The Economic Determinants of Truancy

Simon Burgess, Karen Gardiner and Carol Propper

Contents

CASEpaper 61
September 2002

Centre for Analysis of Social Exclusion
London School of Economics
Houghton Street
London WC2A 2AE
CASE enquiries – tel: 020 7955 6679

Centre for Analysis of Social Exclusion

The ESRC Research Centre for Analysis of Social Exclusion (CASE) was established in October 1997 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science, and benefits from support from STICERD. It is directed by Howard Glennerster, John Hills, Kathleen Kiernan, Julian Le Grand, Anne Power and Carol Propper.

Our Discussion Paper series is available free of charge. We also produce summaries of our research in CASEbriefs, and reports from various conferences and activities in CASEreports. To subscribe to the CASEpaper series, or for further information on the work of the Centre and our seminar series, please contact the Centre Administrator, Jane Dickson, on:

Telephone:	UK+20 7955 6679
Fax:	UK+20 7955 6951
Email:	j.dickson@lse.ac.uk
Web site:	http://sticerd.lse.ac.uk/Case

Editorial Note and Acknowledgements

Simon Burgess is Professor of Economics at the University of Bristol and is an associate of CASE and CEPR. Karen Gardiner is a Research Fellow in CASE and the University of Bristol. Carol Propper is Professor of Economics at the University of Bristol, an associate of CEPR and a co-director of CASE. The authors are grateful for funding and support from the ESRC Centre for Analysis of Social Exclusion at the London School of Economics. They would also like to thank participants in seminars at STICERD, LSE; Universitat Autonoma de Barcelona, Barcelona and the Leverhulme Centre for Market and Public Organisation, University of Bristol; and at the 2002 ESPE conference in Bilbao.

Abstract

Truancy is often seen as irrational behaviour on the part of school age youth. This paper takes the opposite view and models truancy as the solution to a time allocation problem in which youths derive current returns from activities that reduce time spent at school. The model is estimated using a US panel dataset, the National Longitudinal Survey of Youth 1979, and the estimation allows for the possible endogeneity of returns from these competing activities. The results show that truancy is a function of the estimated economic returns from work, crime and school.

JEL classification: I20, J20
Key words: Truancy, returns to education

Address for correspondance:
Simon Burgess
CMPO
Department of Economics
University of Bristol
Bristol BS8 1TN, UK
Tel +44 117 954 6943
Fax + 44 117 954 6997
e-mail: Simon.Burgess@bristol.ac.uk

Non-Technical Summary

Truancy is often viewed as irrational behaviour by young adults. For example, a recent report from Downing Street cited three excuses given by children for not going to school as ‘not liking Mondays’, ‘because of a spot on my nose’ and ‘because my hamster died and we had to get a new one’. However, this view is rather at odds with recent economics research that finds the behaviour of young adults to be a rational response to economic forces. For example, economic factors have been found to be important determinants of leaving home, choosing a partner and having children.

The purpose of the analysis in this paper was to examine whether truanting behaviour is a response to economic incentives. School-age youths face competing uses for their time, including attending school, working for pay, taking leisure and engaging in criminal activity among others. The act of truanting means that a youth has found other uses for her time to be more valuable than school. This is the insight that we follow in this paper, and our approach to understanding the factors that make truancy more or less likely. We set up an economic model of behaviour to study this phenomenon. The different uses of time each bring current rewards and a potential impact on future returns; these are balanced in the decision as to whether to fulfil mandatory schooling requirements or to truant. Given these rewards, and each person’s abilities, teenagers will choose how to spend their time. For some individuals the rewards from working now or engaging in crime are so large, or their return to education so low, that their school attendance drops below the officially mandated level and so they truant.

We estimate the model using a panel dataset from the US: following a cohort of individuals aged 14-21 in 1979 for the following 14 years. This is the National Longitudinal Survey of Youth, 1979 (NLSY79). Our main finding is that economic incentives do matter in determining truanting behaviour. We establish the rates of return different school-aged children would get from being in school, working and engaging in crime and then test to see whether these returns are correlated with playing truant. We find that all three returns are significantly associated with truancy. Those who had higher expected returns from studying were more likely to be in school, whilst those who could command higher returns in the labour market, or who were in areas where the gains from

crime were greater, skipped more school. Other factors, such a family background, also explain truanting behaviour, but the social factors do not wipe out the impact of economic returns.

Our analysis is innovative in that it takes a structural approach to analyse the economic determinants of truancy. It also allows for the influence of a broad range of factors to determine truanting behaviour, beyond that of just the individual and their family. We take in to account the characteristics of the school environment, the local area where the respondent resides and labour market indicators at the state level. These environmental factors are found to be important in estimating our model of truancy.

These findings offer some guide for policy, and support for the current government actions with respect to encouraging young people to stay at school longer. If individuals do truant because the perceived returns from other uses of time is greater than the perceived gain from school, then what is needed is to raise the relative returns from being in school. The government's Educational Maintenance Allowance, which 'pays' young people to stay in school, does just that.

1. Introduction

Governments mandate attendance at high school because there are returns, private and social, from education. Yet a substantial minority of youths absent themselves from school. This may be seen as irrational behaviour: these individuals do not understand the value of schooling.[1] But schooling, even if tuition is free, has an opportunity cost. Time spent in school prevents youths from using this time in other ways, some of which will bring returns, either in the present or in the future. Skipping school can therefore be seen as a rational response of youths to the perceived returns from education compared to the perceived returns from other activities that can be undertaken during school hours. If this is the case, the appropriate policy response is rather different to the policy response that would be warranted if skipping school were simply willful teenage behaviour. If skipping school is because the outside options are better than the option of going to school, then policy needs to be directed towards changing the relative value of these options.

The approach taken in this paper tests the view that truancy is the result of responses to the relative returns from schooling. We put forward a simple economic model of time allocation to various competing activities, one of which is school attendance. We estimate the expected returns for each youth of these various activities and test whether they explain any of the decisions of youths to skip school.

In such a framework, we need to identify the activities youths can undertake whilst being of mandatory school age. Spending time in school is one activity. It brings a later return, which arises because of the correlation between education and the returns from the labour market, but may also bring current returns; for example, from involvement in social activities. Time spent out of school may be used in a variety of ways. Being in paid work is one obvious choice of use of time, crime may be another. Both bring a current return, and both may bring some future return. Young individuals may also be involved in caring

1 A recent UK government survey of reasons for unauthorised absence from school included 'not liking Mondays', 'because of a spot on my nose' and 'my hamster died and I need to buy a new one' (Downing St, 2002). These responses perhaps provide support for common perceptions that truanting is irrational behaviour.

activities for older or younger family members. Finally, they may take leisure. Here we focus on the first two of these activities as substitutes for schooling, and classify the rest as leisure. The model of behaviour we postulate is that those individuals with relatively greater expected returns to work and crime will be induced to spend more time on these activities and therefore exhibit a higher tendency to truant. So truancy (the converse of school attendance) is a function of the returns to school, work and crime. Over and above these three economic parameters we also allow for the influence of other preferences and constraints on the behaviour of school-aged youth.

In our approach we attempt to deal with unobserved heterogeneity and the endogeneity bias that this induces. The decision to truant will be associated with unobserved factors that are correlated with the returns from the various activities the youth may engage in, so estimates of truancy as a function of the actual time spent in work or the actual return will be biased. The solution we adopt is to instrument the returns from schooling, work and crime, exploiting geographical variation in local labour market conditions to help identify the returns from activities other than schooling. Under the standard assumptions that the fitted rates of return are orthogonal to the unobserved heterogeneity (including tastes) conditional on the variables included in the rates of return estimation, the use of fitted values removes the endogeneity bias.

Our analysis uses the National Longitudinal Survey of Youth (NLSY79), a rich panel dataset that follows a cohort of American youth and includes information on the individual, their family, school and local area. This provides a diverse set of variables to be employed as instruments and predictors of the returns variables and has been widely used to study the school to work transition of American youth born in the late 1950s and early 1960s.

Related studies have examined the relationship between truancy and work whilst at school, and the impact of working whilst in school on later wages and labour market outcomes and have sought to disentangle the relative impact of economic factors and heterogeneity. Eckstein and Wolpin (1999) model the determinants of high school[2] drop out also using the NLSY79. This captures a similar decision-making process to

2 'High school' in the US education system corresponds to a UK secondary school.

the one modelled here, but the outcome they examine is the permanent decision to quit school, whereas we focus on the day-to-day decision to attend school. They estimate a dynamic model of high school attendance and work decisions, which assumes that youths choose among work-school combinations in order to maximise their expected lifetime utility at each decision period. They control for both observed and unobserved heterogeneity and find that working whilst enrolled causes a small reduction in school performance (in terms of grades). However, the impact of heterogeneity is large and they estimate that preventing youths from working would only marginally improve graduation rates. When they explore the characteristics of those who drop out they find that these individuals tend to have low ability and motivation and a lower expected value from gaining a high school diploma. If we draw a parallel between drop out and truancy, this suggests that expected returns to education play a role in determining school attendance. It also suggests that heterogeneity, observed and unobserved, play an important role.

Dustmann *et al* (1997) focus on the link between working part time whilst in school and truancy in the UK. They examine two issues. First, they examine whether truancy amongst 16 year olds is associated with longer hours of work by those still in full time education. Second, they examine the extent to which the wages of 16 year olds vary across individuals, and attempt to find reasons for observed differences in wages. They note that the three variables they seek to model – working part time whilst still in education, the wages received for that work, and truancy – are all likely to be related to each other. To allow for the impact of heterogeneity, they jointly estimate truancy and part time labour supply, and separately wage rates and part time work. They find that hours of part-time work are positively related to the decision to truant, but only for females. In the context of our approach, we hypothesize that truancy will be positively related to the returns to work, which are measured in terms of log hourly wages rather than hours, but on the assumption that hours and wages are positively related, the findings that hours are positively related to truancy gives some support to our approach. The other factors that they find to be significant in explaining truanting behaviour are the respondent's ability, parental characteristics and school type. In the context of our approach we can interpret these findings as evidence that those who have higher expected returns to education truant less and that family background is also likely to have a direct impact on attitudes towards truancy.

A larger literature examines the decisions of youths to work whilst enrolled at school. In this literature one focus has been to establish whether there is a positive return to high school employment in terms of future job and earning prospects. A series of papers have examined this issue using the NLSY79. Ruhm (1997) and Light (1995) conclude that working whilst at school brings later advantage in the labour market. In contrast, Hotz *et al* (2002) conclude that the positive effect disappears once corrections for heterogeneity and selectivity biases are included. On the basis of their evidence, Hotz *et al* argue that even with the rich set of controls available in the NLSY79 allowing for only observed heterogeneity will not eliminate the endogeneity bias due to unobserved heterogeneity. Whilst these papers have focused on later returns and our focus is on the *current* returns to working whilst enrolled, the results suggest both the need to control for unobserved heterogeneity and that the primary return from work whilst in school is current income. Finally, Cameron and Heckman (2001) examine the determinants of college attendance amongst males using the NLSY79. Using a dynamic approach they again show the importance of family background in determining this decision, but find a smaller role for financial factors.

We conclude that while the focus of all these papers is related to the present paper, none of them seek directly to estimate a structural model of truancy. Further, the approach to dealing with the important issue of heterogeneity differs from that which we use here. The rest of the paper is structured as followed. We begin with a description of truancy and its correlates from the NLSY79 in terms of time use in section two. This is followed by a detailed description of the data used in the empirical analysis in section three. Section four presents the conceptual framework of our structural model, as well as the estimation strategy. The results are discussed in section five. Section six concludes.

2. Truancy in the NLSY79

We begin with an examination of the relationships between truancy and other activities that school age youth may spend their time doing.[3] We consider paid work and illegal activity, since these are the key substitute activities that we analyse, but we also examine illness and caring responsibilities to the extent the data permit.

The gender breakdown of the number of days truanted in Table 2.1 shows that, in general, women are less likely to truant than men. The differences across the sexes are most pronounced at the lowest and highest levels of truanting. Of women, 55% never truanted in the last year, compared to 51% of men. At the other extreme, only 1% of women truanted more than 51 days, whereas 3% of men fall in to the highest truanting category.

Table 2.1: Truanting behaviour by gender

Number of days truanted in the last year	Men	Women	Total
0	1005 (50.6%)	1043 (54.7%)	2048
1	241 (12.1%)	252 (13.2%)	493
2	182 (9.2%)	169 (8.9%)	351
3 to 5	259 (13.0%)	219 (11.5%)	478
6 to 10	147 (7.4%)	113 (5.9%)	260
11 to 50	97 (4.9%)	90 (4.7%)	187
51 +	57 (2.9%)	22 (1.2%)	79
Total	1988 (51.0%)	1908 (49.0%)	3896

3 See section 3 for further details on the truanting variable. Since this refers to the year prior to the 1980 interview, other variables in the tables relate to the corresponding period. The samples used in these descriptive analyses are, in each case, the maximum number of observations with non-missing values. This varies across the characteristics considered, with the actual numbers indicated in the tables. Throughout the empirical analysis in the paper it is assumed that missing values are randomly distributed.

Table 2.2 presents the breakdowns of truanting by racial groups and shows quite starkly that blacks are the least likely to play truant. These data also suggest that hispanics generally have a greater tendency to truant than whites, although it should be noted that the number of hispanic individuals in the sample is rather small.

Table 2.2: Truanting behaviour and race

Number of days truanted in the last year	Hispanic	Black	White	Total
0	284 (41.8%)	639 (63.3%)	1125 (50.9%)	2048
1	109 (16.1%)	129 (12.8%)	255 (11.6%)	493
2	77 (11.3%)	79 (7.8%)	195 (8.8%)	351
3 to 5	98 (14.4%)	99 (9.8%)	281 (12.7%)	478
6 to 10	54 (8.0%)	35 (3.5%)	171 (7.7%)	260
11 to 50	43 (6.3%)	14 (1.4%)	130 (5.9%)	187
51 +	14 (2.1%)	14 (1.4%)	51 (2.3%)	79
Total	679 (17.4%)	1009 (25.9%)	2208 (56.7%)	3896

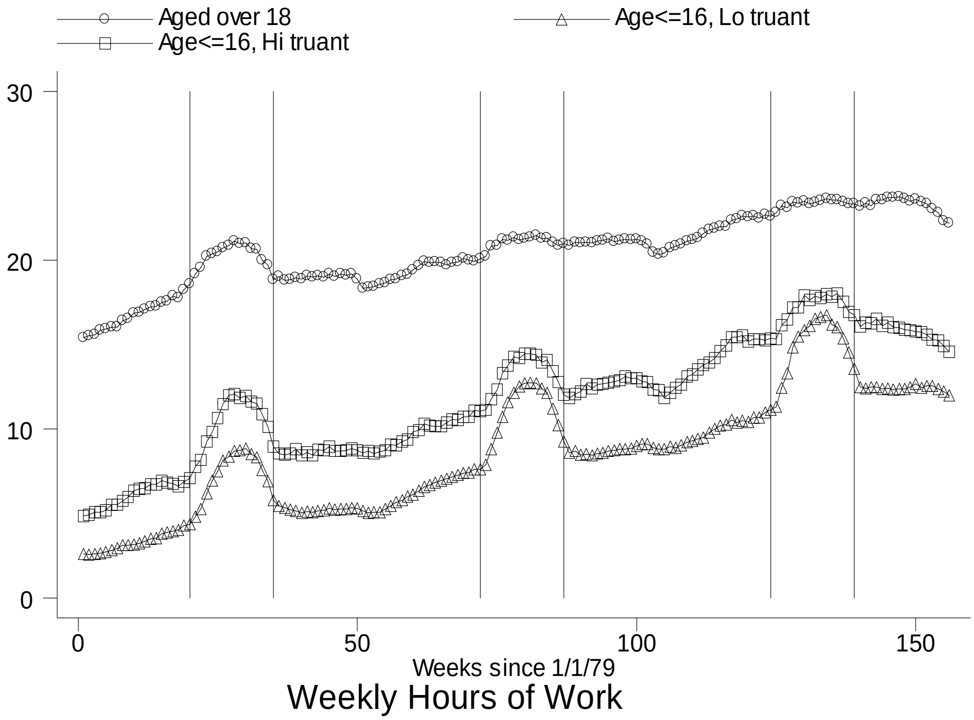
Table 2.3 shows the average annual hours worked during term time by the amount of truanting in the last year. This table shows that mean hours of work increases with truanting, except for the top category. A normalised measure of variation indicates that the variation in hours worked during term time does not increase with the extent of truanting except for the top category. This would suggest that paid work is an activity positively associated with truanting but this relationship is weaker for those who truant the most.

Table 2.3: Truanting behaviour and paid work during term time

Number of days truanted in the last year	Annual hours worked during term time			
	Mean	Standard deviation	Coefficient of variation (std dev / mean)	Number of respondents
0	242.499	346.765	1.43	1,852
1	276.719	331.928	1.20	438
2	341.842	368.823	1.08	310
3 to 5	388.695	447.739	1.15	403
6 to 10	361.067	394.801	1.09	210
11 to 50	457.883	465.455	1.02	120
51 +	361.860	510.666	1.41	43
Total	290.064	375.681	1.30	3,376

It is possible to examine the distribution of paid work over the course of the academic year, and how this varies for different groups of truants. Figure 2.1 shows the distribution of weekly hours of paid work for three groups: those aged over 18, those aged 16 or under who truanted only two days or less in the last year ('lo truant'), and those aged 16 or under who truant more ('hi truant'). The horizontal axis tracks the weeks starting from the beginning of 1979. The vertical bars identify the extended summer vacation when high schools are closed. The graph shows that the older group work substantially more hours per week than the two younger age groups, as we would expect. Among those aged 16 or below, the 'hi truant' group on average spend more hours engaged in paid work, with the difference being greatest during term time.

Figure 2.1: Weekly hours of paid work



Note:
1. Age<= 16 means aged 14 – 16 in 1979. Sample sizes: 5,469 respondents aged over 18; 2,834 respondents aged 16 or under in the 'lo truant' group; 1,244 respondents aged 16 or under in the 'hi truant' group

In addition to paid work, we examine the returns to crime as a determinant of truancy. There are a number of reasons why there may be a link between illegal activity and truancy. Firstly, truancy itself is an illegal activity and hence those who commit one kind of illegal act may also have a greater tendency to commit others. Secondly both criminal activity and schooling may generate returns and a rational agent will decide how to allocate their time between competing activities on the basis of these returns.

There are a number of variables in the NLSY79 which relate to illegal activity and are collected annually, namely, share of income from illegal sources, number of times stopped by the police, number of times charged by the police and the number of physical fights at work or school (all are annual measures). Below results are only presented for illegal income and number of fights, due to large numbers of missing values for the other variables.

The categorical variable for income from crime is defined as the share of total income in the last year that came from illegal sources[4]. Table 2.4 indicates a strong monotonic relationship between the average share of income from illegal sources and the number of truant days. There is also a clear association between the average number of fights and the tendency to truant. (Table 2.5)

Illegal activity is more often engaged in by young males than their female peers. Equivalent activities for young women, in terms of receiving social disapproval, are perhaps teenage sex and pregnancy. From the data available in the NLSY79, it is possible to tell whether young people have their own children living in their household. Using this rather crude measure (we are not able to identify pregnancy) there does not seem to be a link between truancy and the presence of children in the household. A related and possibly female specific reason to be away from school is caring responsibilities for young or elderly relatives. Unfortunately the data do not make it possible to explore the empirical support for such behaviour.

Table 2.4: Truanting behaviour and share* of income from illegal sources

Number of days truanted in the last year	Share of income from illegal sources		
	Mean	Standard deviation	Number of respondents
0	0.204	0.675	1,976
1	0.305	0.724	469
2	0.383	0.799	345
3 to 5	0.394	0.834	465
6 to 10	0.548	0.962	252
11 to 50	0.753	1.102	182
51 +	1.260	1.650	73
Total	0.326	0.812	3,762

*Note that the "Share of income from illegal sources" is not a continuous variable from 0 to 1, but a categorical variable, with values ranging from 0 "None", 1 "Very little" to 5 "Almost all".

4 The raw variable ranges from 1 (no illegal income) to 6 (almost all income in last year from illegal income); it has been recoded so that the value range is instead 0 to 5.

Table 2.5: Truanting behaviour and number of fights

Number of days truanted in the last year	Number of fights		
	Mean	Standard deviation	Number of respondents
0	0.494	0.955	2,045
1	0.677	1.067	493
2	0.903	1.217	351
3 to 5	1.099	1.366	477
6 to 10	1.432	1.522	259
11 to 50	1.695	1.703	187
51 +	2.418	1.959	79
Total	0.787	1.239	3,891

Strictly speaking, absence from school due to illness is not counted as truancy, as it is a valid reason to be away from school. However, we may still find some observed relationship between truancy and illness. For example, pupils who have consistently experienced unavoidable absence from school due to illness may be unable to reap the same returns from education as their more healthy peers. In this case, these individuals may figure that skipping school does not further worsen their chances of academic success, and hence imposes little economic cost. We are limited in our ability to analyse this issue in that the NLSY79 is more likely to pick up recurrent health problems than short-term sickness, since the only relevant question asks about ‘health limitations’ in the past year. Tabulations of the raw data on illness by amount of truanting do not indicate any association.

This description of the raw data has suggested that truanting is linked to participation in both paid work and illegal activity. Given data limitations it is not possible to establish whether the same is true for pregnancy, caring responsibilities or illness. In our economic model, we therefore focus on the crime and current paid employment as the alternatives to full time schooling and do not attempt to measure returns from other possible uses of time of adolescent youth.

3. Data

3.1 The NLSY79

The primary data we use are from the NLSY79[5]. This is a panel dataset running from 1979 containing data on 12,686 people, a representative sample of those aged 14 to 22 in the United States in 1979. Sample members were interviewed annually until 1994 and every other year thereafter; we use data through to 1996. The NLSY79 is an extremely rich dataset. Compared to many longitudinal datasets which follow individuals over time it has the advantage that it does not restrict itself to collecting information on only the individual and their family. Arguably it comprehensively covers the key societal and institutional influences on the development of the adolescents in the sample. In addition to recording characteristics of the young people and their family members, it provides data on the school they attend, and the area where they live. Not only does this provide direct information on the environment within which the adolescents are growing up, but also provides insights into peer group influences such as the truancy rate of their school and the average family income in the area where they live. To provide additional identification of the returns from work, we supplement the NLSY79 with state-level information from the Current Population Survey on regional labour markets.

There are two types of variables that relate to truancy and school absence. Firstly, there is data on the number of days the respondent skipped school in the last year with no real excuse[6]. This is asked of respondents aged 17 or under in 1980. This variable is most close to notions of what we usually mean by 'truanting', in the sense of unauthorised absence. Secondly, there is data on the actual number of days absent in a particular grade year, as recorded in school transcripts. There are four separate variables, one for each grade from grades 9 to 12. The information was collected in 1980 but refers to the year when the pupil was enrolled in the relevant grade. This covers a broader definition of absence, both authorised and unauthorised and, notably, includes absence due to sickness. Whilst taken from official school transcripts, there is relatively low correlation between this variable and

5 For more information see Bureau of Labor Statistics (2000).

6 The variable is banded in to 7 categories: never, once, twice, 3-5, 6-10, 11-50, and 51 times or more.

unauthorised absence (see Appendix for more details). Given that days absent includes sickness, we focus therefore on the self-reported variable of the number of days that the pupil skipped school with no real excuse.

Details and some descriptives of the other NLSY79 variables used in our empirical analysis are presented in Table 3.1 (see p.XX), broken down in to the categories of individual, family, school and local area. In addition to the basic characteristics of sex, race and age, the information about the *individual* mainly refers to their education with some data on crime (fights and illegal income), health status and earnings. In the main the variables concerning the respondent's *family* refer to characteristics of the parents: their education, work and whether they were a member of a professional occupation. Broader measures of family circumstances include whether the respondent lived with both biological parents at age 14, how may siblings they had in 1979 and whether the family lived in urban area. We are also interested in the impact of family resources and so analyse total family income in 1980, by using four dummy variables to represent quartile groups of family income. We also identify individuals not living with at least one biological parent at the time.

The NLSY79 also provides information about the *high school* attended by the respondent. As a measure of school quality we use the percentage of teachers with a master's degree or above. The other school indicator we use is the attendance rate at the school, which is likely to reflect both school quality and pupil composition.

The definition of *local geographical area* for which we have information in the NLSY79 is the county of residence in the US.[7] In Table 3.1 we present information for three county level variables we use – the unemployment rate, crime rate and average family income – as well as an indicator for the state where the respondent was resident.

3.2 The Current Population Survey

To estimate rates of return we use information on state labour market conditions, as captured by relevant wage rates. Our analysis requires that we compute both a youth and adult wage rate. Estimates were

---

7 We find some county level variables are found to have a significant role in explaining truanting behaviour but it should be noted that the individual's county is likely to be too aggregated an area to capture all potential influences of their immediate neighbourhood.

produced using the Current Population Survey (CPS) dataset, and then merged in for the NLSY79 sample on the basis of the relevant characteristics. The wage rates were estimated using pooled CPS data for the months April to November of years 1980 to 1993. The youth wage rate was calculated as the average of log real hourly earnings for those aged 18 or under (213,816 observations), by state, sex and race.

The adult wage rates are predicted values of log real hourly earnings from separate regressions by sex and race for those aged between 21 and 39 and not in education (773,248 observations). The independent variables included in these regressions were age, age squared, year dummies, state dummies and a dummy for residence in an urban area. Matching on these characteristics, our NLSY79 sample were assigned the relevant wage rates from the CPS.

4. Conceptual framework

We model the truancy decision as an optimal time allocation problem. Individuals divide their time between competing uses, of which schooling is one, and will truant if their optimal schooling time is below that legally required. We model this as a once-off decision, rather than a repeated dynamic choice.

4.1 The Model

The activities that we focus on are education (*e*), work (*w*), crime (*c*) and leisure (*l*). Time spent on these are denoted t_j, $j = e, w, c, l$. These are exclusive and exhaustive and so sum to the total available time. Each activity other than leisure generates a return for individual *i* per unit time, denoted r_{ji}, which can be split into current immediate income, p_{ji}, and an impact on future income, π_{ji}, $r_{ji} = p_{ji} + \pi_{ji}$. We assume that the youth has a utility function $U(z, t_l, f, \varepsilon)$, where z denotes her current resources, t_l leisure time, f future potential income, and ε tastes. Current resources depend on the youth's own earnings and any income from crime, and also on family income, y_i:

$$z_i = z(p_{wi}t_{wi} + p_{ci}t_{ci}, y_i).$$

Future potential income depends on characteristics of the youth's family background, b_i, and the investment components of current time allocation:

$f_i = f(\pi_{ei}t_{ei}, \pi_{wi}t_{wi}, \pi_{ci}t_{ci}, b_i)$

The individual maximises utility subject to the time budget. The solution is an optimal allocation of time across activities:

$$t^*_{ji} = t(p_{ei}, p_{wi}, p_{ci}, \pi_{ei}, \pi_{wi}, \pi_{ci}, y_i, b_i, \varepsilon_i) \qquad j = e, w, c, l.$$

We focus particularly on the time spent on schooling, t^*_{ei}. We define an individual as being a truant if $t^*_{ei} < \bar{t}_i$, where $\bar{t}_i$ is the mandated amount of schooling (depending on *i* because it may vary by location). The extent of truanting, i.e. days missed, is given by $dm_i = \max(\bar{t}_i - t^*_{ei}, 0)$.

4.2 Operationalising the model

Each of the three rates of return represent the expected present value of an hour spent on the relevant activity, where this return may generate utility at the present time or in the future. The key element of the returns to education is likely to be the effect on future earning power, but one can also think of other returns such as the current utility from acquiring knowledge. Similarly, with the returns to work, there may be immediate pay offs, as in the wages earned, or future benefits, such as the work experience and skills gained. The gains from criminal activity will be the illegal income generated at the current time but there may also be implications for the future (even if that is a negative return in terms of the greater risk of ending up in prison). The rates of return to the activities are not directly observable so we need to make some assumptions about their nature. We adopt the assumptions outlined in Table 4.1.

Table 4.1: Assumptions on rates of return

Education	Current immediate income, p_{ei}	Assumed zero
	Impact on future income, $\bullet_{ei}$	Measured rate of return to schooling
Work	Current immediate income, p_{wi}	Predicted current wage rate
	Impact on future income, $\bullet_{wi}$	Assumed zero
Crime	Current immediate income, p_{ci}	Predicted current earnings from crime
	Impact on future income, $\bullet_{ci}$	Assumed zero

The justifications for these are as follows. First, we assume any current consumption benefits from education are part of the taste variables. Second, based on recent evidence (Hotz *et al*, 2002), we assume the future returns from work in school are negligible relative to the current ones. Third, we assume that current time spent on crime has no effect on future earnings. It may be the case that current time spent on crime may affect future earnings negatively (due to detachment from the labour market and the risk of time in prison) or positively (due to the skills and experience gained at committing crime which may increase illegal income in the future). Hence, we make the simplifying assumption that future returns are zero.

Thus the final model that we estimate for days of school missed through truanting is:

$$dm_i = \max\left(\bar{t}_i - t(\pi_{ei}, p_{wi}, p_{ci}, y_i, b_i, \varepsilon_i), 0\right) \tag{1}$$

From the model, our priors on the role of the variables are that truancy should be decreasing in the rate of return on education, family income, and any family background characteristics that raise future potential income, and should be increasing in the rates of return to work and crime, and any family background characteristics that reduce future potential income. We discuss the estimation of the rates of return below.

4.3 Heterogeneity

Variations in individuals' truancy across our sample are driven by heterogeneity in economic returns, circumstances and tastes. We can use the NLSY79 data to parameterise some of the heterogeneity in returns, but some idiosyncratic comparative advantages in schooling, work or crime are unobserved. Similarly, differences in tastes are unobserved. The studies referred to in the Introduction, which use the NLSY79, show that it is important to allow for unobserved as well as observed heterogeneity. To allow for this, we instrument the rates of return. From the data we estimate the rates of return to education, work whilst in school and crime whilst in school. We then use the fitted rates of return in our estimates of the determinants of truancy.

This approach means that the idiosyncratic components of the rates of return go into the error term, having made the usual assumption that the fitted rate of return is orthogonal to the idiosyncratic component conditional on the variables included in the rate of return estimation. We also assume that tastes are not correlated with fitted rates of return. Of course, the appropriate theoretical rate of return is the expected rate of return from the standpoint of a 14 year old. So we also assume that the *ex post* observed rate of return for education is equal in expectation to that rate, i.e. is not affected by any decisions the individual took. This is possibly quite a strong assumption.

4.4. Estimation of rates of return

What is required is an *ex ante* measure of the returns. To produce such a measure we use the available sample of individuals engaged in each of the three activities and estimate how returns vary according to their characteristics. It is then possible to predict for each individual in the full sample their expected returns to each activity given their characteristics, whether or not they are actually observed to participate. A range of instrumental variables are used to ensure exogenous variation in these estimates.

For these parameters to have any purchase in explaining truanting behaviour the returns need to be estimated so as to allow for maximum heterogeneity across the individuals in the sample. We are able to exploit the richness of our data source to use a range of characteristics relating to the individual, their family, school and area in order to ensure significant variation in the estimates of returns to the three activities.

We now deal with the specifics of each of the parameters in turn. The returns to education (π_{ei}) are estimated to be the expected future returns from achieving a specified level of education. It is necessary to take this approach since everyone in the sample attends school and even those who truant will acquire human capital as a result of their schooling and therefore reap some economic return. Hence, to capture the trade-offs faced by potential truants we want to include in the model the returns to achieving a certain level of education, which may be jeopardised if school attendance is too low. We focus on the returns to post-high school education and estimate these from the sample who have completed their education and are employed in the adult labour market; we calculate the returns in terms of the log of real hourly earnings[8]. Hence we estimate:

8 Ideally we would want the returns to each of the three main activities (education, work and crime) to be of the same units. In practice, it is not possible with our data to be consistent across the estimated parameters. The

$$\ln\,(real\ hourly\ earnings)_{ei} = \alpha_1 z_i + \alpha_2 e_i + \alpha_3 z_i{*}e_i + \alpha_4 x_i + \xi_{1i} \qquad (2)$$

for those aged 21 or above and not in education, where z_i are fixed characteristics which affect wage rates and are available for the sample who are in education as well as those who have completed their education[9]; e_i is a dummy variable indicating whether individual *i* undertook post high school education[10]; and x_i is a set of variables which are likely to influence wage rates but are not relevant for calculating predicted returns for the group who are currently at high school[11]. Since the purpose of equation (2) is to predict the returns from high school education for school-aged youths, only the z_i variables are interacted with the dummy education variable e_i.

The results are then used to predict the returns that would be generated for the individuals still currently attending school if they were to go on to post high school education, in terms of the impact on log hourly earnings[12]:

$$\pi_{ei} = \hat{\alpha}_3 z_i + \hat{\alpha}_2 \qquad (3)$$

---

returns to school and work are defined in terms of their predicted effect on log hourly earnings. For the returns to crime, the available data enable us to estimate the predicted share of income that comes from illegal sources.

9 The variables we use are gender, race, quartiles of AFQT score normalised by age, whether took remedial maths, percentage of teachers in respondent's school with master's degree or above, parents' education, parents not working, lived with biological parents at age 14, whether ever had health limitations whilst at high school, and county unemployment rate in 1980. All these z_i variables are also included separately as regressors as well as interacted with the variable for post high school education, as shown in equation (2).

10 More precisely, we identify whether the individual has achieved grade 13 or above.

11 The variables in this category are age dummies, year dummies, predicted state level wage rate from the CPS and whether the respondent's family live in urban area.

12 Of the z_i variables, those which uniquely identify the returns to education are quartiles of AFQT score normalised by age, whether took remedial maths, percentage of teachers in respondent's school with master's degree or above, parents not working and whether ever had health limitations whilst at high school.

Note that none of the z_i variables are time varying and are all available for high school students. They include characteristics of the individual such as a measure of their ability (AFQT score), various family background characteristics, an indicator of school quality and the local unemployment rate. The range of z_i variables ensure that our estimates of the predicted returns to post school education vary over the sample of school-aged youths.

A similar approach is used to estimate and then predict the returns to work. Since not everyone engages in paid work whilst they are still enrolled in school, we estimate the returns from this activity using the wages earned by those in the sample who are currently employed, aged 18 or under and still attending school.[13]

Given that we estimate only the current returns from working whilst enrolled (p_{wi}), we use the log of current hourly earnings for those still at high school and regress these on a set of characteristics x_{2i}[14]:

$$\textit{ln (real hourly earnings)}_{wi} = \beta x_{2i} + \xi_{2i} \qquad (4)$$

Again we allow for a range of individual and other characteristics. These include the state level youth wage rate as a measure of labour market conditions, which identifies these estimates. We then estimate the returns to work for all school-aged youths to be the predicted log hourly wage from this equation[15]:

$$p_{wi} = \hat{\beta}\ x_{2i} \qquad (5)$$

---

13 We have not controlled for the possible selection bias in the returns to work arising from the fact that we estimate returns from the possibly non-random subsample of individuals who actually engage in paid work while enrolled. If errors in the return to work are correlated with the errors in the truancy decision, then this censoring may induce bias in our estimates. We cannot *a priori* sign this bias as we don't know the correlation between the variables in the returns equation and in the unobserved component. Selectivity bias is not an issue for the returns to education and crime since these are estimated from the full sample.

14 The set of variables x_{2i} include the youth wage rate, year, race, gender, age, whether the family lived in an urban area and the county unemployment rate.

15 In predicting the returns to work the time-varying variables take their 1980 values since this corresponds to the period to which the truancy data refer.

Similarly, we estimate the current returns to crime from the share of income from illegal activity in 1980, as[16]:

$$\textit{Share of income from illegal activity}_{ci} = \delta x_{3i} + \xi_{3i} \qquad (6)$$

The factors that affect the potential gains from crime, $x_{3i,}$ capture the opportunities for illegal activity in the area where the individual resides, such as the county crime rate. The variables that identify the returns from crime are the country crime rate and the average county income. The results from equation (6) are then used to predict the returns from crime for the full sample[17]:

$$p_{ci} = \hat{\delta}\ x_{3i} \qquad (7)$$

We use the estimates (3), (5) and (7) in the estimation of (1). The dependent variable is the truanting variable described above, the number of days an individual skipped school in the last year with no real excuse. In addition to the predicted rates of return, we also allow for the impact of family resources (y_i), observed factors measuring tastes / preferences (k_i) for truanting[18] and the minimum mandated school attendance (t_i^-). Family resources may act as a budget constraint and influence the need for financial returns in the present (from work or crime) rather than the future (when returns from schooling pay off); we measure this using family income in 1980. We use a dummy variable to identify those students who were not living with at least one parent in 1980, since this group may not have had a full claim on family resources. For other students we identify the relevant quartile group of family income in 1980. The minimum mandated school attendance (ti^-) varies across states of the US; we allow for this by having state dummies in the truancy regression. These state dummies may also capture other state-

---

16 This equation is estimated as an ordered probit of the share of income from illegal sources for those at high school and aged 18 or under in 1980 on race, gender, age, family lived in an urban area, county crime rate, county unemployment rate, and county average family income.

17 We use the linear prediction from the ordered probit regression.

18 These are gender, race, attendance rate at the individual's school (this is intended to reflect both school quality and the pupil's peer group which may directly affect their preferences for truanting), parents' education, whether head of household was professional, lived with biological parents, number of siblings and whether family lived in an urban area.

level influences such as the effect on truanting behaviour of differences between states in penalties associated with truanting offences.

5. Results

We present our estimates of the structural model of truancy and then illustrate the impact of various observables on the level of truancy by means of hypothetical examples, since the observables may have both a direct and an indirect (through the rates of return) impact.

5.1 Estimates of the structural model

We first discuss the estimates of the predicted the returns to education, work and crime. These are given in the Appendix in Tables A3, A4 and A5. The regression results for the returns to education, given in Table A3, show which factors are significant predictors of later adult earnings. Most of the coefficients on variables that are included as levels are significant and of the expected sign and accord with prior work using the NLSY79. For example, log hourly real earnings are positively associated with the individual's ability (as measured by AFQT score), parents' education, the regional wage rate and the percentage of teachers in the respondent's school with a master's degree or above.

What is of importance for the purposes of our modelling are the coefficients on the interaction terms with post high school education, since these are used to predict the returns to education. Focusing on those interaction coefficients that are significant, we find that the returns to post high school education are higher for women than men, and larger for blacks than whites or hispanics. Being of high ability is associated with bigger gains from attending college compared with those in the second and third quartile groups of AFQT score. However, the better educated are the respondent's parents, the smaller is the earnings benefit from undertaking post high school education. Five of the interaction coefficients we use to predict the returns to education are individually significant and those that uniquely identify the estimated returns parameter are jointly significant[19].

19 The F statistic is 2.43, with a significance level of 0.018.

A smaller set of variables are used to estimate the returns to work and crime but again the instruments for both are significant[20]. The results presented in Tables A4 and A5 indicate that most of the coefficients are significant and of the expected sign. We find that being a female or black reduces the gains from working whilst enrolled, but the youth wage rate and living in an urban area are positively associated with hourly earnings. These positive coefficients point to the significant influence of local labour market conditions.

As for work, women also experience lower returns to crime than men, but blacks are found to earn more from crime than either whites or hispanics. Several characteristics of the local area prove to be important for the returns to illegal activity. The richer the residents in an area, as measured by average family income, the greater are the returns to crime. One reason for this may be that theft of personal property will be more lucrative when the victims are well off. We find that the gains from illegal activity are negatively associated with the county crime rate. Perhaps people in high crime areas are more security conscious and therefore crime pays less well in such areas. It is also possible that high school youths experience smaller returns in high crime areas because they have to compete with hardened criminals. Finally, the county unemployment rate is also positively related to the returns to crime. It could be that adult criminals disproportionately live in more disadvantaged areas but commit crime in richer areas, whereas youths involved in crime only operate in their local environs. Hence, youths living in high unemployment areas may have a greater chance to establish contact with criminals who can inform them of opportunities for illegal gain.

Using these estimates, we estimate (1) using an ordered probit regression. The results are given in Table 5.1. We find that the returns to education, work and crime are all significant and have the predicted signs[21]. Hence, we find that the number of days spent truanting rises

---

20 The F statistic for the significance of the instrument for the returns to work is 4.51, with a significance level of 0.034; the equivalent chi-squared statistic from the ordered probit results for the returns to crime is 23.01 with a significance level of 0.000.

21 We have not adjusted the standard errors in Table 5.1 to take account of the fact that the measures of returns are estimated rather than random variables.

with the returns to work and crime, and falls as the returns to education increase. This clearly identifies a role for economic incentives in determining truanting behaviour: those who expect larger relative gains from non-school activities are less likely to attend school.

The other explanatory variables allow for further direct influences on truanting behaviour; we find a number of them to be significant. There is a significantly positive coefficient for women, although it is important to note that gender also operates in the model through its influence on the estimates of all the three returns parameters. Since women are predicted to have higher returns to education than men and lower returns to work and crime, the economic components of the model alone would point to a much smaller tendency to truant for females than males. The positive direct influence of being a woman has an offsetting effect. The differences in truancy by sex are explored further using the examples below.

Parental and family background characteristics are found to have a significant direct relationship with the probability of truanting, in addition to their indirect effects through the estimated returns to education, work and crime. Having well-educated parents and living with both biological parents at age fourteen were found to be associated with lower returns to education which would feed through to a greater tendency to truant; these same variables have a significantly negative direct effect which works in the opposite direction. Two other variables, the head of the family being in a professional occupation and *not* living with at least one biological parent in 1980, are both significant and the coefficients suggest that a stable and affluent family background tends to reduce the likelihood of truanting.

In our model we explicitly allowed for income constraints to influence truancy behaviour. However the results provide no evidence that those on low family income give greater weight to the current gains from non-school activities versus the future gains from education.

There is also a prior stage of the analysis to estimate wage rates from the CPS; again no adjustment has been made to the standard errors.

Table 5.1: Structural model of truancy

Dependent variable: Number of days truanted from school (categorical variable)

Regressor	Coefficient (standard error in brackets)
Expected returns to schooling	-2.3025** (0.5307)
Expected returns to work whilst enrolled	4.2309** (0.9291)
Expected returns to crime whilst enrolled	1.4374** (0.5126)
Woman	1.2349** (0.2631)
Black	0.2164 (0.2057)
Hispanic	-0.0244 (0.1331)
Number of siblings	0.0063 (0.0149)
Lived with both biological parents at age 14	-0.2820** (0.0923)
Parents' education grades	-0.0425** (0.0156)
Whether head of household was professional	-0.2488** (0.1121)
Family income quartile 1 (in 1980 if living with at least one biological parent)	-0.0567 (0.1165)
Family income quartile 2 (in 1980 if living with at least one biological parent)	-0.0612 (0.1127)
Family income quartile 3 (in 1980 if living with at least one biological parent)	-0.1210 (0.0934)
Not living with at least one biological parent in 1980	0.4606** (0.2190)
Low attendance rate at respondent's school	-0.0612 (0.1276)
High attendance rate at respondent's school	-0.0722 (0.1113)
Whether family lived in urban area	-0.2810** (0.1350)

**: $p<0.05$; *: $0.05<p<0.10$

Notes:

1. The sample is those with non-missing values for all observations (1,258).
2. The dependent variable is a categorical variable indicating the number of times truanted from high school in last year (available in 1980 only).
3. Controls for states of the Union were included; coefficients not presented in the table.
4. Calculation of standard errors allows for non-constant variance.
5. Omitted categories are 'male', 'white', 'family income quartile 4 (in 1980 if not living with at least one biological parent)' and 'middle attendance rate at respondent's school'.

Table 5.2: Hypothetical examples based on results from structural model

Person type	Predicted returns to education (1)	Predicted returns to work (1)	Predicted returns to crime (1)	Predicted prob of truanting (2 days + per year)
Mean values	0.164	1.492	0.659	0.194
Woman + other mean values	0.215	1.444	0.421	0.189
Man + other mean values	0.118	1.536	0.876	0.198
Black + other mean values	0.258	1.390	0.840	0.143
White + other mean values	0.140	1.517	0.611	0.210
Living with bio parents at age 14 + other mean values	0.158	1.492	0.659	0.182
Youth wage rate at 90th percentile + other mean values	0.164	1.528	0.659	0.239
White male in bottom quartile of AFQT scores, with parental education at 10th percentile, not living with both bio parents at age 14 and ill during school years + other mean values	0.085	1.561	0.828	0.351
Black female with AFQT score in top quartile, did not take remedial maths, parental education at 90th percentile, both parents working, professionally employed head of household, living with at least one biological parent in 1980, high school and unemployment at 10th percentile + other mean values	0.335	1.349	0.511	0.044

Notes:

1. Estimated returns to education, work and crime are all linear predictions (even though the returns to crime are generated by an ordered probit). The units are as follows: the return to education is a rate of return, so a value of 0.164 means a 16.4% rate of return to going to college. The returns to work is a log hourly wage, so a value of 1.492 means $4.45 per hour. The return to crime has no interpretable units in this form – it is the underlying 'intensity' to move to a higher share category.
2. The predicted probability of truanting is the probability of truanting given mean values (rather than the mean value of the probabilities). Using mean values, unless otherwise stated in examples, we predict the returns to education, work and crime and the resulting probability of truanting using the estimated parameters of the model. Since truancy is a categorical variable, the prediction generates the probability of being in each category; these

probabilities are then summed to determine the predicted probability of truanting more than two days a year.

3. Mean values are those for the final sample of the truancy model, 1,258 individuals.

Beyond the influences of parents and family, we considered the direct impact of living in an urban area. This increases the returns to work and crime and therefore indirectly has an upward influence on truanting too; the direct effect is an offsetting negative association with truanting behaviour. The other factor outside the realms of the family that is included in the model is the attendance rate at the respondent's school. Our empirical results do not indicate it having any significant explanatory power.

5.2 Illustrative examples

We examine the results from the structural model in more detail by calculating the probability of truanting as well as the returns to education, work and crime for a range of hypothetical examples. The starting point is to take an individual with mean values for all the variables in the model, and use the estimated parameters of the model to predict their returns to education, work and crime and the predicted probability of truanting more than two days a year. We then go on to select specific values for some variables to determine how this affects the predicted returns and probability of truanting.

In the case of a hypothetical person with mean characteristics their predicted probability of truanting more than two days a year is 19%. Women are estimated to have a slightly lower probability of truanting than men (19% compared to 20%); this confirms the findings of our descriptive analysis of truancy by gender. The difference in the returns to education, work and crime by gender all work in the direction of reducing the likelihood of truanting for women relative to men, this is only partly offset by the large positive direct effect of being a women on the probability of truanting.

The impact of race also reflects the results from the raw data. The estimated probabilities for blacks of truanting more than two days a year are only 14% compared with 21% for whites. This result implies that the influence of higher returns from education and the lower returns from working whilst at school for blacks strongly dominate the effects of race captured in the returns to crime and direct preferences for truancy.

Living with both biological parents when aged fourteen reduces the returns the education slightly below the mean but the overall effect on truancy is small but negative when we take into account the direct effect of family background.

The youth wage rate has an influence on predicted truancy levels through the returns to working whilst at school. When the youth wage rate is set at the 90^{th} percentile for the sample instead of the mean, the predicted probability of truanting rises from 18% to 23%. Interestingly from a policy perspective, this suggests that students facing more buoyant youth labour markets will tend to truant more.

Finally, we illustrate the predictions of the model for two composite hypothetical examples that can portray the extremes. For a white male with low ability (bottom quartile) and parental education (10^{th} percentile), who did not live with both biological parents at age 14 and who was ill during school years, his predicted probability of truanting would be 0.35. As well as the direct influence on truanting behaviour of these characteristics, this is driven by lower returns to education and higher returns to work and crime than for someone with mean characteristics.

At the other extreme if we take the case of a black female with high AFQT score (top quartile), who did not take remedial maths at school, has highly educated parents (90^{th} percentile) who both worked, a professionally employed head of household, living with at least one biological parent in 1980, at a school with high attendance, and living in a low unemployment county (10^{th} percentile), her predicted probability of truanting more than two days a year is 0.04. Compared to the mean example, she has substantially higher predicted returns to education and lower returns to work and crime.

6. Conclusions

The purpose of the analysis in this paper was to examine whether truanting behaviour is a response to economic incentives. We have taken a structural approach using a model of optimal time allocation and predicted the returns to education, work and crime. We then estimate a model of truancy with the predicted returns as explanatory variables. Our main finding is that economic incentives do matter in determining truanting behaviour. When we analyse a sample from the NLSY79, we find that truancy is negatively related to the returns to education and positively linked to the returns to work and crime. In addition to a significant role for economic factors, we also identify the importance of family background. There is an influence through the predicted returns to schooling, one of the main economic factors, but we also find a strong impact of family background on direct preferences and attitudes towards truancy.

Our analysis is innovative in that it takes a structural approach to analyse the economic determinants of truancy. It also allows for the influence of a broad range of factors to determine truanting behaviour, beyond that of just the individual and their family. We take in to account the characteristics of the school environment, the local area where the respondent resides and labour market indicators at the state level. These environmental factors are found to be important in estimating our model of truancy.

Finally, the main conclusion from this analysis for policy seems to be that economic incentives matter in determining the behaviour of potential truants. Whilst it is a difficult challenge to identify how policy might significantly affect the returns individuals can expect to receive from education versus alternative uses of their time, our results suggest that direct financial incentives to attend school may have some efficacy in reducing levels of truancy in that these may reduce the relative returns from other competing uses of time. The government's Educational Maintenance Allowance, which 'pays' young people to stay in school, does just that.

References

Bureau of Labor Statistics (2000) *The National Longitudinal Surveys Handbook 2000*, US Department of Labor.

Cameron, S V and Heckman, J J (2001) "The Dynamics of Educational Attainment for Black, Hispanic and White Males" *Journal of Political Economy*, vol 109, no 3, 455-498.

Downing Street (2002) "Truancy Sweeps a Success" Press Release, 18th June 2002, http://www.number-10.gov.uk/output/page5357.asp

Dustmann, C., Rajah, N. and Smith, S. (1997), "Teenage Truancy, Part-time Working and Wages", *Journal of Population Economics*, vol. 10, pp 425-442.

Eckstein, Z. and Wolpin, K. I. (1999) "Why Youths Drop Out of High School: the Impact of Preferences, Opportunities and Abilities", *Econometrica*, vol. 67, no. 6, pp. 1295-1339.

Hotz, V. J., Xu, L. C., Tienda, M. and Ahituv, A. (2002) "Are there Returns to the Wages of Young Men from Working while in School?", *The Review of Economics and Statistics*, May, vol. 84, no. 2, pp 221-236

Light, A. (1995) "High School Employment", Bureau of Labor Statistics Discussion Paper NLS 95-27, US Department of Labor

Ruhm, C. J. (1997) "Is High School Employment Consumption or Investment?", *Journal of Labor Economics*, vol. 15, no.4, pp 735-776

Appendix

A1. The data sample

For the purposes of estimating the structural model we use a sample of 1,258 observations which represents 30.5% of those in the NLSY79 aged under eighteen in 1980 (for whom truancy information was collected). When we compared the characteristics of our subsample with all those individuals aged under eighteen we found that the gender breakdown was very similar but that our estimation sample has more white individuals with higher family income than the fuller sample.

A2. The dependent variable

Table A1 presents some basic information about the truancy and days absent variables in the NLSY79. There are two notable differences between the truancy and days absent variables. Firstly, the truancy data is only available for younger individuals. Secondly, given that we want to be able to match the information on truancy / absence with other data, there are a larger number of usable observations for the truancy variable.

Table A1: Variables on truancy and school absence in the NLSY79

	Number of respondents with non-missing values	Age of respondents in 1979	Highest grade completed in 1979	Non-missing with current information*
Truanting	3896	14-17	1-12	3896 (all)
Days absent in grade 9	4676	14-22	5-16	1511
Days absent in grade 10	4629	14-22	5-16	2153
Days absent in grade 11	4310	14-22	4-16	2550
Days absent in grade 12	3870	14-22	4-16	3273

*This indicates the number of respondents where the truancy / absence refers to 1979 or later, so other current information such as health status etc. that refers to the same period would be available in the dataset.

Table A2 suggests that there is not a very high correlation between the truant and days absent variables, and perhaps more surprisingly, between the days absent variables for different grades, although it is somewhat higher.

Table A2: Correlation between truancy and number of days absent in various grades

	Truancy	Days absent in grade 9	Days absent in grade 10	Days absent in grade 11	Days absent in grade 12
Truancy	1.000				
Days absent in grade 9	0.196	1.000			
Days absent in grade 10	0.236	0.536	1.000		
Days absent in grade 11	0.274	0.461	0.553	1.000	
Days absent in grade 12	0.242	0.347	0.396	0.464	1.000

Notes:

1. Based on 1,117 respondents with non-missing values for all variables.
2. Note that the truant variable is categorical (7 bands), while the days absent variables are continuous.

For the purposes of this analysis of truancy we use the variable on truancy, since the model we propose is intended to explain the determinants of unauthorised absence. This choice of dependent variable is supported by our preliminary description of the raw data in section 2, since we find stronger associations with paid work and illegal activity for truancy than for the number of days absent. Further practical considerations are that the data on truancy was collected during one particular year of the survey (unlike the information on days absent) and hence is more straightforward to match to other information from the survey. As noted above, the available sample with truanting information is larger once we take account of the need to match with other data.

A3. Estimates of the returns to education, paid work and crime

Table A3: Future returns to education

Independent variable	Coefficient (standard error in brackets)
Education (any post high school)	0.3169** (0.0949)
Woman	-0.2140** (0.0204)
Woman*education	0.0972** (0.0264)
Black	0.0230 (0.0294)
Black*education	0.118** (0.0395)
Hispanic	0.1779** (0.0317)
Hispanic*education	0.0207 (0.0459)
AFQT score quartile 1	-0.2746** (0.0345)
AFQT score quartile 1*education	-0.0486 (0.0526)
AFQT score quartile 2	-0.1273** (0.0339)
AFQT score quartile 2*education	-0.1186** (0.0462)
AFQT score quartile 3	0.0022 (0.0328)
AFQT score quartile 3*education	-0.1300** (0.0399)
Ill in school	-0.0497 (0.0469)
Ill in school*education	-0.0968 (0.0693)
Took remedial maths in school	-0.0507* (0.0269)
Remedial maths*education	0.0144 (0.0502)
Parents' education grades	0.0135** (0.0037)
Parents' grades*education	-0.0089* (0.0052)
Parents not working	-0.0358** (0.0181)
Parents not working*education	0.0275 (0.0267)
Lived with both biological parents	0.0821** (0.0226)
Lived with bio parents*education	-0.0276 (0.0364)
Percentage of teachers in respondent's school with master's degree or above	0.0015** (0.0004)
Percentage of teachers in respondent's school with master's degree or above*education	-0.0006 (0.0006)
County unemployment rate in 1980	-0.0142** (0.0056)
County unemployment rate*education	0.0005 (0.0083)
Whether lived in urban area	0.0647** (0.0174)
Wage rate for those aged 21+	0.5730** (0.0613)
Constant	0.9747** (0.1849)

**: p<0.05; *: 0.05<p<0.10

Notes:

1. This regression is based on repeated observations for the sample aged 21 or above and not in education (33,932 observations).
2. The dependent variable is the log of real hourly earnings.
3. Controls for year and age were included; coefficients not presented in the table.
4. Calculation of standard errors allows for non-constant variance and correlation across repeated observations for each individual.
5. Omitted categories are ‘male’, ‘white’ and ‘AFQT score quartile 4’.
6. The predicted returns to education (where this is attending college post high school) are estimated for all observations in the sample to be the sum of each coefficient on the interaction terms multiplied by the relevant variable value (i.e. coeff[woman*education] * woman + coeff[black*education] * black + …) plus the coefficient on attending post high school education (ie. …+ coeff[education]).

Table A4: Current returns to working whilst enrolled in school

Independent variable	Coefficient (standard error in brackets)
Woman	-0.0925** (0.0235)
Black	-0.1268** (0.0274)
Hispanic	-0.0150 (0.0298)
Whether family lived in urban area	0.0449* (0.0259)
County unemployment rate in 1980	-0.0037 (0.0062)
Wage rate for those aged <=18	0.3062** (0.1442)
Constant	0.9314** (0.2604)

**: p<0.05; *: 0.05<p<0.10

Notes:

1. This is based on repeated observations for the sample aged 18 or below and enrolled in high school (4,986 observations).
2. The dependent variable is the log of real hourly earnings.
3. Controls for year and age were included; coefficients not presented in the table.
4. Calculation of standard errors allows for non-constant variance and correlation across repeated observations for each individual.
5. Omitted categories are ‘male’ and ‘white’.
6. The estimated return to working whilst enrolled in high school education is the predicted value from this regression.

Table A5: Current returns to crime whilst enrolled in school

Independent variable	Coefficient (standard error in brackets)
Woman	-0.4556** (0.0481)
Black	0.2295** (0.0565)
Hispanic	0.0463 (0.0673)
Whether live in urban area	0.1107 (0.0788)
County unemployment rate in 1980	0.0503** (0.0141)
County crime rate in 1980	-0.0249* (0.0132)
County average family income in 1980	0.0706** (0.0148)

**: p<0.05;   *: 0.05<p<0.10

Notes:

1. This is estimated for the sample aged 18 or below and enrolled in high school in 1980 (3,702 observations).
2. The dependent variable is the share of support in the last year (a categorical variable, not a straightforward share), which comes from criminal income (available in 1980 only).
3. Controls for age were included; coefficients not presented in the table.
4. Calculation of standard errors allows for non-constant variance.
5. Omitted categories are ‘male’ and ‘white’.
6. The estimated return to crime is estimated to be the predicted linear value from this regression

<table>
<tr><th colspan="6">Table 3.1 .Variables from NLSY79 for analysis of truancy</th></tr>
<tr><th>Variable used in analysis</th><th>Definition of relevant variables in raw data</th><th>Definition of constructed variable we use</th><th colspan="2">Mean of constructed variable</th><th>Standard deviation of constructed variable</th></tr>
<tr><td colspan="6">Individual characteristics</td></tr>
<tr><td>Woman</td><td>Fixed over time.
Categorical variable for gender</td><td>Fixed over time.
Dummy variable coded ‘1’ for female</td><td colspan="2">0.477</td><td>0.500</td></tr>
<tr><td>Black</td><td>Fixed over time.
Categorical variable for race</td><td>Fixed over time.
Dummy variable coded ‘1’ for black</td><td colspan="2">0.178</td><td>0.383</td></tr>
<tr><td>Hispanic</td><td>Fixed over time.
Categorical variable for race</td><td>Fixed over time.
Dummy variable coded ‘1’ for hispanic</td><td colspan="2">0.148</td><td>0.355</td></tr>
<tr><td rowspan="4">Age</td><td rowspan="4">Time-varying.</td><td rowspan="4">Time-varying.
Dummy variables for ages 15 to 18</td><td>Age 15</td><td>0.204</td><td>0.403</td></tr>
<tr><td>Age 16</td><td>0.386</td><td>0.487</td></tr>
<tr><td>Age 17</td><td>0.397</td><td>0.489</td></tr>
<tr><td>Age 18</td><td>0.013</td><td>0.112</td></tr>
<tr><td>Enrolled in high school</td><td>Time-varying.
Current enrolment status</td><td>Time-varying.
Dummy variable coded ‘1’ if enrolled in high school</td><td colspan="2">*</td><td>*</td></tr>
<tr><td>Not enrolled in any education</td><td>Time-varying.
Current enrolment status</td><td>Time-varying.
Dummy variable coded ‘1’ if not enrolled in any education</td><td colspan="2">*</td><td>*</td></tr>
</table>

Table 3.1 .Variables from NLSY79 for analysis of truancy				
Variable used in analysis	**Definition of relevant variables in raw data**	**Definition of constructed variable we use**	**Mean of constructed variable**	**Standard deviation of constructed variable**
AFQT score normalised by age (quartile 1)	Fixed over time. AFQT score normalised by age, recorded in 1980 only	Fixed over time. Dummy variable coded '1' if in bottom quartile of distribution of AFQT for final sample	0.250	0.433
AFQT score normalised by age (quartile 2)	Fixed over time. AFQT score normalised by age, recorded in 1980 only	Fixed over time. Dummy variable coded '1' if in second quartile of distribution of AFQT for final sample	0.249	0.430
AFQT score normalised by age (quartile 3)	Fixed over time. AFQT score normalised by age, recorded in 1980 only	Fixed over time. Dummy variable coded '1' if in third quartile of distribution of AFQT for final sample	0.248	0.432
Whether took remedial maths during high school	Fixed over time. Recorded in 1979 only	As in raw data	0.149	0.356
Ever had health limitations during high school years (grades 1 to 12)	Time-varying. Had health limitations in previous year	Fixed over time. Dummy variable coded '1' if had health limitations in any year when enrolled at high school	0.079	0.269
Whether currently/previously in education post high school (grade 13 or above)	Time-varying. Current highest grade completed	Time-varying. Dummy variable coded '1' if highest grade completed 13 or above	*	*

Table 3.1 .Variables from NLSY79 for analysis of truancy				
Variable used in analysis	**Definition of relevant variables in raw data**	**Definition of constructed variable we use**	**Mean of constructed variable**	**Standard deviation of constructed variable**
Number of physical fights at school or work in past year	Fixed over time. Recorded in 1980 only	As in raw data	0.652	1.125
Share of total income in past year which came from illegal activities	Fixed over time. Categorical variable ranging from 1 to 6, recorded in 1980 only	Fixed over time. Recoded so categories range from 0 to 5; '0' = no income from illegal sources; '5' = almost all income from illegal sources	0.280	0.725
Log of real hourly earnings	Time-varying. Variables on annual hours and annual real earnings recorded separately in all years where relevant	Time-varying. Constructed from raw data	1.557	0.647
Family characteristics				
Parents' education	Fixed over time. Recorded in 1979 only – highest grade completed separately for mother and father	Fixed over time. Average of parents' grades	11.369	3.116
Parents not working	Fixed over time. Recorded in 1979 only – work status in 1978, separately for mother and father	Fixed over time. Dummy variable coded '1' if either parent was not working	0.413	0.493

Table 3.1 .Variables from NLSY79 for analysis of truancy				
Variable used in analysis	**Definition of relevant variables in raw data**	**Definition of constructed variable we use**	**Mean of constructed variable**	**Standard deviation of constructed variable**
Whether head of household was professional	Fixed over time. Recorded when respondent was age 14 only – whether head was professional, separately for male and female heads	Fixed over time. Dummy variable coded '1' if head (whether male or female) was professional	0.165	0.371
Whether respondent lived with both biological parents	Fixed over time. Recorded when respondent was aged 14 only – categorical variable for family structure	Fixed over time. Dummy variable coded '1' if lived with both biological parents	0.796	0.403
Number of siblings	Fixed over time. Recorded in 1979 only	As in raw data	3.467	2.355
Whether family lived in urban area	Time-varying	As in raw data	0.729	0.445
Family income in 1980 if lived with at least one biological parent (quartile 1)	Time-varying. Total family income	Fixed over time. Dummy variable coded '1' if in bottom quartile of the distribution of family income in 1980 for those living with at least one biological parent in 1980; based on distribution for final sample	0.246	0.431

Table 3.1 .Variables from NLSY79 for analysis of truancy				
Variable used in analysis	**Definition of relevant variables in raw data**	**Definition of constructed variable we use**	**Mean of constructed variable**	**Standard deviation of constructed variable**
Family income in 1980 if lived with at least one biological parent (quartile 2)	Time-varying. Total family income	Fixed over time. Dummy variable coded '1' if in second quartile of the distribution of family income in 1980 for those living with at least one biological parent in 1980; based on distribution for final sample	0.250	0.415
Family income in 1980 if lived with at least one biological parent (quartile 3)	Time-varying. Total family income	Fixed over time. Dummy variable coded '1' if in third quartile of the distribution of family income in 1980 for those living with at least one biological parent in 1980; based on distribution for final sample	0.253	0.451
Not living with at least one biological parent in 1980	Time-varying. With whom respondent currently lives	Fixed over time. Dummy variable coded '1' if not living with at least one biological parent in 1980	0.025	0.158
School characteristics				
Percentage of teachers in respondent's school with master's degrees or above	Fixed over time. Recorded in 1979 only	As in raw data	47.298	21.882

Table 3.1 .Variables from NLSY79 for analysis of truancy				
Variable used in analysis	**Definition of relevant variables in raw data**	**Definition of constructed variable we use**	**Mean of constructed variable**	**Standard deviation of constructed variable**
Low attendance rate at respondent's school	Fixed over time. Attendance rate at respondent's school, recorded in 1979 only	Fixed over time. Dummy variable coded '1' if in bottom decile of the distribution of attendance rate for final sample	0.099	0.298
High attendance rate at respondent's school	Fixed over time. Recorded in 1979 only	Fixed over time. Dummy variable coded '1' if in top decile of the distribution of attendance rate for final sample	0.106	0.308
Area characteristics				
County unemployment rate in 1980	Time-varying. Not updated annually and not recorded in all years	Fixed over time. Value in 1980	4.512	1.704
County crime rate in 1980	Time-varying. Not updated annually and not recorded in all years	Fixed over time. Value in 1980, divided by 1000 for computational convenience	4.658	2.696
County average family income in 1980	Time-varying. Not updated annually and not recorded in all years	Fixed over time. Value in 1980, divided by 1000 for computational convenience	9.261	2.101

Table 3.1 .Variables from NLSY79 for analysis of truancy				
Variable used in analysis	**Definition of relevant variables in raw data**	**Definition of constructed variable we use**	**Mean of constructed variable**	**Standard deviation of constructed variable**
State of the Union where resident in 1980	Time-varying. Categorical variable, not updated annually and not recorded in all years	Fixed over time. Dummy variables for each state coded '1' if state where resident in 1980	-	-

Notes:

1. Variables described in the table have been used at various stages of the analysis and therefore different samples may be relevant. For illustrative purposes the means and standard deviations presented here are values for the sample of 1,258 individuals in 1980 for whom the truancy structural model has been estimated.

* For this sample descriptive statistics for certain variables are not presented, since there is no variation within this particular sample.