriction to achieve identification of the education effect.

The second issue of concern stems from potential heterogeneity in education effects. Estimation of educational impacts on earnings has typically involved instrumental variables estimators. However, different instruments will highlight educational impacts for different parts of the population and if there are differences in responses to education across these groups then different estimates will emerge (Card (1999)). In our case, we are interested in the responsiveness of a particular subset of society: individuals who were exposed to welfare as children. This may differ greatly from the estimates for individuals from all backgrounds obtained in other studies. In addition, we are interested in whether education can keep children of welfare recipients off welfare altogether while the earlier literature just focusses on effects on spell lengths for people who are already in receipt.

3) Data

The data we use in our investigation is compiled from a linkage of three data sources. The first source is BC Ministry of Education records on grade 12 students for the years 1991 through 1996. These records contain information on all individuals enrolled in grade 12 in a BC high school at the start of November of the given year. For each student we know whether they failed a previous year of school and whether they graduate from high school. We also know the high school the individual attended. From this we can identify the principal at the school when the student is in grade 12 and can also use the records to calculate historical graduation rates for the school. Finally, we know the individual's grades in up to five courses in grade 11 and, if they complete grade 12, their final high school Grade Point Average (GPA), which is a weighted average of their course marks. These are important components in our attempt to identify the causal link between graduation and IA recipiency. We keep only observations on individuals who either graduate from high school by age 19.5 or never graduate in our sample period. This gives individuals on average about a year to complete high school even if they did not complete it on the regular schedule with their cohort. Few drop-outs complete high school after age 19.5 in BC and attempts to include them in the analysis changed reduced form data patterns very little but were costly in terms of the analytical complexity required to address their special patterns.

We link the data for the individuals in the high school to the 1996 Census tract records through the location of the high school attended in grade 12. With this link, we can use Census data on characteristics of the population in the area surrounding the school to control for neighbourhood effects. Since we do not have direct information on the income or education level of the individuals' families, the Census tract data provides an indirect means of controlling for these types of effects on both the graduation and IA use outcomes. It should be noted, though, that our focus on families that receive IA already effectively constitutes a substantial control for family income.

We also link the high school data to administrative data from the IA programme. The link is done by matching name and birth date between the two sets of files. Given a successful link, we have access to IA histories for the individual up to May 1999. We use that to create a series of dummy variables equalling one if the individual was in receipt of IA at ages 19.5, 20.5, 21.5, 22.5, 23.5 and 24.5. We start at age 19 because individuals younger than 19 are not eligible to collect IA in their own right. We observe IA use for the full set of ages only for the 1991 and 1992 high school cohorts. For the subsequent cohorts we are restricted in the ages observed by the end date of our data. We also know whether the individual came into contact with the IA system as a child (i.e., at any time before grade 12) as a member of a family receiving IA payments, and whether they came into contact with other parts of the child welfare system. We focus on individuals from IA recipient families but have data on all high school students in BC for the years listed above and use a 10% random sample from the full dataset to generate some comparative tabulations. The non-IA family data contains 16,425 observations while the IA family dataset contains 16,806 observations, of which 9272 are female and 7534 are male. We stratify all our results by gender.

The main difficulty with this data is associated with geographic mobility. If an individual enters a BC high school in grade 12 but then migrates out of the province, we do not know about the migration and will count him or her as a non-IA user because s/he will never appear in the BC IA administrative data. If migrants are randomly selected then this does not cause any problems. However, if, as seems likely, graduates are more likely to migrate then we may over-estimate the effect of graduation on welfare use. Whether we actually over-estimate depends on whether migrants would have used IA if they had remained in BC. If migrants are self-selected for positive qualities then there may be little or no distortion from the fact they leave the province. In any case, these potential selection effects should be kept in mind when examining our results.

Table 1 contains IA recipiency rates at age 19.5 broken down by gender and family IA history. The table supports findings in other papers that there is a strong inter-generational correlation in welfare use. In particular, a female drop-out from a family that received IA before she entered grade 12 has a probability of receiving IA at age 19.5 that is nearly three times that for a female from a non-IA family. Among high school graduates, the difference is even greater: female graduates from non-IA families have only a .019 probability of receiving IA at age 19.5 while graduates with IA backgrounds have a .094 probability. Males have uniformly lower IA probabilities but show IA family effects that are of a similar proportional size as for females.

Table 1 also shows very clear effects related to high school graduation. Graduating from high school rather than dropping out in grade 12 reduces IA probabilities by almost .15 and almost .10 for females and males from IA families, respectively. The effects for individuals from non-IA backgrounds are also substantial but much smaller in absolute size than those for IA family students.

Thus, increasing high school graduation rates may be an avenue for lifting children of IA recipients out of patterns that result in their using IA as well.

As we discussed earlier, we are interested in whether improvements in educational outcomes can alter IA outcomes for children from IA backgrounds. One reason for focussing on educational outcomes may be because coming from an IA background itself has adverse effects on educational attainment. In that case, educational policies can help reverse these negative effects. Table 2 shows graduation rates by gender and IA family background. The results from the US literature on family welfare background are mimicked here: for both men and women, being from an IA family is associated with graduation rates that are over 13% lower.

As stated earlier, our data allows us to investigate IA receipt at every age between 19.5 and 24.5. The dashed line in Figure 1 plots the simple difference between the IA usage rate for drop-outs and the IA rate for graduates at every age for IA family females from the cohort entering grade 12 in 1991. Thus, this figure follows the age profile of graduation effects for one cohort. The effects start at approximately .17 at age 19.5, rise to .19 at age 22.5 and then fall to .16 at 24.5. The solid line repeats this exercise for males. Here the pattern is much flatter, with the difference staying at values of .11 across most of the age range. Again, though, there is some tendency for the graduation effect to grow just after eligibility is attained and to decrease slightly at older ages. While it is tempting to speculate about life-cycle patterns of IA use based on these figures, the fact that only one cohort is followed in the figures means that we cannot separate life-cycle effects from the effects of changing macro conditions. In the estimation that follows we will attempt to address this.

4) Estimation Framework

4.1) The Identification Problem and Solutions

The identification issues we seek to address can best be explained in the context of a model of the joint determination of high school graduation and future IA recipiency status. The approach we take is similar to that specified in Aakvik, Heckman and Vytlacil (2000) and builds on Heckman

(1981) and, to some extent, Heckman and Singer (1984).

We begin by specifying the process determining high school graduation. One way to do this is to consider the members of our sample (students in school in November of their grade 12 year) as performing a utility maximization exercise. Thus, individuals consider the effort required and future expected utility following from either graduating from grade 12 or dropping out. The students graduate if they both stay in school and their level of accomplishment surpasses some threshold. Their level of accomplishment (i.e. their observable outcomes such as test scores) is a function of their effort. The amount of effort needed to achieve a given level of accomplishment is a function of the student's individual ability, family emotional support, family financial or resource support, and the resources provided by the school. Providing effort negatively affects the student's utility and this is part of the cost of choosing to graduate. Against this is weighed the expected future utility stream stemming from labour market outcomes arising from completing high school as well as any emotional returns from family and friends. If the individual drops out of high school, s/he expends effort in searching for and then working at a job (though s/he may choose to put little effort into either or both). The labour market outcomes available to drop-outs will determine a stream of expected future utility. The individual will either graduate or drop out from school depending on which option generates the highest net utility (the discounted future utility stream minus the disutility from current effort). For both options, future income streams may include periods of IA receipt, and part of the return to graduating high school may involve better access to jobs and therefore reduced need for IA.

We represent the choice of graduating from high school with an index function corresponding to the difference between the net utility from graduating and dropping out,

1) $$I_{it}^{H} = x_{it}\alpha + u_{it}$$

where: individual i is observed to graduate in year t (represented by $D_{it}^{H} = 1$) if $I_{it}^{H} > 0$ and is observed to drop out ($D_{it}^{H} = 0$) if $I_{it}^{H} \leq 0$; α is a parameter vector; u_{it} is an error term, the properties of which we discuss in detail below; and x_{it} is a vector of observable characteristics determining graduation, including factors relating to returns to effort in school and costs to that effort (e.g., parental resource inputs, family environment, school resource inputs and approaches, and neighbourhood environment), the state of the local labour market, and IA system parameters.

The second process concerning us is the one determining IA receipt in each year after the individual becomes eligible for IA. To keep the model simple, assume that an individual moves onto IA if her financial resources pass below a specified threshold and the perceived social costs to taking up IA do not exceed the monetary benefits. Assume further that her financial resources are a function of her employment status and the wage she can earn in the labour market. For younger individuals (even those old enough to qualify for IA benefits in their own right), parental resources are probably also relevant. That is, if the individual is out of work and has no saving from which to support herself, she may receive financial support from her parents before applying for IA. Thus, the resources available to an individual are a function of her education level, accumulated labour market experience, the state of the local labour market, and parental resources. The IA process can be summarized in the equations:

$$2a) \quad I^{I,G}_{it\tau} = z_{it\tau}\beta_{G,\tau} + e^{G}_{it\tau}$$

$$2b) \quad I^{I,D}_{it\tau} = z_{it\tau}\beta_{D,\tau} + e^{D}_{it\tau}$$

where the function in 2a) pertains to IA outcomes if the individual is a graduate and the function in 2b) is relevant if she is a drop-out. Thus, if individual i is a graduate from potential graduating year t then she is observed to be receiving IA at age τ (represented by $D_{it\tau}^{I,G} = 1$) if $I_{it\tau}^{I,G} > 0$ and is observed not to be in receipt ($D_{it\tau}^{I,G} = 0$) if $I_{it\tau}^{I,G} \leq 0$. The dummy variables corresponding to IA receipt if the individual drops out ($D_{it\tau}^{I,D}$) are defined analogously. Both $\beta_{G,\tau}$ and $\beta_{D,\tau}$ are parameter vectors that are allowed to vary with age, and $e_{it\tau}^{G}$ and $e_{it\tau}^{D}$ are disturbance terms the properties of which are discussed below. The z vector includes the factors specified earlier for the x vector, plus factors affecting individual decisions on whether to take up IA, including previous contact with the system. This specification allows for high school graduation to change the process determining IA receipt completely and corresponds to permitting heterogeneity in returns to education.

As the model is specified to this point, the equations given by 1), 2a) and 2b) form a recursive system: education outcomes affect future IA outcomes but the reverse is not true. However, it is unlikely that we observe all relevant factors in these two sets of decisions. A common assumption in the education-earnings literature is that there is a person-specific, time-invariant factor (often described as "ability") that affects both education and future labour market outcomes. In our case, we can represent this by specifying the error structure as follows:

$$3a) \quad u_{it} = \omega_i + \eta_{it}$$

$$3b) \quad e^G_{it\tau} = \omega_i \lambda_{G,\tau} + \epsilon^G_{it\tau}$$

$$3c) \quad e^D_{it\tau} = \omega_i \lambda_{D,\tau} + \epsilon^D_{it\tau}$$

where: ω_i is a person-specific, time-invariant factor; η_{it}, $\epsilon_{it\tau}{}^G$, and $\epsilon_{it\tau}{}^D$ are error terms independent across time and age, and with respect to one another; and the λ_G's and λ_D's are parameters. The λ's capture potential correlations among the error terms in equations 1) and 2).

Finally, in the estimation that follows, we have observations on variables we interpret as proxies for ω_i, including individual test scores from grade 11. We bring those into the analysis through the following specification:

$$4) \quad \omega_i = a_i \gamma + \theta_i$$

where, γ is a parameter vector , a_i is a vector of observable proxies, and θ_i is an unobservable component of ability. In the tradition of standard proxy approaches, we interpret γ as capturing the projection of ω_i onto the space spanned by a_i, and θ_i as an orthogonal component of ω_i. We do not give this equation a behavioural interpretation.

Given this structure, if we have consistent estimates of $\beta_{G,\tau}$, $\beta_{D,\tau}$, γ , $\lambda_{G,\tau}$ and $\lambda_{D,\tau}$ as well as estimates of parameters defining the distribution of θ_i then we can generate estimates of treatment effects related to specific policy and other economic questions. We provide more detail on the treatment effects we are interested in and how to construct them in section 6. One might consider obtaining estimates of these parameters from simple probit estimation of equations 2a) and 2b). However, if the λ_G's and λ_D's are non-zero then simple estimation of 2a) and 2b) will not yield consistent estimates of the β_G's and β_D's, and thus of the impact of graduation on IA receipt. The difficulty arises because graduates and drop-outs will be systematically different with respect to ω_i, and because ω_i also affects IA outcomes. As a result, any observed differences in IA receipt between graduates and drop outs may reflect differences in ω_i rather than causal effects of high school graduation. Indeed, one could estimate differences in IA outcomes between graduates and drop-outs even if graduation had no causal effect on future IA receipt.

We adopt three different approaches to this standard identification problem. The first is to use proxies for ω_i . Those proxies are the elements of the a_i vector in 4). Introducing a_i will solve the endogeneity problem if $\theta_i = 0$, otherwise θ_i remains in the error terms of both the high school and IA equations, creating problems. Note that ω_i corresponds to the part of ability that is common across the high school and graduation equations. The η_{it}, $\epsilon_{it\tau}{}^G$, and $\epsilon_{it\tau}{}^D$ terms may contain types of ability that are specific to the outcome corresponding to those error terms. Thus, the requirement that $\theta_i = 0$ is the same as the standard requirement for a good proxy: that it captures the part of the variation in the omitted variable that is common to both the dependent variable and any right hand side variables. If we have a good proxy, we can estimate 2a) and 2b) as standard probits with the elements of a_i included as regressors and obtain consistent estimates of the key parameters.

Our second approach is to find an instrument for high school graduation: a variable that is in x_i but does not belong in $z_{i\tau}$. With a structure where there is heterogeneity across individuals in impacts, different instruments emphasize different parts of that heterogeneity distribution and generate different average effect estimates (Heckman, Lalonde and Smith (1999)). The third approach, which is less common, is to address the problem with ω_i econometrically, as an unobservable, omitted factor. In this approach, a likelihood function is specified in which individual contributions to the likelihood are conditioned on ω_i. The model is completed by integrating over the distribution of ω_i, the parameters of which form part of the estimation problem. As in the first approach, the goal is to condition on the omitted factor. Aakvik et. al. (2000) use this approach to study vocational rehabilitation programmes in Norway and argue that this can be viewed as a matching estimator where ω_i is the key matching variable but is unobserved. In principle, this approach does not require exclusion restrictions to be identified given the error structure set out in 3a) - 3c). However, our main estimation combines approaches 1, 2 and 3, implementing the Aakvik et. al. (2000) type estimator but also using an instrument for education as well as proxies for unobserved heterogeneity. Essentially, we take an approach that allows for the possibility that $\theta_i \neq 0$ in equation 4) and thus that we need to use other means to fully address endogeneity.

4.2) Implementing the Estimator

We implement an estimator based on equations 1) - 4) and the assumption that the error terms η_{it}, $\epsilon_{it\tau}{}^{G}$, and $\epsilon_{it\tau}{}^{D}$ are independently and normally distributed. Thus, conditional upon θ_i, we can specify our model as a simple set of Probits: one corresponding to whether the individual graduates from grade 12 and separate Probits corresponding to IA receipt for either graduates or drop-outs (depending on which the individual is) for each year the individual is both age 19 or more and in our sample. An individual's contribution to the likelihood function consists of the set of Probit contribution values relevant for her, calculated for a specific value of θ_i. For example, for an individual who graduates from high school, stays in the sample until age 21.5, and uses IA at age 19.5 but not in subsequent years, the contribution to the likelihood function for a specific value, θ^*, is given by:

$$6)\quad l_i(\theta^*) = Pr(\eta_{it} > -x_{it}\alpha - a_i\gamma - \theta^*) * Pr(\epsilon^G_{it19} > -z_{it19}\beta_{G,19} - a_i\delta_{G,19} - \theta^*\lambda_{G,19}) * Pr(\epsilon^G_{it20} \le -z_{it20}\beta_{G,20} - a_i\delta_{G,20} - \theta^*\lambda_{G,20}) * Pr(\epsilon^G_{it21} \le -z_{it21}\beta_{G,21} - a_i\delta_{G,21} - \theta^*\lambda_{G,21})$$

We then multiply this contribution by the probability associated with θ^*, repeat the exercise for all different values of θ_i and then integrate across the full range of θ_i. In practice, we use the type of approach described in Heckman and Singer (1984), where we specify the distribution of θ_i non-parametrically as a discrete distribution with a finite set of mass points and associated probabilities. We estimate the locations of those mass points and the associated probabilities.[4] We do not impose the cross-equation restrictions linking the δ parameters (the coefficients on the a_i variables int eh IA processes) to the γ and λ parameters. The full likelihood function is specified in Appendix A.

Our estimator is similar in spirit to that in Carneiro et. al. (2003), except that they estimate an expanded system that includes equation 4) and they have more than one unobserved factor. In the end, this means that they estimate a distribution for ω, relying heavily on the observable indicator, *a*, while we estimate only the distribution for θ, which is the part of ω remaining after using *a* as a proxy. As in indicator function approaches in standard regression analysis, their approach requires instruments or multiple indicators to obtain consistency. Ours is a proxy approach in which we need instruments or other solutions to the extent that the proxy is imperfect. Note that these kinds of estimators can be interpreted in a more reduced form manner in which incorporating and estimating the distribution of θ_i is just a less parametric, flexible way of modelling the joint distribution of u_{it}, $e_{it\tau}^G$ and $e_{it\tau}^D$ since mixtures of normals can mimic other distributions.

The x_i vector includes variables intended to capture the factors contributing to high school graduation described earlier.[5] Unfortunately, we do not have direct observations on parental income or education. We approach this problem in two ways. First, all of our estimation is done for children from families which received IA at some time before the child entered grade 12, thus narrowing the families under observation to a group with low income. Second, we use characteristics of the neighbourhood where the individual lived in grade 12 related to education and income. These latter variables can also be seen as capturing neighbourhood attitudes toward education. The full set of elements of x_i are as follows. From Census data we create the neighbourhood average income level (from which we then construct a dummy variable corresponding to neighbourhoods with average income above the 75th percentile of the neighbourhood income distribution and another corresponding to neighbourhoods with average income below the 25th percentile), the proportion of families in the neighbourhood with a lone parent, the proportion of individuals with less than a grade 9 education, and the proportion of individuals who are not immigrants. We also include the employment rate for the neighbourhood to capture labour market opportunities. All of these variables correspond to June 1996 (the Census date). We also include individual level indicators for whether the individual was of Native ethnic origin and whether English was the first language spoken at home. Both of these variables were constructed from IA administrative records. Finally, we include a set of dummy variables corresponding to the potential graduation year in order to capture time variation in the provincial labour market, education policy and IA parameters that might affect graduation decisions.

We use two set of variables to proxy for ω_i. The first are variables reflecting whether the child came from a troubled household, on the assumption that coming from such a background has persistent negative effects on education and labour market outcomes through channels such as the child's ability to concentrate and maintain stable relationships. From provincial administrative records, we construct a dummy variable equalling one if the province had to come into a household to assess it for problems related to the child at any time before the child entered grade 12 (ASSESS), and a dummy variable equalling one if the assessment led to the province needing to provide counselling and other services (FAMSRV).

Our second set of proxy variables are built from the child's education records. The first is a dummy variable equalling one if the child failed any grade before grade 12 (PFAIL). The second, GR11F, is a fitted grade 11 grade average and requires some extra explanation. We have data for each individual on grades from a series of grade 11 exams. The number of exam grades we have in the dataset varies by individuals and there is no one grade (e.g., math) present for all individuals. Thus, we have to aggregate the grade data in a way that makes use of the varied individual data. We also adjust the fitted average grade to eliminate differential grade inflation across schools. The details on the construction of this variable are given in Appendix B.[6]

Our second identification approach involves specifying an instrument: a variable included in x but not in z. To fulfill this role, we introduce a set of dummy variables corresponding to each principal employed in a BC school in our sample period. The principals in the BC school system are purposefully shifted across schools every few years. Thus, in our sample period, 265 principals are employed and 65% of them change schools at least once in the period. The timing and location of the shifts are determined by the school boards and are not random. As a result, we need to separate principals' impacts on graduation from school specific effects. The latter might reflect problems in specific neighbourhoods that are persistent and could affect future IA outcomes, making them inappropriate as instruments. However, principal specific effects on graduation rates are unlikely to directly affect future IA outcomes. We take advantage of the switching of principals across schools to separate school effects from principals effects, including the average graduation rate for the high school calculated over the six years of our sample as a separate regressor in all equations.[7] Principal effects are then identified as the difference between the average graduation rates corresponding to a specific principal and the long term average rates at the schools at which s/he served.

We view this as a useful instrument because it captures variation that potentially reflects differences in policies that affect graduation. Thus, the average effect of high school graduation on future IA receipt we identify using this variable corresponds to the effect one would obtain from changing from the policies used by bad principals to the policies used by good principals. This seems to us to be a relevant type of variation to use in evaluating education outcomes in our policy context. Indeed, the education literature on effective schools places considerable weight on the impact of good principals on student outcomes (Raptis and Fleming(2002)). The 5th percentile of the distribution of principal-specific raw average graduation rates is .74 while the 95th percentile is .98. Thus, there is considerable variation in graduation rates across principals.

The z vector includes all the variables specified above except the principal dummies. We also include a set of dummy variables corresponding to the calendar year in which the observation on IA receipt is made. For these variables, we impose the restriction that the corresponding elements of $\beta_{G,\tau}$, for graduates, are the same for all values of τ. The analogous restriction is imposed on the $\beta_{D,\tau}$ vectors for drop-outs. Thus, we estimate year effects that are common across age groups, allowing us to capture labour market shifts and/or changes in the IA system. The coefficients corresponding to Native status are also imposed to be the same for all values of τ because of a lack of observations to identify this coefficient at separate ages.

A final concern in the IA processes is with dynamics. The inclusion of the θ_i factor implies persistence in IA outcomes across ages. However, we would also like to capture potential structural dependence. To do this, we include a single lag of the IA dummy variable in all the z vectors except z_{it19} (the vector relevant for the age 19.5 process). In some specifications, we also include a variable which takes a value of 1 if the individual has ever used IA in any previous year except in z_{it19} and z_{it20}. Thus, the latter variable is identified relative to the first by differences in outcomes between individuals who did not use IA in the previous year but did use it in the past and individuals who did not use IA in the previous year and did not use in the past. Note that we observe individuals from the time they first become eligible for IA, that is from the start of the IA determination process. We follow Heckman (1981) in treating the initial IA outcome by effectively giving it its own distribution, which is conditioned on θ. Because we control for θ_i and thus unobserved heterogeneity, the coefficients on the IA history variables can be interpreted as reflecting structural dependence in a similar manner to the way functions of duration can be interpreted when using a Heckman and Singer(1984) estimator with duration models. Recall that we assume that, conditioning on θ_i, the $\epsilon_{it\tau}^{G}$, and $\epsilon_{it\tau}^{D}$ are independent across time, which is necessary to obtain the implication that the lagged dependent variable effects reflect causal impacts. Without that assumption, the coefficients on the lagged dependent variables will reflect both causal impacts and time series processes in the errors. We are interested in the dynamics of the process regardless of whether it bears a causal interpretation and so do not attempt to investigate this issue further.

5.0) Results: Parameter Estimates

5.1) Probit Style Estimates

We begin with estimates from running simple Probits on IA use at each age (with the cross-equation restrictions related to calendar year and Native status effects mentioned above) plus a Probit on graduation. Given that we include the test score, past failure and family background variables, which we interpret as proxies for ω_i, this approach yields consistent estimates for the relevant parameters under the assumption that these are a good set of proxies (i.e., that $\theta_i = 0$). It would also yield consistent estimates if the λ's equalled zero and, hence, there was no endogeneity problem. We will call this the Homogeneous model, though there could be plenty of relevant heterogeneity in treatment effects coming through the z_i and a_i vectors.

The first two columns of Table 3 contain the derivatives for the probability of high school graduation from the Homogeneous model for females and males, respectively. The derivatives are calculated for a typical person type[8] and correspond, in each case, to a switching of the corresponding variable from 0 to 1 if it is a dummy variable or increasing it by 1 standard deviation if it is a continuous variable. The first point of interest from the table is that none of the Census tract level variables are statistically significant or economically substantial. This suggests that once we control for our rich set of individual characteristics and for school average graduation rates, neighbourhood characteristics have no strong role in the graduation outcome. On the other hand, having been in a family situation where the province had to intervene is associated with a substantial reduction in the graduation rate. Because any household with an intervention would also have had a family services call, the total effect of being in a family with an intervention is the sum of the "Family Services Call" and "Other Intervention" effects and amounts to an over .09 reduction in the graduation rate. The finding that neighbourhood characteristics do not affect longer term outcomes but family conditions do, fits with results in Oreopoulos(2003) on low income families in Toronto.

Previous school outcomes have a strong relationship with graduation. A one standard deviation increase in the average grade 11 grades increases the graduation rate by 14% for females and 16% for males, and having failed a previous grade is associated with statistically significant declines in the probability of graduating.

The other main variables of interest are the school principal dummies. In figure 2, we plot the smoothed kernel density of the estimated principal effects from the female model. These are constructed as fitted probabilities for each principal using the estimated Probit coefficients, the covariate values for the base person defined in footnote 6 and each principal's own dummy coefficient, and thus abstract from differences in the student body facing each principal. The average estimated value is .78. Note that we will see variation in these estimated effects even if there is not true heterogeneity among principals because of sampling variability. We calculate a "true" standard deviation of principal effects by subtracting an estimate of the sampling variability from the variance of the distribution in figure 2.[9] That true standard deviation is .076, implying substantial differences between good and bad principals in their impacts on graduation. The figure indicates that the main mass of the effects lie between .6 and .98 but there are some rates below .6, with the minimum being 0.3. These low values correspond to principals at small schools where small changes in the number of graduates imply large movements in the proportion graduating.

Our model contains separate sets of parameters for the IA processes at each of six ages for both drop-outs and graduates. Rather than present all of those parameters, we present the estimated marginal effects for the IA take up processes for the youngest (19.5) and oldest (24.5) groups in our data. Table 4a contains the estimated effects for 19.5 year old graduates, with the first two columns containing the results from the Homogeneous model. With a few exceptions, the neighbourhood characteristics again seem to play relatively minor roles in determining the outcome of interest. The main exceptions are that having a higher percentage native born in the census tract (i.e., fewer immigrants) is associated with higher IA use, and being in a region with a higher employment rate lowers IA use. As with the high school graduation process, the proxies for unobserved ability and background have economically substantial and statistically significant impacts. For example, having failed a grade before grade 12 is associated with an increase in the probability of using IA at age 19.5 of .066 for females and .035 for males. More substantially, the total impact associated with growing up in a family where there was a provincial intervention is a doubling of the IA use rate compared to those from non-intervention families. As a general rule, these heterogeneity variables have a larger impact on females than males. Finally, the year effects (which are not shown here) accord with well known patterns in the BC welfare system. The estimates indicate little in terms of significant changes in overall IA use up to 1996 but reflect sharp drops in use in the remaining years of the sample. The fall likely reflects a combination of the impacts of a significant tightening of the IA system in 1996 and some improvements in the labour market.

The first two columns of Table 4b contains a similar set of derivatives but for 19.5 year old drop outs. For the drop outs, the neighbourhood characteristics are even less important, with even differences in local employment rates having little impact. On the other hand, the family intervention variables have much larger effects. Having had Family Services check on the household is associated with a 10 percentage point increase in the IA usage rate for females. On the other hand, grade 11 grades play a smaller role and having failed a grade plays about the same size role as for graduates. The year effects again suggest large drop-offs in IA use after 1996.

Table 5a contains the probability derivatives for 24.5 year old graduates. Comparing the results to those in Table 4a, the most noticeable difference is that the family background and school results variables are smaller, being generally not statistically significant and often economically insubstantial. Thus, background characteristics that are important just after leaving school appear less important with time. Individual heterogeneity does not appear to take the kind of time constant form that is often assumed.

The strongest estimated effects are associated with the lagged IA and IA history variables. The lagged IA variable indicates that a female sample member who used IA at age 23.5 has a probability of using IA at age 24.5 that is .26 higher than if she had not used IA in the previous year. Having ever used IA in the past raised the probability of using IA at age 24.5 by .16 for females and .11 for males. Thus, there appear to be strong dynamic effects from IA use.

5.2) Estimates From the Full Model

We next implement the full model including the unobserved heterogeneity distribution. We found that we fit the data best with three points of support for the θ_i distribution.[10][11] The third and fourth columns of Table 3 contain probability derivatives of the high school graduation outcome from the full model for females and males.[12] The estimated derivatives are extremely similar to those from the Homogeneous model in both their absolute size and their significance patterns. The last three rows of the table are presented to help understand the role being played by the estimated unobserved heterogeneity. The first of these rows corresponds to the fitted probability generated for the base case person using the "base" value of θ_i (the case where θ_i is normalized to zero). The other two rows correspond to the differences between fitted probabilities obtained for the base case person with one of the two estimated θ_i values and the fitted probability given in the first of the three rows. From these rows, it is apparent that the basic person type has a relatively high (.86 for males) probability of graduating and that there is a reasonable level of unobserved heterogeneity, with one type having a rate that is .11 lower than the base case for both males and females.

We turn next to the probability derivatives for the IA processes based on the full model. The estimated derivatives for 19.5 year old graduates in columns 3 and 4 of Table 4a are very much like what was observed for the homogeneous model. Again, the strongest estimated effects are from the family background and school results variables. As with the high school graduation equation, we report the fitted probability for the basic person type with the base value for θ_i and the differences between that probability and the fitted probabilities associated with each of the estimated values for θ_i. Recall that in this case, the θ_i values are multiplied by $\lambda_{G,19}$. The estimated values for the λ's as well as the underlying estimates of the θ distribution are given in Table 6. We found, for males, that it is difficult to separately identify age specific λ values along with dynamic effects. As a result, we present male estimates from our most stable specification in which we restrict the λ's to be the same across ages and drop the IA history variable. For the female estimates $\lambda_{G,19}$ is negative, with the implication that the types of individuals with lower probabilities of graduating have higher probabilities of receiving IA. This would fit with a notion of θ_i as something like an ability which improves both schooling and labour market outcomes. On the other hand, the estimate of $\lambda_{G,19}$ for males is positive, which does not make immediate intuitive sense. However, it might arise in situations where there are good jobs available for drop-outs (such as in resource based economies). In that case, job offers may be made to more able individuals before they complete high school, implying an unobserved trait that is associated with a greater probability of dropping out and a lower probability of using IA. What makes this plausible is the nearly zero returns to completing high school for young males in Canada, at least at the start of this period (Donald et. al. (2001)). Whatever is behind the differing error correlations for males and females, the $\lambda_{G,19}$ parameters are sufficiently large to imply that small unobserved heterogeneity in the graduation process is translated into much larger heterogeneity in the IA process. Thus, for females, a person with the base type observed characteristics has probabilities of taking up IA at age 19.5 ranging from .03 to .53, depending on their θ_i type.

The derivatives for 19.5 year old drop outs based on the full model (shown in columns 3 and 4 of Table 4b) are again much like those from the Homogeneous model and, paralleling the results for graduates, include a negative $\lambda_{D,19}$ value for females and a positive value for males. We again observe substantial impacts of unobserved heterogeneity, with males now having the larger range.

Introducing the unobserved heterogeneity component for 24.5 year old graduates has several substantial effects relative to the results from the Homogeneous model, seen in the probability derivatives in Table 5a. First, the absolute values of the estimated family background and schooling results variables are larger, though still not always statistically significant. This is particularly the case for females. Second, the lagged IA use and IA history variables are much smaller in absolute value and, for females, switch signs. Third, the unobserved heterogeneity impacts become much larger. For females, there is now one person type with virtually zero probability of receiving IA and one type with a probability of almost. This fits with a general pattern in which the λ coefficient values increase with age.

Changes in the IA history coefficients when moving from the Homogeneous to the Full model potentially point to the conclusion that the strong IA history effects observed in the Homogeneous model results are, in part, unobserved heterogeneity effects. However, we had some difficulty in identifying the λ coefficients at older ages for males and, in fact, had to restrict the λ values across age groups to be equal. Imposing this restriction did not change any estimated coefficients apart from those on the IA history variables. We conclude from this that unobserved heterogeneity versus state dependence is not well identified at older ages (particularly for males). There is clearly persistence in the data but we do not have a convincing means for determining its source. Fortunately, the estimated treatment effects in which we are interested do not vary much with changes in how we incorporate this persistence. Thus, the heterogeneity/state dependence debate does not appear to be important for our purposes and we do not pursue it further.

The probability derivatives for drop-outs at age 24.5, shown in Table 5b, share a similar, though not identical, pattern to those for graduates. In particular, the history variable impacts are again much smaller, but for females the lagged IA variable effect remains substantial and positive.

6) Treatment Effects

6.1) Defining Treatment Effects

Our ultimate goal in estimating the models just described is to obtain estimates of the impact of graduating from high school on IA use for various groups. In the terminology of the training impact literature, we are interested in estimating treatment effects, where the treatment in this case is graduation and the outcome we focus on is IA receipt. We can meet that goal using the estimated parameters from our model. As Aakvik et. al.(2000) point out in deriving what is essentially this model in a different context, if we have consistent estimates of all of the parameters in equations 2)-4) (including the parameters determining the distribution of θ_i) then we can obtain the joint distribution of the graduation and IA outcomes for any individual and, from that, any specified treatment effect. For example, the average treatment effect at age τ for the set of individuals in group j, all of whom share the vector $\{z_{j\tau}, a_j, \theta_j\}$ is given by:

$$
\begin{aligned}
5)\quad Diff_{j\tau} &= E(D^{I,G}_{j\tau}) - E(D^{I,D}_{j\tau}) \\
&= Pr(\epsilon^{G}_{ij\tau} > -z_{j\tau}\beta_{G,\tau} - a_j\delta_{G,\tau} - \theta_j\lambda_{G,\tau}) - Pr(\epsilon^{D}_{ij\tau} > -z_{j\tau}\beta_{D,\tau} - a_j\delta_{D,\tau} - \theta_j\lambda_{D,\tau}) \\
&= (1 - F(-z_{j\tau}\beta_{G,\tau} - a_j\delta_{G,\tau} - \theta_j\lambda_{G,\tau})) - (1 - F(-z_{j\tau}\beta_{D,\tau} - a_j\delta_{D,\tau} - \theta_j\lambda_{D,\tau}))
\end{aligned}
$$

where, we have suppressed the t subscript to denote that we are seeking treatment effects controlling for macro time effects, the δ vectors correspond to combinations of the γ vector and the λ parameters, and F is the cumulative distribution function corresponding to the ϵ's. The only complication arises because θ_j is unobservable. We describe how we deal with that below.

A key insight from the recent treatment effect literature is that approaches to defining and estimating treatment effects hinge critically upon whether those effects differ across individuals. If the effects do exhibit heterogeneity then we turn to averages across groups of individuals and the central decision is which group we should focus on to make relevant policy statements or for understanding the economic processes behind the data. If the I's (latent variables) in equations 2a) and 2b) were directly observable then the question of whether there is heterogeneity in treatment effects would be completely captured in the parameters in equations 2)- 4). In particular, all individuals would face the same treatment effect if $\beta_{G,\tau} = \beta_{D,\tau}$ (apart from the intercept coefficients), $\delta_{G,\tau} = \delta_{D,\tau}$,and $\lambda_{G,\tau} = \lambda_{D,\tau}$. The situation is more complicated with qualitative dependent variables because even if $\beta_{G,\tau} = \beta_{D,\tau}$ (apart from the intercept coefficients), $\delta_{G,\tau} = \delta_{D,\tau}$,and $\lambda_{G,\tau} = \lambda_{D,\tau}$, average differentials of the type given in 5) would vary across groups with different values of $z_{j\tau}$,a_j and θ_j simply because F is a nonlinear function whose level will depend on the values of $z_{j\tau}$,a_j and θ_j . We want to differentiate this type of variation in $\mathrm{Diff}_{j\tau}$ from what we might call meaningful heterogeneity. We define meaningful heterogeneity as follows. Consider two groups, A and B, characterized by the vectors $\{z_{A\tau}, a_A, \theta_A\}$ and $\{z_{B\tau}, a_B, \theta_B\}$. Suppose the two groups have the same probability of using IA at age τ if they drop out of school. If A and B have different probabilities of using IA if they graduate (and therefore different $\mathrm{Diff}_{j\tau}$ values) then there is meaningful heterogeneity in the treatment effect. For this type of heterogeneity not to exist for any two such groups with any characteristic vectors, we must have the same condition as before: $\beta_{G,\tau} = \beta_{D,\tau}$ (apart from the intercept coefficients), $\delta_{G,\tau} = \delta_{D,\tau}$,and $\lambda_{G,\tau} = \lambda_{D,\tau}$. This defines heterogeneity at a given point on the F function and thus allows for differentials to vary along the F function.

Assuming there is meaningful heterogeneity, we need to decide for which groups we would like to calculate average treatment effects. The groups are typically defined based on conditioning related to the treatment outcome (rather than just on a value of the $\{z_{j\tau}, a_j, \theta_j\}$ vector as was done in 5)). In our case, we estimate three types of average treatment effects. The first is the average treatment effect for the whole population (also called the average treatment effect for a randomly selected individual). This average effect has a substantial intellectual history since it is what one obtains from the Heckman two step estimator and, as we show below, it is useful for discussing the relative importance of heterogeneity and true causal impacts behind observed differences between actual graduates and drop-outs. However, its policy usefulness is doubtful because we are not considering policies in which individuals are selected at random to graduate from high school. The second is the average treatment effect for drop-outs (also called the effect of treatment on the untreated). The ultimate goal of policy in this area is likely to try to get drop-outs to stay in school through graduation, and this effect corresponds to the average impact for people in that policy target group. Another common effect is the effect of treatment on the treated (i.e., the impact for those who graduated). This allows us to see whether high school pays off for those who have graduated. Since we are not considering policies in which we withdraw public schooling, this seems less relevant than treatment on the untreated in our case. Finally, local average treatment effect (LATE) estimators estimate the average effect for individuals who would change their graduation status in response to a force which exogenously shifts the graduation rate (an instrument). As we discuss below, our main instrument consists of school principal fixed effects and, thus, we can estimate a LATE corresponding to the average treatment effect among drop-outs who would be influenced to graduate high school by a better principal.

6.2.1) Homogeneous Model - Benchmark Impacts

Before generating our three treatment effects, we begin with some benchmark estimated IA impact profiles constructed from the Homogeneous Model estimates. We simulate the profiles in a series of steps. First, we generate a fitted probability of graduating from high school for each youth in the sample, given their characteristics and the estimates from the high school graduation equation. This calculated probability is compared with a random draw from a uniform distribution on [0, 1] to assign an appropriate graduation status to each sample member. Those youth assigned to graduate are then used to simulate IA take-up based on their individual characteristics and using the parameter estimates from the IA take-up equations for graduates. A probability of IA use is constructed at each age, and these calculated probabilities are compared to random draws from U[0, 1]. If an individual is predicted to take-up IA at any age, that then impacts the value of the IA history variables included at older ages when constructing probabilities of IA take-up. The same procedure is used to simulate IA take-up for youths predicted to drop-out. These simulations are conducted 200 times for each individual in the sample. The average of the simulated probabilities across all individuals and across these 200 repetitions are then calculated, for graduating from high school, and for IA take-up at each age for predicted graduates and predicted dropouts.

In figures 2a and 2b, we plot the fitted probability of using IA for graduates and dropouts, and the difference between the two, along with the raw IA use proportions from our sample. To match the sample probabilities, we only use an individual's simulated history at a given age if that person was in our sample at that age. The resulting fitted probabilities[13] are very close to the observed IA use patterns. These model benchmark patterns thus match closely the main patterns in the data that we are interested in. The difference in IA use probabilities between graduates and dropouts is the main focus in our analysis, as we attempt to uncover whether this is the causal impact of graduation.

The IA use profiles in figures 2a and 2b are relatively strongly negatively sloped for both males and females. This might arise because IA use declines with age but might also reflect influences from the types of macro patterns seen in Tables 4 and 5. For this reason, we construct our remaining profiles with all the time dummy variables set to zero in order to net out macro effects.

We construct confidence intervals for these simulations using a parametric bootstrap based on the estimated parameter vector, its variance-covariance matrix and an assumption of normality. We perform the simulation exercise behind the IA profile estimates for each parameter vector draw. We repeat this procedure for each of 500 "draws" from the estimated parameter distribution and use the results to derive confidence intervals for our impact estimates.

The result of this process is given in the top left panel of Figure 4 for males and in the top left panel of Figure 5 for females. The thicker lines in these figures corresponds to the estimated graduation effect while the two thinner lines correspond to the 90% confidence interval (the 5th and 95th percentiles from the bootstrapped impact distribution). For males, the impact is quite flat with respect to age at a level just above .1. This implies that high school drop-outs have nearly a 10 percentage point higher likelihood of receiving IA at any age than graduates. Given the IA usage rate of .16 for drop-outs from IA families listed in Table 1, this represents a very substantial effect. The confidence interval indicates these effects are significantly different from zero at the 90% confidence level and they are, in fact, also significant at the 95% level. For females, the estimated impacts are approximately double those for males and rise in the last few years of our age window. These impacts amount to at least 3/4 of the observed drop-out IA use shown in Table 1. The estimated impact for females is also statistically significant at any conventional significance level.

6.2.2) Homogeneous Model - Treatment Effects

The initial impact estimates just discussed would provide consistent treatment effect estimates if there were no heterogeneity in treatment effects and the proxy variables fully addressed the endogeneity issue. As discussed earlier, we can test the no heterogeneity assumption by testing the joint set of restrictions, $\beta_{G,\tau} = \beta_{D,\tau}$ (apart from the intercept) and $\delta_{G,\tau} = \delta_{D,\tau}$. Tests of that set of restrictions rejects them at any conventional significance level for both males and females.[14] Given this, the initial predictions amount to comparing graduate and drop-out IA outcomes and, since the two groups differ in observable characteristics, this generates an estimate that may not represent the actual treatment effect for any individual.

We next construct the average treatment effect for a random individual. To do this, we put everyone in the sample (whether that person is a graduate or a drop-out) through the IA take-up simulation exercise as if they were a drop-out and then repeat it as if they were a graduate. This provides us with two complete sets of histories: one that would occur if everyone graduated and one that would occur if everyone dropped out. Averaging each of these sets of histories at each age and then taking the difference between the averages yields our estimated treatment effect for a random person. This is presented in the second panel in the left column in Figure 4 for males and Figure 5 for females. For males, the estimated effect rises slightly from about .067 at younger ages to just over .1 by age 23 and is statistically significant at either the 90% or 95% level of confidence. Thus, between about a fifth to a third of the apparent effect shown in the first panel of the column is due to differences between graduates and drop-outs in observable characteristics. For females, the effect for a random person also rises from about 2/3 to about 4/5 of the effect in the top panel. Once again the female effects are larger and have a stronger age slope than those for males. The female effects are statistically significant at the 95% level of confidence at most ages.

The third panel in the columns presents the estimates of the effect of treatment on the untreated. To construct this estimate, we begin with the sub-sample of youth who actually drop out during grade 12. For each member of this sub-sample, we predict a graduation status in the same manner as we did in the exercise behind the first panel. Those who are predicted to drop out become our "untreated" sample. We simulate an IA usage path for each member of this untreated sample using the parameters corresponding to the IA process given graduation and another path using the parameters from the IA process given dropping out. The difference between the averages for these paths at each age is our estimate of treatment on the untreated. The resultant estimate for males ranges between 4 and 9% but is statistically significant at the 10% significance level only at age 19.5. This path is lower than that for a random individual, indicating that graduates must have higher graduation effects than non-graduates. This is an example of the underlying heterogeneity in treatment effects and would correspond to those who benefit most from graduation being the ones who actually choose it. For females, in contrast, the treatment on the untreated effect is about .1 for most years, with a rise again at older ages. The greater similarity between the random treatment and treatment on the untreated effects suggests weaker heterogeneity in effects among females.

One point of interest is to examine how special is our "untreated" sample. We can do this by examining averages for observable covariates among the “untreated” and comparing them to those for the overall sample. It is particularly interesting to compare differences in average test scores because, in our structure, these most closely represent what would be unobserved heterogeneity if we did not have test scores in our data. For our sample as a whole, the mean value for the test score variable is 2.01 for males and 2.26 for females. In the untreated sample the means are 1.34 and 1.47, respectively. Thus, the untreated group has decidedly lower ability as measured by this variable.

Our third treatment effect is the LATE associated with the principals instrumental variable. To construct this effect, we started with the untreated sample drawn as in the treatment on the untreated effect calculation. In constructing the probability of graduating for each person in that sample we set the dummy variable associated with their actual principal to one, implying that the estimated effect for that principal was part of the fitted probability calculation. In forming the relevant sample for the LATE, we formed new fitted probabilities of graduating for the untreated sample after increasing the principal effect that is relevant for them by the equivalent of one “true” standard deviation in the distribution of the estimated principal effects. We again compared the fitted probability with a draw from a uniform distribution to assign graduate status. Approximately half the untreated sample are predicted to graduate after the improvement in principal effect. It is these switchers who are affected by the instrument and it is the average treatment for them that constitutes the LATE. Thus, we simulated IA histories for this group both as if they were graduates and as if they were drop-outs and formed a treatment effect based on those histories. That is reported in the bottom panel of the left columns in Figures 4 and 5. For males, the LATE differs very little from the treatment on the untreated effect. This likely arises because the principal variables are relatively uncorrelated with the other covariates so that a subsample chosen based on changes in principal effects does not have a substantially different covariate distribution relative to the total drop out sample. Since our method for forming the LATE essentially consists of integrating over individual treatment effects with respect to the sample covariate distribution, the implication is that if covariate distributions are the same across groups then so will be their average treatment effects (see Heckman et. al.(2000) for a discussion of this point). This is corroborated by the fact that the average test score for those who would be induced to graduate by a better principal is 1.6, only slightly higher than that for the untreated sample as a whole.

For females, the LATE estimates are also not substantially different from the estimated treatment on the untreated effects. The estimates again suggest effects of graduation on the order of a reduction in the probability of receiving IA by .1 to .15. These effects are statistically significant at the 10% significance level at all ages and at the 5% level at the youngest age.

6.3 Estimated Effects Based on the Full Model

We repeat this simulation exercise using the results from the full model. Figures 3a and 3b contain the fitted probabilities and differences from the model including time effects and compares those to the actual sample probabilities. As with the Homogeneous model, the fitted difference in probability of using IA between graduates and drop-outs observed in simple sample means is very well matched by the fitted model values. Thus, the full model also appears to perform well.

In the second column of figures in Figures 4 and 5 we repeat our treatment effects estimation exercise based on the full model estimates. Our approach for constructing the treatment effects from the full model is the same as that from the homogeneous model with adjustments to address the presence of θ_i terms. In particular, each person is assigned one of the three estimated values for θ_i according to the estimated probability associated with each. That value of θ_i then effectively becomes one of their covariates for the remainder of the simulation.

What is most noticeable about these estimates is the wider confidence intervals compared to those derived from the homogeneous model in the first column. This arises because the parameters associated with the θ_i distribution are less precisely estimated than other model parameters, particularly for males. We suspect this is true, in part, because our "proxy" variables (such as the grade 11 average grade) are picking up much of the relevant heterogeneity.

For males, the initial fitted (with time effects removed) simulation is much the same as that from the homogeneous model. We again reject the restriction that the drop-out and graduate coefficient vectors (the β's, δ's and λ's) are equal, indicating that there is heterogeneity in returns to graduation. Interestingly, though, we cannot reject the restriction that the λ_G parameter equals the λ_D parameter. That is, there is no evidence of heterogeneity in returns to graduation in an unobserved dimension.[15] That, in turn, means that differences across treatment effects stem from heterogeneity in returns to observable variables. The remaining treatment effects are somewhat larger than those from the homogeneous model and show a stronger upturn at older ages. Comparing the observed (time trends removed) difference in IA use between actual graduates and drop-outs in the top panel and the causal effect for a random individual in the second panel allows us to evaluate the importance of causality versus (observed and unobserved) heterogeneity in the observed differences. The random person effects are larger than the observed differences at all ages. This implies that observed differences between drop-outs and graduates are completely attributable to causal impacts. Fro the random person effects and the two other treatment effects for males, there is strong evidence that graduation pays off more as people get older. It appears that, given the combination of observed and unobserved characteristics for the people in our sample, building experience is particularly important as a route to them fully implementing their education. All the treatment effects are statistically significant at the 10% significance level, with the random person effects being significant at the 5% level as well.

The heterogeneity patterns are again interesting in their own right. As in the Homogeneous model, the random person effect is above the treatment on the untreated effect at young ages for males. As before, this suggests that the effect of high school graduation is larger for those who do actually graduate. In contrast, and again mirroring the Homogeneous model results, the treatment on the untreated and LATE estimates are not very different, suggesting little heterogeneity among drop-outs. Again, this is corroborated by the average test score which is 2.01 for the whole sample, 1.34

for the untreated sample and 1.47 for the LATE sample. Since we assign a θ_i value to each person in a given simulation, we can also see whether the distribution of θ_i values differs across sub-samples. For males, there is almost no difference in the proportion of each sub-sample with each of the three possible values of θ_i.

Unobserved heterogeneity appears to play a much larger role for females than males. Introducing that heterogeneity into the estimation yields treatment effect profiles that are more substantially different from the Homogeneous model in both shape and level for females compared to males. For females, the estimated treatment effects take on something closer to a U shape, with the treatment on the untreated effects being statistically significantly different from zero at the 10% level at both the oldest and youngest ages in our sample. Further, the random individual effect falls from about 60% of the observed difference between actual graduates and drop-outs at age 19 to about 40% at ages 20 and 21 before rising to about 2/3 by age 24. As a result, while the observed differences in IA use between actual graduates and drop-outs are substantially larger for females than males, the causal impacts of graduation for a random person are, if anything, larger for males. The reason for larger observed differentials for females lies in differences in observed and unobserved characteristics. Since both the observed differences and the random individual effects are decidedly larger for females when using the Homogeneous model, it appears the gender differential arises mainly from differences in unobserved characteristics. For males, the fact that the estimates from the Homogeneous and the Full models are similar at most ages suggests that variables such as the test score capture much of the relevant (usually) unobserved heterogeneity. For females, though, there is still substantial unobserved heterogeneity even after controlling for these variables. One wonders, though there is no way to investigate it with this data, whether the extra female heterogeneity is related to child bearing.

Examining the treatment on the untreated and LATE estimates for females, we again conclude that graduation has its greatest causal impact at older ages. However, for females there also seems to be an effect of graduation observed just after leaving school with a low impact point around age 22. While results are not always well defined, the overall implication seems to be that graduation does have a causal impact in reducing IA use but that such an effect does not become substantial until about five years after school leaving. As with the males, comparisons across treatment effect estimates for females indicates that the effect of graduation is larger for graduates than for drop-outs and that there does not appear to be large heterogeneity among the drop-outs.

6.4) The Distribution of Impacts

The differences in the β and δ vectors between drop-outs and graduates imply there is heterogeneity in the impact of graduation. We investigate this heterogeneity further, focussing on the treatment on the untreated effects. To do this, we first form a sample of "untreated" individuals and put them through both the drop-out and graduate relevant IA processes 200 times.[16] For each individual, we then take the difference between their frequency of IA receipt in those 200 simulations as a drop-out and as a graduate at each age. In effect, we are just examining the distribution of fitted probabilities across individuals of using IA at each age but the dynamics make obtaining those fitted probabilities somewhat complicated since the probability at one age depends on whether the individual received IA at earlier ages. The implied heterogeneity is large, with the 10th percentile of the estimated effects being slightly negative and the 90th percentile being about .2 for both males and females.

The overall heterogeneity is generated from the heterogeneity in impacts across different observed characteristics of individuals. As a result, we can investigate what distinguishes high impact from middle and low impact individuals. To do this, we select three sub-samples: low impact (those whose estimated impact falls in between the 5th and 15th percentiles of the estimated impact distribution that underlies figure 6); middle impact (those with estimated impacts between the 45th and 55th percentiles); and high impact (those with estimated impacts between the 85th and 95th percentiles). In Table 7, we present means for various observable characteristics for each of these sub-samples, separately for males and females. Often in other studies, the impacts of education and training are greatest for the least disadvantaged groups, generating a tug of war between who we need to help and who would get the most advantage from a dollar spent. Some of this is evident in Table 7, as the high impact group has higher average grade 11 grades and lower probabilities of having failed an earlier grade than the low and middle impact groups for both males and females. However, for males, the proportion from families in which the province has intervened rises considerably with the level of the graduation impact. For females the same is true for family services visits but not other impacts. For both genders as well, the proportion from low average income neighbourhoods rises substantially with the impact level, particularly at young ages. Thus, Table 7 suggests that targeting educational resources at children in low income neighbourhood households where the province has needed to intervene may be particularly beneficial.

7) Conclusions

In this paper, we study the impact of high school graduation on the probability individuals from welfare backgrounds use welfare themselves at each age between 19 and 24. We make use of a unique dataset linking high school and welfare records for a group of individuals entering grade 12 in the province of British Columbia in the period 1991 to 1996. We address potential endogeneity issues using a combination of proxy variables, instruments and an econometric framework incorporating unobserved heterogeneity.

We find that there are substantial differences in IA use between high school graduates and drop outs from IA families. For females, drop outs have raw average IA take-up rates of over .2 while graduates have rates under .1. However, there is good reason to believe that this does not represent the causal impact of graduation on welfare use. People who are more likely to graduate may be high ability people who would be less likely to use welfare even if they did not graduate. Our estimates indicate that there is substantial heterogeneity in the actual causal impact of graduation on IA take-up and, as a result, we obtain different average treatment effects for different groups. For females, among drop-outs, graduation would cause a substantial reduction in IA use (on the order of between .1 and .15), with effects rising with age and the magnitude of the impact depending on how we control for heterogeneity. For males, impacts are of the same order of magnitude and also rise with age. Causal effects for graduates appear to be much larger, indicating that it is those who get the best return from graduating who actually do so. However, we also investigate heterogeneity in graduation impacts more directly and find that impacts are larger for students who come from families in which the province had to visit the household because of problems in the family and from low income neighbourhoods. This suggests there may be an opportunity to have sizeable impacts by targeting education resources at these types of families. Finally, differences in unobserved heterogeneity between drop-outs and graduates appear to play a substantial role in observed raw differences in IA take-up between the two groups for females while our (quite complete) set of observed heterogeneity controls play a larger role for males.

Table 1
Income Assistance Receipt by Gender, Education,
Family Background

	Drop-Out	Graduates	# of Observations
Females			
IA Family	0.24	0.094	9272
Non-IA Family	0.084	0.019	8495
Males			
IA Family	0.16	0.065	7534
Non-IA Family	0.046	0.009	7930

Table 2:
Graduation Rates By Gender, Family

	Graduation Rate
Females	
IA Family	72.3
Non-IA Family	85.4
Males	
IA Family	65.1
Non-IA Family	80.2

Table 3
Probability Derivatives for the High School Graduation Equation

	Homogeneous Model		Full Model	
	Females	Males	Females	Males
Base probability	0.791*** (0.062)	0.780*** (0.059)	0.784*** (0.063)	0.798*** (0.056)
Aboriginal background	-0.039* (0.021)	-0.038 (0.026)	-0.040* (0.022)	-0.034 (0.025)
Non-English at home	0.032* (0.018)	0.001 (0.020)	0.032* (0.018)	0.000 (0.021)
% One parent families (cd)	0.000 (0.007)	-0.013 (0.008)	-0.000 (0.007)	-0.011 (0.008)
% Canadian born (cd)	0.001 (0.013)	0.015 (0.014)	-0.000 (0.013)	0.016 (0.014)
% More than grade 9 education (cd)	-0.001 (0.010)	0.003 (0.011)	-0.002 (0.011)	0.002 (0.012)
Employment rate (cd)	-0.007 (0.009)	0.007 (0.011)	-0.007 (0.010)	0.005 (0.011)
High ave. income (cd)	-0.004 (0.017)	0.011 (0.019)	-0.003 (0.018)	0.009 (0.019)
Low ave. income (cd)	-0.034* (0.019)	0.035* (0.020)	-0.035* (0.019)	0.030 (0.020)
Family services call	-0.036** (0.014)	-0.028* (0.016)	-0.037** (0.015)	-0.025 (0.016)
Other intervention	-0.052*** (0.014)	-0.033** (0.014)	-0.052*** (0.014)	-0.030** (0.014)
Grade 11 grades	0.144*** (0.035)	0.159*** (0.035)	0.148*** (0.035)	0.149*** (0.034)
Failed a grade or more	-0.050*** (0.014)	-0.028** (0.012)	-0.052*** (0.015)	-0.028** (0.012)
School mean graduation rate	-0.002 (0.020)	0.045*** (0.017)	-0.004 (0.021)	0.042** (0.017)
Cohort Dummies	yes	yes	yes	yes
Base (zero theta)			0.804*** (0.060)	0.864*** (0.051)
Theta one			-0.114** (0.035)	-0.111** (0.044)
Theta two			-0.048** (0.024)	0.071*** (0.022)

Significant at the 10% (*), 5% (**), or 1% (***) level of significance. Standard errors are in parentheses

Table 4a:
Probability Derivatives for IA Take-Up at Age 19.5 - High School Graduates

	Homogeneous Model		Full Model	
	Females	Males	Females	Males
Base probability	0.079*** (0.009)	0.059*** (0.009)	0.089*** (0.010)	0.058*** (0.009)
Aboriginal background	0.042*** (0.011)	0.032*** (0.012)	0.064*** (0.015)	0.042** (0.017)
Non-English at home	-0.029** (0.011)	-0.012 (0.012)	-0.030** (0.012)	-0.011 (0.014)
% One parent families (cd)	0.004 (0.004)	0.010** (0.005)	0.007 (0.004)	0.008* (0.005)
% Canadian born (cd)	0.017*** (0.005)	0.003 (0.005)	0.020*** (0.005)	0.004 (0.005)
% More than grade 9 education (cd)	0.005 (0.005)	-0.008* (0.005)	0.000 (0.005)	-0.008 (0.005)
Employment rate (cd)	-0.009** (0.004)	-0.011*** (0.004)	-0.010** (0.004)	-0.011** (0.004)
High ave. income (cd)	0.002 (0.011)	-0.002 (0.010)	-0.000 (0.011)	-0.002 (0.010)
Low ave. income (cd)	-0.005 (0.009)	-0.018** (0.008)	-0.007 (0.009)	-0.018** (0.008)
Family services call	0.037*** (0.011)	0.036*** (0.014)	0.042*** (0.012)	0.031** (0.014)
Other intervention	0.048*** (0.011)	0.024** (0.011)	0.047*** (0.011)	0.021* (0.011)
Grade 11 grades	-0.026*** (0.004)	-0.019*** (0.004)	-0.028*** (0.004)	-0.014*** (0.004)
Failed a grade or more	0.066*** (0.012)	0.035*** (0.010)	0.071*** (0.012)	0.033*** (0.011)
School mean graduation rate	-0.001 (0.004)	-0.003 (0.004)	-0.000 (0.004)	-0.002 (0.004)
Year Effects	Yes	Yes	Yes	Yes
Base (zero theta)			0.028*** (0.006)	0.105*** (0.025)
Theta one			0.503*** (0.053)	-0.100*** (0.023)
Theta two			0.129*** (0.017)	0.445*** (0.057)

Significantly different from zero at the 10% (*), 5% (**), or 1% (***) level of significance
Standard errors are in parentheses.

Table 4b:
Probability Derivatives for IA Take-Up at Age 19.5 - High School Dropouts

	Homogeneous Model		Full Model	
	Females	Males	Females	Males
Base probability	0.198*** (0.022)	0.122*** (0.018)	0.179*** (0.021)	0.171*** (0.030)
Aboriginal background	0.016 (0.016)	0.039** (0.017)	0.040* (0.023)	0.051** (0.026)
Non-English at home	0.020 (0.029)	-0.038* (0.020)	0.023 (0.028)	-0.048* (0.027)
% One parent families (cd)	0.002 (0.008)	-0.006 (0.006)	0.001 (0.008)	-0.007 (0.009)
% Canadian born (cd)	0.021** (0.009)	0.005 (0.007)	0.020** (0.009)	0.007 (0.010)
% More than grade 9 education (cd)	0.013 (0.012)	0.004 (0.009)	0.009 (0.011)	0.002 (0.012)
Employment rate (cd)	-0.004 (0.010)	-0.001 (0.008)	-0.003 (0.010)	-0.002 (0.011)
High ave. income (cd)	0.002 (0.024)	-0.023 (0.017)	0.002 (0.023)	-0.028 (0.023)
Low ave. income (cd)	0.016 (0.022)	-0.002 (0.017)	0.022 (0.021)	0.003 (0.022)
Family services call	0.096*** (0.023)	0.085*** (0.023)	0.088*** (0.022)	0.091*** (0.027)
Other intervention	0.033* (0.019)	0.047*** (0.018)	0.036** (0.018)	0.061*** (0.023)
Grade 11 grades	-0.015 (0.009)	-0.014* (0.007)	-0.018** (0.009)	0.004 (0.013)
Failed a grade or more	0.070*** (0.019)	0.031** (0.014)	0.067*** (0.018)	0.029* (0.017)
School mean graduation rate	-0.000 (0.008)	-0.015** (0.006)	-0.000 (0.008)	-0.017** (0.008)
Year Effects	Yes	Yes	Yes	Yes
Theta one			0.528*** (0.069)	-0.316*** (0.052)
Theta two			0.191*** (0.029)	0.504*** (0.047)

Significantly different from zero at the 10% (*), 5% (**), or 1% (***) level of significance
Standard errors are in parentheses.

Table 5a:
Probability Derivatives for IA Take-Up at Age 24.5 - High School Graduates

	Homogeneous Model		Full Model	
	Females	Males	Females	Males
Base probability	0.028*** (0.011)	0.042*** (0.016)	0.100*** (0.021)	0.105*** (0.035)
Aboriginal background	0.020** (0.008)	0.025** (0.012)	0.037*** (0.010)	0.063** (0.028)
Non-English at home	0.026 (0.031)	0.023 (0.037)	-0.026 (0.027)	0.023 (0.072)
% One parent families (cd)	-0.007* (0.004)	-0.002 (0.008)	-0.002 (0.009)	0.000 (0.018)
% Canadian born (cd)	-0.001 (0.006)	-0.003 (0.008)	0.014 (0.012)	0.003 (0.018)
% More than grade 9 education (cd)	-0.014** (0.005)	0.007 (0.011)	-0.034*** (0.010)	0.008 (0.023)
Employment rate (cd)	-0.008* (0.004)	0.015 (0.012)	-0.025*** (0.009)	0.018 (0.017)
High ave. income (cd)	-0.009 (0.010)	-0.008 (0.019)	-0.016 (0.020)	-0.012 (0.046)
Low ave. income (cd)	0.004 (0.012)	-0.013 (0.016)	-0.014 (0.019)	-0.039 (0.030)
Family services call	-0.006 (0.009)	0.029 (0.028)	0.019 (0.024)	0.059 (0.061)
Other intervention	0.028* (0.016)	-0.015 (0.016)	0.081*** (0.028)	-0.030 (0.047)
Grade 11 grades	-0.012** (0.005)	-0.006 (0.008)	-0.037*** (0.011)	-0.018 (0.017)
Failed a grade or more	-0.007 (0.008)	0.014 (0.018)	0.076** (0.038)	0.036 (0.040)
School mean graduation rate	-0.008** (0.004)	0.005 (0.008)	-0.011* (0.007)	0.007 (0.014)
Lagged IA take-up	0.244*** (0.066)	0.303*** (0.094)	-0.040** (0.016)	0.160** (0.069)
History of IA take-up	0.157*** (0.035)	0.112*** (0.040)	-0.048** (0.023)	-
Year Effects	Yes	Yes	Yes	Yes
Base (zero theta)			0.000 (0.000)	0.206*** (0.077)
Theta one			0.995*** (0.010)	-0.192*** (0.068)
Theta two			0.184*** (0.067)	0.506*** (0.044)

Significantly different from zero at the 10% (*), 5% (**), or 1% (***) level of significance.

Table 5b:
Probability Derivatives for IA Take-Up at Age 24.5 - High School Dropouts

	Homogeneous Model		Full Model	
	Females	Males	Females	Males
Base probability	0.092*** (0.032)	0.080** (0.033)	0.238*** (0.056)	0.294*** (0.083)
Aboriginal background	0.009 (0.010)	0.029* (0.016)	0.037 (0.023)	0.068** (0.034)
Non-English at home	-0.037 (0.039)	0.035 (0.061)	-0.088 (0.071)	0.025 (0.133)
% One parent families (cd)	0.015 (0.014)	0.017 (0.014)	0.022 (0.021)	0.026 (0.028)
% Canadian born (cd)	0.005 (0.014)	0.014 (0.016)	0.020 (0.025)	0.022 (0.032)
% More than grade 9 education (cd)	-0.004 (0.017)	-0.008 (0.016)	0.026 (0.032)	-0.027 (0.041)
Employment rate (cd)	-0.003 (0.016)	-0.012 (0.013)	-0.006 (0.028)	-0.040 (0.031)
High avg. income (cd)	0.043 (0.049)	-0.030 (0.031)	0.087 (0.073)	-0.088 (0.084)
Low avg. income (cd)	0.012 (0.033)	-0.011 (0.027)	0.002 (0.051)	-0.021 (0.059)
Family services call	0.010 (0.030)	0.086* (0.050)	0.027 (0.049)	0.185** (0.081)
Other intervention	0.011 (0.028)	-0.028 (0.024)	0.041 (0.045)	-0.018 (0.066)
Grade 11 grades	-0.003 (0.016)	-0.012 (0.013)	-0.003 (0.026)	0.013 (0.045)
Failed a grade or more	-0.008 (0.025)	-0.018 (0.022)	0.015 (0.050)	-0.039 (0.058)
School mean graduation rate	-0.020* (0.011)	0.013 (0.014)	-0.033* (0.019)	0.022 (0.033)
Lagged IA take-up	0.531*** (0.078)	0.314*** (0.091)	0.240** (0.123)	0.184*** (0.067)
History of IA take-up	0.120** (0.048)	0.162*** (0.052)	-0.015 (0.058)	-
Year Effects	Yes	Yes	Yes	Yes
Base (zero theta)			0.076 (0.050)	0.570*** (0.134)
Theta one			0.873*** (0.113)	-0.466*** (0.080)
Theta two			0.410*** (0.133)	0.377*** (0.095)

Significantly different from zero at the 10% (*), 5% (**), or 1% (***) level of significance.

Table 6:
Estimates of Mass Points, Their Associated Probabilities and the Lambdas

	Females	Males
Mass point one	-0.359** (0.159)	-0.414*** (0.159)
Mass point two	-0.164** (0.074)	0.415** (0.163)
Prob. on 1st mass point	0.045*** (0.005)	0.621*** (0.042)
Prob. on 2nd mass point	0.299*** (0.025)	0.035*** (0.009)
Graduate lambdas		
Age 19.5	-5.55** (2.49)	3.32*** (1.28)
Age 20.5	-4.84** (2.20)	as above
Age 21.5	-6.89** (3.13)	as above
Age 22.5	-8.51** (3.88)	as above
Age 23.5	-22.44** (11.08)	as above
Age 24.5	-17.69** (8.29)	as above
Dropout lambdas		
Age 19.5	-4.50** (2.24)	3.47*** (1.31)
Age 20.5	-6.32** (3.08)	as above
Age 21.5	-8.29** (4.04)	as above
Age 22.5	-15.07** (7.45)	as above
Age 23.5	-13.67* (7.30)	as above
Age 24.5	-8.54* (4.80)	as above

* Significantly different from zero at the 10% level of significance
** Significantly different from zero at the 5% level of significance
*** Significantly different from zero at the 1% level of significance

Table 7a
Characteristics of Individuals by Age and Level of the Impact of Graduation on IA Receipt
Males

	Age 19.5			Age 24.5		
Characteristic	Low Impact	Middle Impact	High Impact	Low Impact	Middle Impact	High Impact
Native	0.096	0.052	0.072	0.09	0.07	0.08
Non-Engl. At Home	0.26	0.11	0.032	0.11	0.11	0.09
Family Services Call	0.17	0.26	0.56	0.11	0.19	0.33
Other Intervention	0.074	0.12	0.43	0.21	0.32	0.3
Grade 11 grades	1.15	1.17	1.36	1.11	1.2	1.27
Failed a year or more	0.64	0.41	0.51	0.8	0.46	0.24
Low Avg. Income	0.13	0.29	0.29	0.1	0.19	0.29
High Avg. Income	0.3	0.13	0.09	0.46	0.15	0.04
Obs.	135	135	140	125	136	136

Table 7b
Characteristics of Individuals by Age and Level of the Impact of Graduation on IA Receipt
Females

	Age 19.5			Age 24.5		
Characteristic	Low Impact	Middle Impact	High Impact	Low Impact	Middle Impact	High Impact
Native	0.32	0.071	0.029	0.26	0.12	0.07
Non-Engl. At Home	0.01	0.04	0.23	0.06	0.18	0.059
Family Services Call	0.16	0.25	0.53	0.23	0.26	0.37
Other Intervention	0.55	0.33	0.37	0.69	0.36	0.17
Grade 11 grades	1.12	1.35	1.44	1.1	1.27	1.47
Failed a year or more	0.66	0.4	0.47	0.59	0.49	0.39
Low Avg. Income	0.14	0.22	0.41	0.2	0.29	0.22
High Avg. Income	0.09	0.25	0.17	0.07	0.2	0.3
Obs.	98	127	137	130	122	138

References

Aakvik, A., J.J. Heckman and E.J. Vytlacil (2000). “Treatment Effects for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs,” NBER Technical Working Paper 262, September 2000.

An, C., R. Haveman and B. Wolfe (1993). “Teen Out-of-Wedlock Births and Welfare Receipt: The Role of Childhood Events and Economic Circumstances,” Review of Economics and Statistics, Vol. 75, No. 2, pp. 195-208.

Antel, J. (1992). “The Inter-generational Transfer of Welfare Dependency: Some Statistical Evidence,” Review of Economics and Statistics, Vol. 74, No. 3, pp. 467-73.

Barrett, G. 2000). “The Effect of Educational Attainment on Welfare Dependence: Evidence from Canada,” Journal of Public Economics, Vol. 77, pp. 209-232.

Barrett, G. (1998). "The Duration of Welfare Spells and State Dependence: Evidence from British Columbia," Department of Economics, University of New South Wales.

Barrett, G. and M. Cragg (1998). "An Untold Story: The Characteristics of Welfare Use in British Columbia," Canadian Journal of Economics, Vol. 31, No. 1, pp. 165-188.

Beaulieu, N., J-Y. Duclos, B. Fortin and M. Rouleau (2001). “An Econometric Analysis of Intergenerational Reliance on Social Assistance,” Department of Economics, Universite Laval.

Blank, R.M. (1989). "Analyzing the Length of Welfare Spells," Journal of Public Economics, vol. 39, pp. 245-73.

Boggess, S. (1998). “Family Structure, Economic Status, and Educational Attainment,” Journal of Population Economics, Vol. 11, pp. 205-22.

Card, David (1999). “The Causal Effect of Education on Earnings,” in O. Ashenfelter and D. Card (eds) The Handbook of Labor Economics, Vol. IIIa, Amsterdam: Elsevier.

Carneiro, P., K. T. Hansen and J.J. Heckman.(2003). “Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice,” NBER Working Paper 9546, March.

Donald, S.G., D.A. Green and H.J. Paarsch (2000). “Differences in Wage Distributions Between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates,” Review of Economic Studies, 67(4), pp. 609-33.

Gottschalk, P. (1990). “AFDC Participation Across Generations,” American Economic Review, Vol. 80, No. 2, pp.367-71.

Gottschalk, P. (1996). “ Is the Correlation in Welfare Participation Across Generations Spurious?” Journal of Public Economics, Vol. 63, pp. 1-25.

Heckman, J.J. (1981). “Heterogeneity and State Dependence,” in Studies in Labor Markets, S. Rosen, ed.. Chicago: The University of Chicago Press.

Heckman, J.J. (1997). "Instrumental Variables: A Study of Implict Behavioral Assumptions Used in Making Program Evaluations," Journal of Human Resources, Vol 32, no. 3, pp. 443-462.

Heckman, J.J., R.J. LaLonde, and J.A. Smith (1999). “The Economics and Econometrics of Active Labor Market Programs,” forthcoming in Handbook of Labor Economics, Vol. III, O. Ashenfelter and D. Card, eds..

Heckman, J.J. and B. Singer (1984). “A Method for Minimising the Impact of Distributional Assumptions in Econometric Models for Duration Data,” Econometrica, Vol. 22, no. 2, pp. 271-320.

Heckman, J.J., J. Smith and N. Clements (1997). “Social Experiments: Accounting for Heterogeneity in Programme Impacts,” Review of Economic Studies, Vol. 64, pp. 487-535.

Heckman, J.J., J.L. Tobias and E. Vytlacil (2000). “Simple Estimators for Treatment Parameters in a Latent Variables Framework with an Application to Estimating the Returns to Schooling,” NBER Working Paper 7950, October.

Ku, I. and R.D. Plotnick (2000). “Do Children From Welfare Families Obtain Less Education?” Institute for Research on Poverty, Discussion Ppaer no. 1217-00, December.

Levine, P.B. and D.J. Zimmerman (1996). "The Intergenerational Correlation in AFDC Participation: Welfare Trap or Poverty Trap?" Institute for Research on Poverty Discussion Paper #1100-96, July.

MaCurdy, T.(2000). "How Does Welfare Influence Young Women's School and Work Achievements?" Department of Economics, Stanford University, May 2000.

Meghir, C. (2002). "The Effect of School Quality on Educational Attainment and Wages." Review of Economics and Statistics.

______Oreopoulis, P. (2003). "The Long-run Consequences of Living in a Poor Neighborhood." The Quarterly Journal of Economics, Vol. 118, no. 4, pp. 1533-75.

______Raptis, H. and T. Fleming (2002). "Effective Education: Reframing Effective Schools Research." C.D. Howe Institute Commentary.

Vytlacil, E. (2000). "Independence, Monotonicity and Latent Variable Models: An Equivalence Result" Stanford University.

Appendix A
Likelihood functions

Here is the estimated log-likelihood function for the Homogeneous Model:

$$L_H = \sum_{i=1}^{N} \log([1 - \Phi(-x_{it}\alpha - a_i\gamma)]^{D_i^H} * \Phi(-x_{it}\alpha - a_i\gamma)^{1-D_i^H} *$$
$$(\prod_{\tau=19}^{24} ([1 - \Phi(-z_{it\tau}\beta_{G,\tau} - a_i\delta_{G,\tau})]^{D_{i\tau}^{I,G}} * \Phi(-z_{it\tau}\beta_{G,\tau} - a_i\delta_{G,\tau})^{1-D_{i\tau}^{I,G}})^{O_{i\tau}})^{D_i^H} *$$
$$(\prod_{\tau=19}^{24} ([1 - \Phi(-z_{it\tau}\beta_{D,\tau} - a_i\delta_{D,\tau})]^{D_{i\tau}^{I,D}} * \Phi(-z_{it\tau}\beta_{D,\tau} - a_i\delta_{D,\tau})^{1-D_{i\tau}^{I,D}})^{O_{i\tau}})^{1-D_i^H})$$

where:

(a) D_i^H equals one if individual i (sample size N) graduated from high school, zero otherwise,

(b) $D_{i\tau}^{I,G}$ and $D_{i\tau}^{I,D}$ equals one if individual i took up IA at age τ, zero otherwise, for graduates and drop-outs respectively,

(c) $O_{i\tau}$ equals one if individual i is observed at age τ in our sample, zero otherwise, and

(d) the function $\Phi(..)$ denotes the standard normal cumulative distribution function.

For the Full Model, the log-likelihood function is as follows:

$$L_F = \sum_{i=1}^{N} \log(\sum_{j=1}^{J} P_j * ([1 - \Phi(-x_{it}\alpha - a_i\gamma - \theta_j)]^{D_i^H} * \Phi(-x_{it}\alpha - a_i\gamma - \theta_j)^{1-D_i^H} *$$
$$(\prod_{\tau=19}^{24} ([1 - \Phi(-z_{it\tau}\beta_{G,\tau} - a_i\delta_{G,\tau} - \theta_j\lambda_{G,\tau})]^{D_{i\tau}^{I,G}} * \Phi(-z_{it\tau}\beta_{G,\tau} - a_i\delta_{G,\tau} - \theta_j\lambda_{G,\tau})^{1-D_{i\tau}^{I,G}})^{O_{i\tau}})^{D_i^H} *$$
$$(\prod_{\tau=19}^{24} ([1 - \Phi(-z_{it\tau}\beta_{D,\tau} - a_i\delta_{D,\tau} - \theta_j\lambda_{D,\tau})]^{D_{i\tau}^{I,D}} * \Phi(-z_{it\tau}\beta_{D,\tau} - a_i\delta_{D,\tau} - \theta_j\lambda_{D,\tau})^{1-D_{i\tau}^{I,D}})^{O_{i\tau}})^{1-D_i^H}))$$

where:

(e) the sum of the probabilities of each mass point equals one, i.e. $\sum_{j=1}^{J} P_j = 1$,

(f) $P_j \geq 0$ for all j,

(g) J is the number of mass points used (three in our estimations), and

(h) the last mass point θ_J equals zero.

Appendix B
Construction of the Average Grade Variable (GR11F)

As discussed in the paper, we have data for each individual on grades from a series of grade 11 exams, and we need a way to aggregate the grade data in a way that makes use of the varied individual data. To do this, we make use of the grade 12 Grade Point Average (GPA), which is calculated in a standard way across the province. We have the grade 12 GPA only for individuals who graduated from grade 12 since the province only calculates it for that group. We regress the grade 12 GPA on a large set of dummy variables corresponding to five grade ranges for each of the five types of grade 11 exams upon which we have data. The regression is specified in such a way that we can construct a predicted GPA whether the individual has information on only one exam or on up to five. The exercise is complicated by the fact that the grade 12 GPA is only observed for individuals who graduated from grade 12 and because the GPA has a maximum value of 4. Thus, we estimate a doubly censored (at zero, for those individuals for whom we do not have a GPA value, and four) Tobit model relating GPA to grade 11 grades. Using the estimated coefficients from this regression we construct a preliminary fitted GPA score for every individual in our dataset based on their grade 11 grades.

The second stage of our construction addresses potential grade inflation in some schools relative to others. To address this, we use data on school specific average marks on province wide exams administered to evaluate relative school performance. Since these exams are the same across all schools, they provide a standard measure of the average ability of students in the school, independent of school specific grade inflation. To estimate the extent of school grade inflation we regress the average fitted GPA score for the school (constructed as just described) on the school's province-wide exam grade average. The residual from this regression, which we will call RESIDM, represents the extent to which a given school tends to give high or low grades relative to the average across schools. We next re-estimate the censored Tobit GPA regression including the same exam mark dummy variables as before plus the RESIDM value corresponding to the individual's high school. We then create a new fitted GPA score based on the coefficients from this newest Tobit estimation and the individual's grade 11 grades, with the RESIDM variable set to zero. This provides an average grade for each individual, GR11F, which is purged of relative grade inflation across schools.

Endnotes

1. Note that the outcome of this investigation would be useful even if the cross-generational correlation did not reflect a causal link. In that case, we would be asking whether education policy is effective in helping a set of families who are particularly needy.

2. Throughout the remainder of the paper we will use IA to denote BC's assistance system and the more generic term, welfare, to denote assistance programmes in other jurisdictions.

3. Unemployables include individuals who cannot work due to medical conditions, single parents with one dependent child under the age of six or two or more dependent children under the age of twelve, individuals over age 65, or a single parent with a disabled child.

4. In estimation, we normalize one of the θ's to zero and the variances of η_{it}, $\epsilon_{it\tau}^{G}$, and $\epsilon_{it\tau}^{D}$ to one. We estimate the probabilities of all but one of the mass points directly but form the probability of the last mass point as 1 minus the sum of the estimated probabilities associated with the other mass points.

5. Our set of variables affecting graduation is very similar to that used in Meghir(2002).

6. As stated earlier, we do not view 4) as a behavioural equation. If one were to view it as behavioural then it is likely that ω determines the education proxy variables rather than vice versa and equation 4) would be mis-specified. However, one could argue that the family intervention variables do reflect deeper factors that could causally affect ω. Thus, we could estimate the system omitting the education proxies and giving the γ parameters a behavioural interpretation. When we do this the estimated impacts are substantially higher than those obtained when we include the test score and failure proxies but, for females, lower than when we do not use any proxies at all.

7.Using the principal effects and the instrumental variables approach leads us to two further sample cuts, which we maintain throughout the remaining analysis. First, we used only principals with whom there were associated at least five students in the sample. This resulted in cutting the number of principals associated with the sample to 248 for the female sample and 241 for the male sample. We also dropped the student observations associated with the principals we cut. Second, we estimated a simple probit for high school graduation using the variables just described and formed fitted probabilities of graduation for both those who actually dropped out and those who actually graduated. We dropped high school graduate observations for individuals whose fitted probability of graduation was above the highest fitted probability observed in the drop out sample (.995) and dropped drop out observations for individuals whose fitted probability was below the lowest fitted probability observed in the graduate sample (.06). We do this in keeping with results from the matching literature arguing that reliance on observations receiving a treatment for whom there is no close match among the non-treated puts too much faith in specific functional forms. Once we drop observations for both reasons, our female sample falls from the 9272 observations mentioned earlier to 9005 and our male sample falls from 7534 to 7309.

8.This "typical" person is constructed using sample average values for the Census tract variables, the grade 11 fitted grades, and mean school graduation rate. Otherwise the person speaks English as their first language at home, lives in a census tract with average income in the middle two quantiles, has never had an intervention into the household from the province, did not fail any previous grades and is not from a First Nations background. All year dummies are set to zero and the principal dummy corresponding to the principal with an effect nearest to the median estimated principal effect is set to one.

9. Specifically, assuming that sampling error and true heterogeneity are independent, the bariance or our estimated principal effects can be written as the sum of the variance of the true heterogeneity and the sampling error variance. Fro a principal with a true graduation rate of p and with whom there is associated N sample members, the sampling variance is p(1 - p)/N. We estimate p for each principal using the fitted graduation effect described in the text. We then average over the calculated sampling variances for each principal to get an overall sampling variance estimate. Finally, we subtract that from the variance of our estimated principal effects to get our estimate of the variance of the "true" heterogeneity.

10.More specifically, we estimated models with two, three and four points of support. Moving from two to three points of support, we found that we could estimate the third point and its associated mass without problems. However, when we tried to add a fourth point of support, the location parameter assumed essentially the same value as one of the points of support when we used only three and the probability associated with that former point was split in two. We, therefore, settled on three points of support.

11.We tried a very large number of starting values for the parameter values associated with the θ_i distribution and the λ's. The starting values chosen for the θ's had no real impact on the maxima achieved in estimating the models. However, starting λ values did lead to local maxima being found. Estimates starting from negative λ values reached one maximum. Estimates starting from positive λ values reached another, while those starting from zero λ values reached a third local maximum. For females, the estimates starting from negative starting values for the λ's reached the highest maximized function value, with the final estimates for the λ's all being negative. For the male sample, the maximal function value was achieved from the estimates starting from positive starting values for the λ's. The optimized parameter estimates for the λ's here were also positive.

12. The derivatives are based on the same typical person type as in the first two columns. To address heterogeneity in θ_i, we construct three baseline probability of graduation estimates using typical person values for the covariates but with θ_i set to each of the three estimated support values. These three baseline probabilities are then combined by multiplying each one by the estimated probability associated with the particular θ_i value. The same calculations are made after, for example, changing the NATV dummy from zero to one, leaving all other variables at their baseline levels. The difference between the two weighted average probabilities is then the relevant probability derivative for the NATV variable.

13. To match the sample probabilities, we only use an individual's simulated history at a given age if that person was in our sample at that age.

14. The test statistic for females takes a value of 172.33 and that for males takes a value of 131.62. The statistics are distributed as $\chi^2(88)$, which has a critical value at the 1% significance level of 121.76.

15. The test statistic corresponding to the restriction that the complete vector of coefficients (the β's, δ's and λ's) is the same for graduates and drop outs takes a value of 168.73 for females and 125.76 for males. This test statistic is distributed as $\chi^2(94)$ for females and $\chi^2(92)$ for males. The relevant 1% critical values are 128.8 and 126.76, respectively.

16. In the simulations for the treatment on the untreated effects in the earlier figures, we drew new samples in each iteration. Here we use a single sample in which we keep actual drop-outs whose

predicted probability of dropping out is above the average predicted probability for all drop-outs. This generates a single sample that mimics the average of the multiple samples drawn to create the earlier figures. In the Full model, we actually create the fitted graduation probability three times (once for each value of θ) and then average using the estimated probabilities associated with the respective mass points. Similarly, in generating the predicted IA probabilities in Figure 6 for the full model, we ran simulations three times (again, once for each value of θ) and then averaged across them using the estimated probabilities.

Figure 1

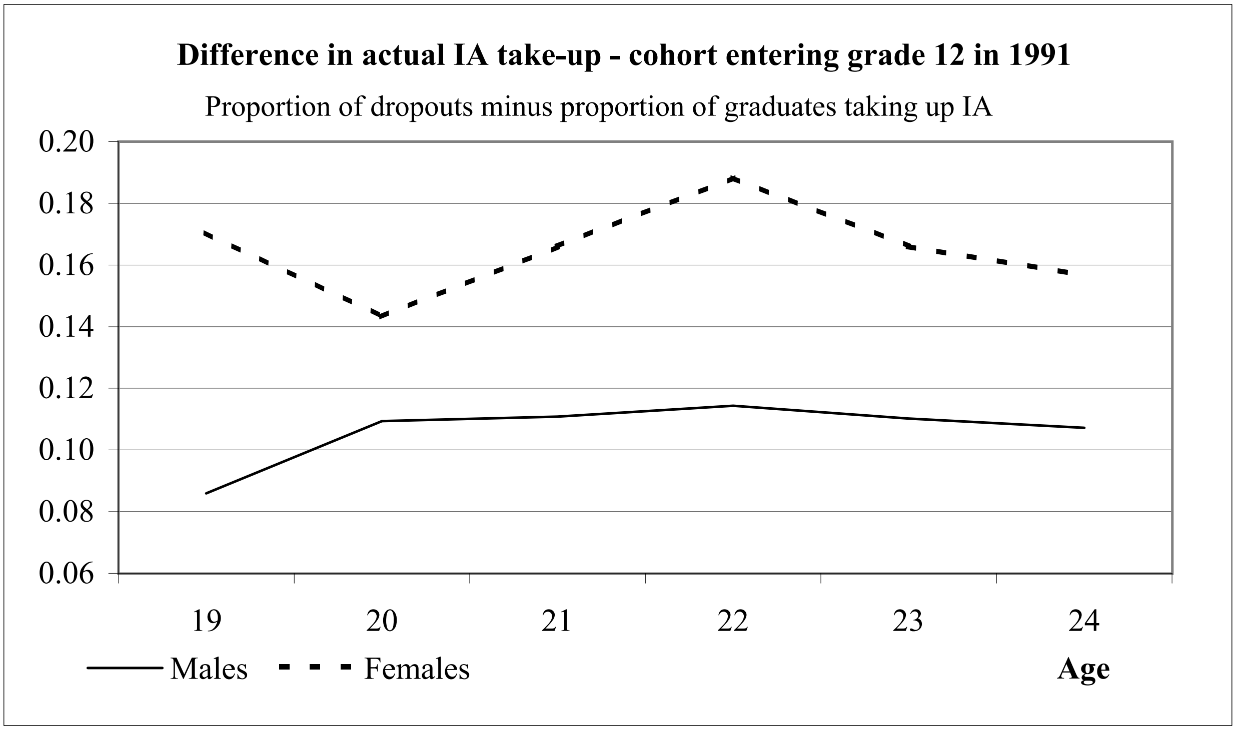

Figure 2

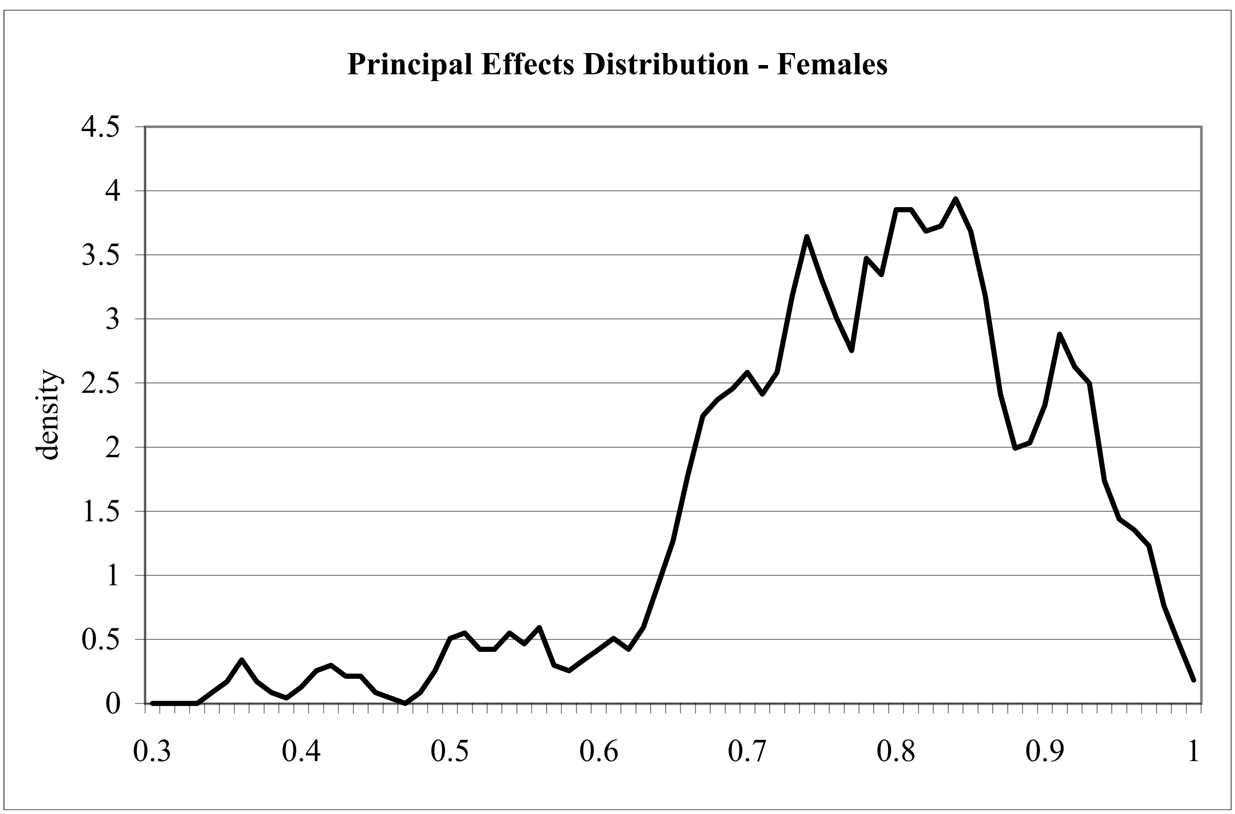

Figure 3a

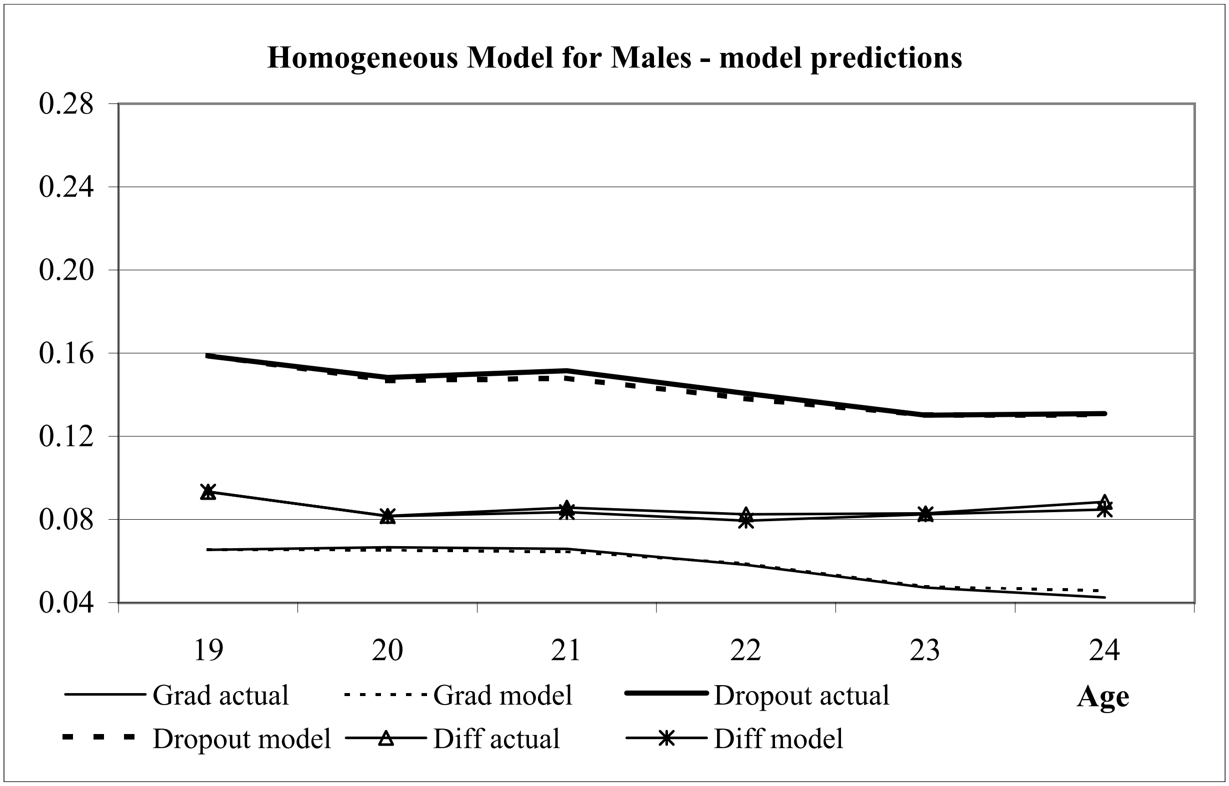

Figure 3b

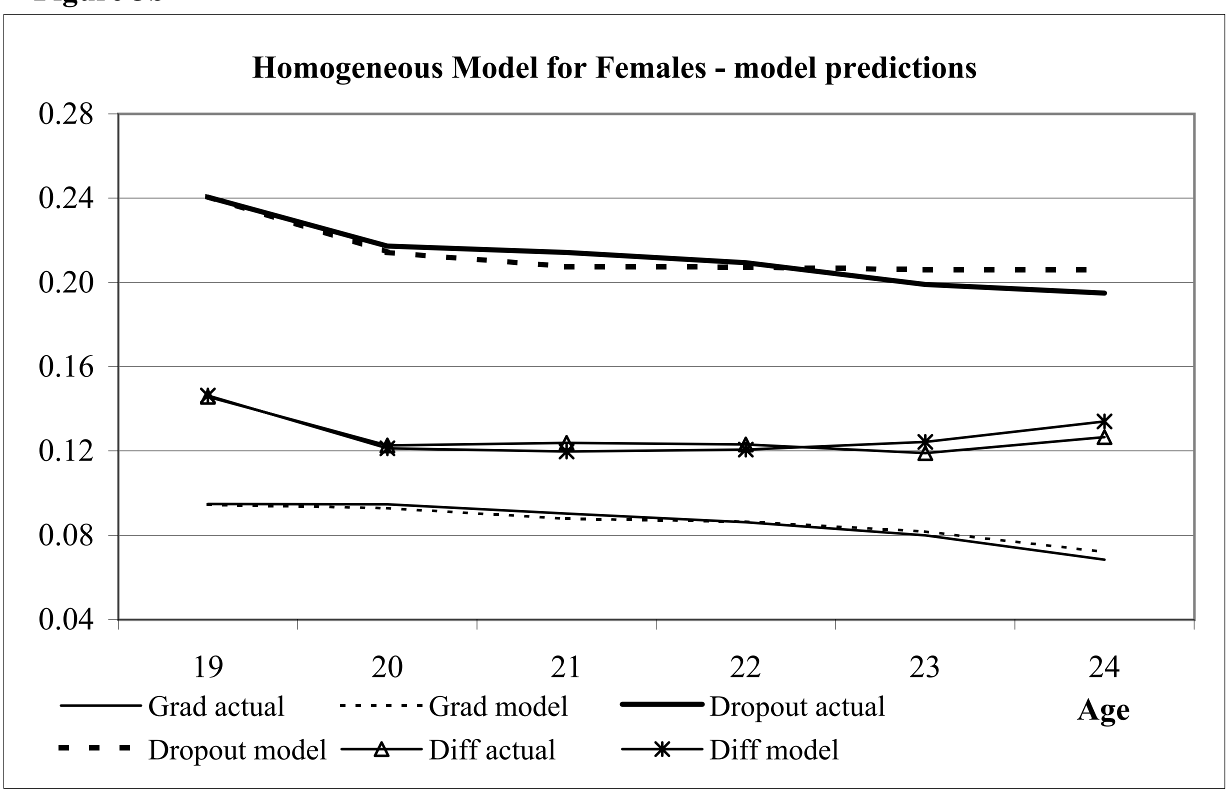

Figure 4a

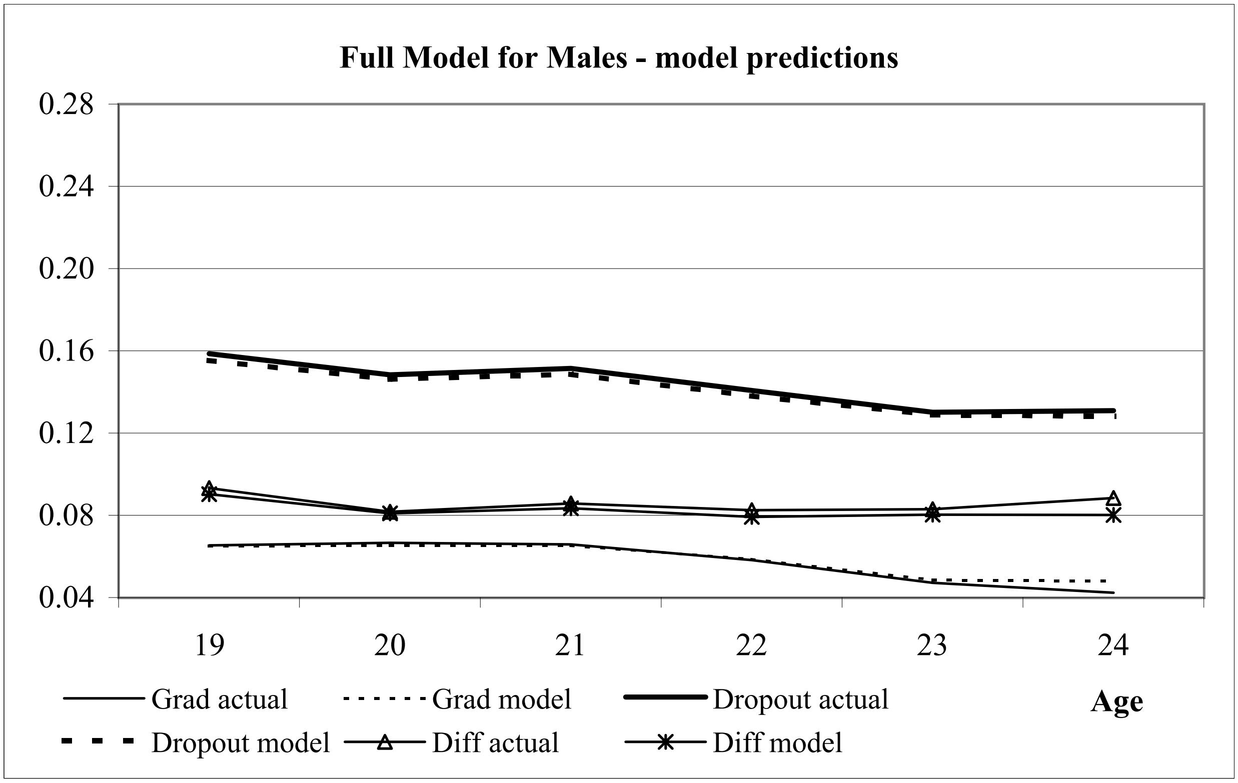

Figure 4b

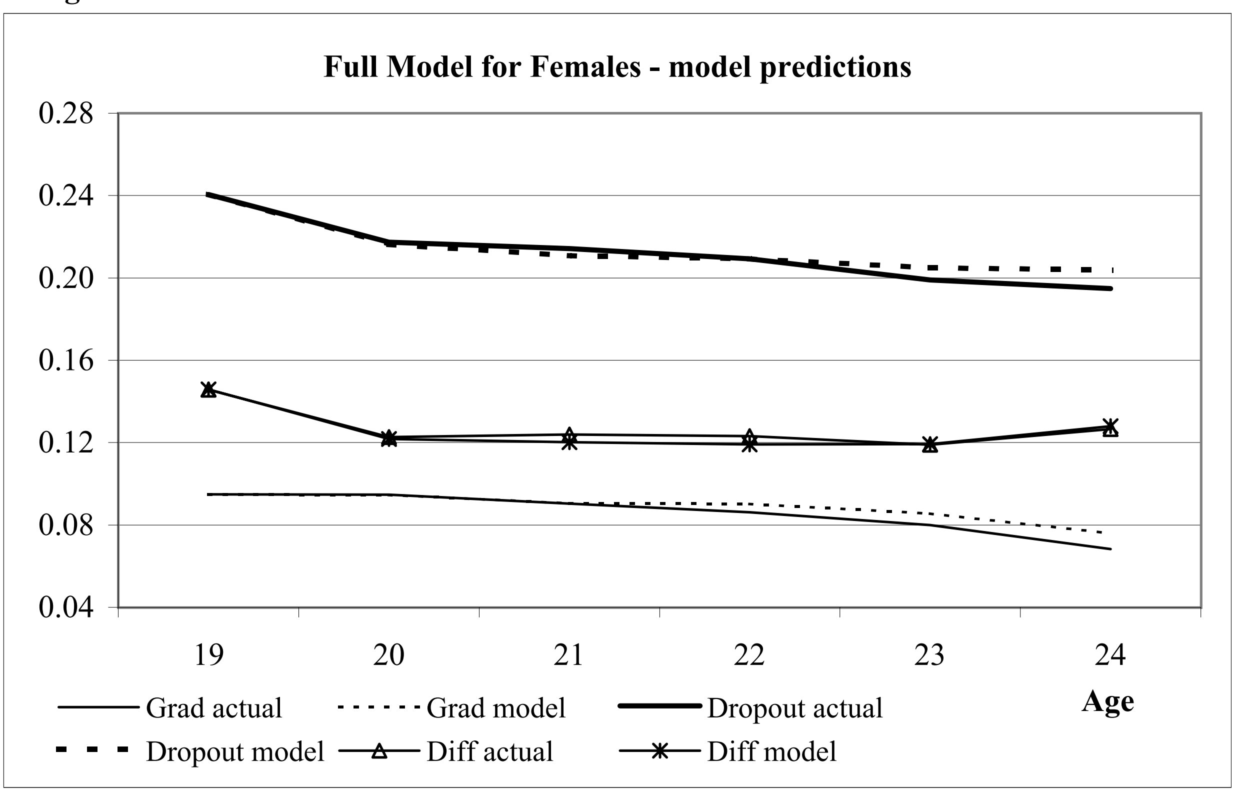

Figure 5 **MALES - estimated effects**

(time trends removed - 90% interval)

Homogeneous Model **Full Model**

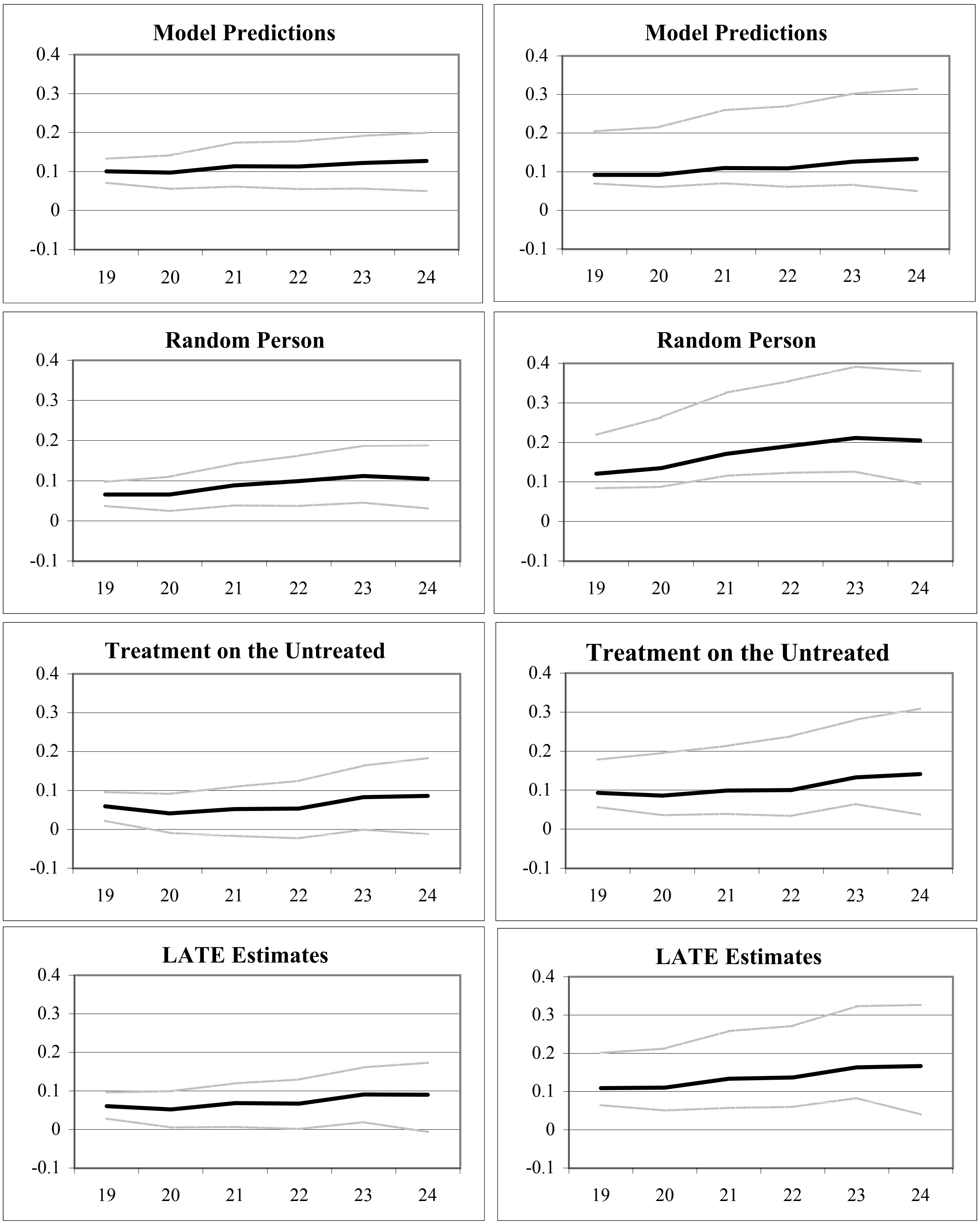

Figure 6 **FEMALES - estimated effects**

(time trends removed - 90% interval)

Homogeneous Model **Full Model**