# The influence of neighbourhood effects on education decisions in a nationally funded education system : the case of Australia


Henry G. Overman and Alex Heath

# The Influence of Neighbourhood Effects on Education Decisions in a Nationally Funded Education System

Henry G. Overman*‡
*London School of Economics*

Alex Heath*§
*London School of Economics*

July 2000

Abstract: Empirical papers studying the effects of neighbourhood characteristics on socio-economic variables have predominantly used US data. We argue that the local nature of the US schooling system means that neighbourhood effects on education decisions may act through fiscal or social channels. We use data for a nationally funded public schooling system to identify neighbourhood effects in an environment where the level of school funding is independent of neighbourhood composition. We identify two different types of neighbourhood effects on school dropout. First, teenagers are more likely to dropout if the average dropout rate in the neighbourhood is high. Second, teenagers are more likely to dropout if they live in neighbourhoods with a high percentage of adults with vocational qualifications.

Key words: neighbourhood effects, education decisions, externalities.
JEL classification: I20, J24, R12.

*Common thanks.

‡Also affiliated with the Centre for Economic Performance at the London School of Economics. Funding from the United Kingdom Economic and Social Research Council is gratefully acknowledged.

§Also affiliated with STICERD at the London School of Economics.

Correspondence addresses:

Alex Heath
STICERD
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom
heatha@rba.gov.au

Henry G. Overman
Department of Geography and Environment
London School of Economics
Houghton St
LondonWC2A 2AE
United Kingdom
h.g.overman@lse.ac.uk
http://cep.lse.ac.uk/~overman

## 1. Introduction

This chapter considers empirical evidence on the importance of geographical neighbourhoods for social and economic outcomes. Casual observation suggests that children who grow up in 'bad neighbourhoods' tend to have worse outcomes on a range of social indicators. They accumulate less human capital, drop out of school earlier and have a higher risk of involvement in criminal activity. Young women are more likely to get pregnant in their teenage years, and tend to form single parent households after the birth of their child. However, the fact that neighbourhood characteristics appear to be related to individual behaviour may result from the tendency of families with similar characteristics to live close to each other[1]. This chapter considers teenage dropout rates to examine whether concentrations of poorer families in bad neighbourhoods may exacerbate individual and family effects.

Traditionally, an individual's education decisions have been treated as a function of personal characteristics, the family environment and macroeconomic conditions. Recently, however, there has been a rapid increase in the number of empirical studies analysing how the immediate geographical environment affects behaviour, above and beyond the effects of family background and macroeconomic conditions. This recent expansion in the empirical literature investigating the existence of neighbourhood effects has, in part, been driven by the availability of data allowing individuals to be located in relatively small geographic areas which can be thought of as neighbourhoods. Most of these data sources are concerned with the experience of those living in inner city areas and the suburbs of major US cities. Comparable evidence from outside the US is not readily available[2]. This chapter provides such evidence, for a nationally funded school system using data from the early 1990's.

We use Australian data to examine the earliest education decision available to Australian teenagers – whether or not to complete high school. We combine the Australian Youth Survey with neighbourhood data derived from the 1991 Australian census to create a data set of individuals, with information on their personal characteristics, family background and their immediate environment. Although the results are interesting in their own right, they also have implications for our understanding of neighbourhood effects in a much wider context. One of the main problems facing studies of neighbourhood effects using US data is that the locally funded nature of the US education system makes it difficult to distinguish between true neighbourhood effects and differential tangible inputs into the schooling system. In contrast, the distribution of funding across Australian secondary schools is relatively equitable.

Why should policy makers be interested in socio–economic neighbourhood effects? If neighbourhood effects exist, the ability of families to sort across neighbourhoods may lead to costs for families in low income neighbourhoods that outweigh the benefit to families in high income neighbourhoods. Policy makers may want to intervene to ensure that externalities arising from the presence of neighbourhood effects are internalised. The subsequent increase in efficiency could

[1] That is, to sort across neighbourhoods according to socio–economic criterion.

[2] Notable exceptions include Robertson and Symons (1996) and Meghir (1997) who use data for the UK, although Meghir (1997) does not directly consider neighbourhood effects, and Robertson and Symons (1996) concentrate on peer group effects.

increase welfare for all. In addition, equity considerations suggest that, 'A system that allows the accidents of geography and birth to determine the quality of education received by an individual is inimical to the idea of equal opportunity in the market place'[3]. The policy response will depend on the mechanisms through which neighbourhoods influence education outcomes, as well as the strength of these neighbourhood effects.

The structure of this chapter is as follows. The next section discusses the existing theoretical and empirical literature regarding the existence of neighbourhood effects, and the mechanisms through which they may operate. Section 3 provides information about the data used in the subsequent analysis. In particular, we detail the Australian Youth Survey and how it links to the 1991 Australian Census data. Other practical considerations such as neighbourhood definitions and variable selection are also highlighted. Section 4 explores the importance of neighbourhood effects for the school leaving (dropout) decisions of Australian teenagers. The results suggest that there are significant exogenous and endogenous social effects on Australian drop out rates. Policy conclusions are drawn in Section 5.

## 2. The existing literature

A number of papers suggest mechanisms through which neighbourhood effects might arise. Namely, in situations where some aspects of the individuals information set may depend on location, or where an individual's payoff or optimal strategy may be influenced by the action of others in their neighbourhood. These ideas provide theoretical support for the existence of neighbourhood effects. Establishing empirical support for neighbourhood effects has proved difficult. Data availability, measurement and identification problems have all dogged attempts to test for neighbourhood effects. We return to these issues below.

Struefert (1991) presents a theory of role models to explain why teenage education decisions may be affected by neighbourhood composition. He assumes that children infer the returns to effort at school by examining the outcomes of adults in their neighbourhood, and base their education decisions on this information. Thus, the distribution of education across neighbourhoods can influence the education decisions of future generations.

Montgomery (1991) provides an alternative 'social networks' explanation. His model assumes that the unemployed have different productivity levels, but that, without further information, they are observationally equivalent to potential employers. By introducing a social structure in which workers with similar productivity levels are more likely to associate with each other, it becomes possible for employers to increase the probability of hiring a high productivity worker by employing people recommended by current high productivity workers. By increasing information flows, social networks relieve adverse selection problems and increase efficiency. To the extent that social networks are localised it is possible that some neighbourhoods will provide their job seeking residents with better job information networks than others. For example, high unemployment

[3](Fernandez and Rogerson, 1988, p136).

areas are likely to have less active job information networks, which will decrease the probability of receiving job offers and may decrease the incentives to leave school early[4].

These two models help explain how the composition of the neighbourhood may affect an individual's decisions. We label these spillovers exogenous neighbourhood effects. We are also interested in considering endogenous neighbourhood effects, where the propensity of staying on at school is an increasing function of other teenagers' propensities to stay on at school. In the terminology of Cooper and John (1988), we are interested in the existence of strategic complementarities. The presence of strategic complementarities also raises the possibility of multiple equilibria across otherwise identical neighbourhoods. Banerjee and Besley (1990) use this idea to model the importance of peer effects on education achievement. These endogenous effects are also implicit in the ethnographic evidence described in Akerlof (1997), which suggests that an individual's payoffs to completing school can be severely diminished if peer group members do not complete school.

A rapidly growing literature has also found empirical support for the existence of neighbourhood effects. Jencks and Mayer (1990) provide a detailed survey of the early literature. One of the best sources of data for looking at the influence of the neighbourhoods on education outcomes is the 1968 sample of the University of Michigan Panel Study of Income Dynamics (PSID) combined with the 1970 Census Fifth Count for Zip Codes. This provides a sample of young male heads of household who were 23–32 years old in 1978 and who were living with at least one of their parents in one of 188 Standard Metropolitan Statistical Areas in 1968. The neighbourhood data consist of a number of socio–economic indicators recorded by five digit zip code.

Two representative papers are Datcher (1982) and Corcoran, Gordan, Laren, and Solon (1992). Both find strong intergenerational links between father's 1968 income and son's subsequent economic status. However, neither report a strong impact of neighbourhood variables on son's income over and above family background effects. Corcoran *et al.* (1992) conclude that a likely reason for these problems is the presence of measurement error and omitted variable bias.

Crane (1997) finds evidence of neighbourhood effects which are especially important in low income neighbourhoods. He suggests that the extremely bad outcomes observed in inner city areas of major US cities, can be explained by epidemic or contagion effects, triggered after some critical level of social problems is reached. After this point, outcomes in these neighbourhoods deteriorate rapidly as susceptibility to these problems increases. He tests this hypothesis by estimating a piecewise linear logit model using the Census Bureau's 1970 Public Use Microdata Samples, and finds that the probability of dropping out of school is much higher than background characteristics suggest for teenagers in the lowest 5% of the neighbourhood distribution.

Case and Katz (1991) explicitly allow for the possibility of strategic interaction between agents in their analysis of the influence of neighbourhoods on the outcome of youths in low income neighbourhoods in inner city Boston. They look at the influence of peer behaviour and the characteristics of older members of the neighbourhood on several outcome variables including teen pregnancy, drug abuse, church attendance, involvement in crime and drop out rates. They find that there are

[4] This may also help explain the increasing concentration in Australia of unemployment in low status neighbourhoods between the 1976 and 1991 Censuses (Gregory and Hunter (1995)). For further evidence on this see Heath (1998).

significant neighbourhood effects, even after a large array of family background characteristics are taken into account. Interestingly, they find that child behaviour is strongly influenced by similar behaviour of the neighbourhood adult population. High rates of neighbourhood crime bias children towards criminality, high neighbourhood church attendance biases children in other, more saintly, directions.

## 3. The AYS and Australian neighbourhoods.

The Australian Youth Survey is compiled by the Australian Department of Employment, Education and Training[5]. The data covers the period from 1989 to 1994. The first wave, sampled in 1989, consists of 5350 sixteen to nineteen year olds. In each subsequent year roughly 1500 sixteen year olds are interviewed for the first time, and all other panel members are re–interviewed where possible. Our sample includes teenagers who were in the final year of high school, or were in the same cohort but left school at an earlier stage. In this sample, the probability of leaving school early is 30 percent, which is consistent with aggregate retention rates over this period.

Extensive individual and family background information is collected, including details of educational outcomes and labour market experience for both the respondent and the members of their household. Unfortunately, parent's income is not well measured. Child reported income figures are available, but the response rate is relatively low and the quality of the data is questionable. There is, however, detailed information about the occupational status of parents and their education levels, both of which are likely to be good proxies for income, especially permanent income. These variables are also likely to provide information about the parents' likely attitudes to education. Information on other important variables is also available, including the number of siblings and the type of school attended.

Most importantly for our purposes, the AYS provides detailed geographic information. As well as providing information about which state the respondent lives in, and the section of state[6]. the respondent lived in before they were 14 years old, the AYS allows individuals to be located by their geographic neighbourhood in most years. In 1989 and 1990 the information is recorded by 1986–defined collection districts (CD), which are small neighbourhoods containing, on average, 465 individuals. The postcode where the interview took place is available for re–interviewees in 1991 and all people interviewed from 1992 to 1994. Postcodes are larger than CDs, but there is a mapping from 1986–defined CDs to 1991 defined postcodes. The average postcode has 5558 residents over the age of 15 years. The largest postcode has a population of 62885; the smallest has less than a hundred residents. The distribution is highly skewed with 90 percent of postcodes with fewer than 15131 residents.

The following analysis is restricted to major urban areas for two reasons[7]. The first is that the ABS introduces sampling error into small postcodes to ensure confidentiality. By excluding non–major urban areas, most affected postcodes will be excluded. The second reason is that the concept of the neighbourhood underlying the economic models above is related to physical proximity.

[5]Now known as the Department of Employment, Workplace Relations and Small Business.

[6]Section of state is categorised as either capital city, other city, country town or rural area.

[7]Major urban areas are defined as cities with greater than 100 000 in population.

Consequently, low density population areas, such as rural areas, do not conform easily to the concepts underlying our analysis.

The childhood postcode is defined as the postcode where the individual was interviewed when they were 16, as this is the earliest recorded neighbourhood information. The postcode information for 16 year–olds is missing in 1991, and these individuals are allocated their 17 year old postcode from the subsequent interview. This is also done for the 17 year olds in 1989 to increase the available sample. Childhood postcodes are only defined if the children are living with one or both of their parents. This is standard practice in the literature, but may cause biases if the decision to move out from the family home is a function of the endogenous variable. For example, if children who leave school at 16 are also more likely to move out then we under sample this group of children. It should be noted here, that respondents who reported that they had spent most of their life until 14 overseas are excluded from the analysis. This does not significantly affect any of the formal analysis.

We have information on a range of neighbourhood characteristics at the postcode level from the 1991 Australian Census. This includes information about male and female education attainment, household and personal income, and labour force status. We also have a neighbourhood socio–economic status (SES) variable based on 1991 Census data which was constructed at CD level by Hunter (1996).

## 4. Empirical model and results

The purpose of our analysis is to estimate the effects of neighbourhood on education decisions. In particular, we look at the first free education decision available to Australian teenagers: whether to complete high school, or to leave at the legal minimum age of 15 years. In Section 4.1 we develop an empirical model within the framework presented by Manski (1993, 1995) to formalise the different mechanisms through which neighbourhood effects could operate. In Section 4.2 we present the results of estimation that ignores neighbourhood effects. In Sections 4.3 and 4.4 we add neighbourhood variables and discuss the importance of neighbourhood effects for our understanding of teenage education decisions. In Section 4.5 we discuss alternative interpretations of these results.

### *4.1 A common framework*

Our empirical work is based on the model originally proposed by Manski (1993, 1995), summarised as follows:

$$y_i^* = \alpha + z_i\beta + E(z|x_i)'\gamma + \delta E(y^*|x_i) + \varepsilon_i; \tag{1}$$

where $y_i^*$ is the underlying propensity to leave school before the final year of high school for individual i; $z_i$ are the personal background and family characteristics of individual i; $x_i$ is the postcode neighbourhood of individual i; $E(z|x_i)$ are the average characteristics of the individuals in that neighbourhood; $E(y^*|x_i)$ is the probability of being an early school leaver in that neighbourhood; and $\varepsilon_i$ is the error term which contains all the unobserved factors which affect individual i's propensity to leave school before the final year.

Thus $E(z|x_i)$ captures exogenous neighbourhood effects, and $E(y^*|x_i)$ captures endogenous neighbourhood effects.

### *4.2 Individual effects*

We start our analysis by estimating the model assuming that neighbourhood effects are not important (i.e. assuming that $\gamma = \delta = 0$ in equation 4.1). This specification has been considered in earlier literature and has been quite successful in explaining teenage education decisions (Miller and Volker (1987)). Because we do not observe the propensity to leave school early but the final decision, which is a binary variable, we estimate this model using probit. The results are presented in the first two columns of 1.

Because the probit model is non–linear, the estimated coefficients will provide information about the direction of the effect an independent variable has on the probability of leaving school early, but the magnitude of the effect depends on where the probability is evaluated. To facilitate comparison we present results in marginal effects form. The marginal effect can be interpreted as the impact a one unit change in the variable will have on the probability of leaving school early, given that the probability is initially evaluated at the sample mean. For dummy variables, marked with an asterisk, the reported marginal effect will be the change in the probability of being an early school leaver if the individual has that characteristic rather than the reference characteristic given by the omitted group. The results in table 1 are expressed in marginal effects form.

Note that most of the variables have the expected effect. Males are 8 percentage points more likely to leave school early than females, and older cohort members are more likely to leave school early than younger ones, perhaps reflecting the effects of repeating earlier school years. Teenagers with more brothers and sisters are more likely to leave school early possibly reflecting financial constraints.

Teenagers who do not have English as their first language are significantly more likely to complete high school than teenagers who are born overseas. These effects are independently significant, so that teenagers who were born overseas and did not learn English as their first language are 15.7 percentage points more likely to complete high school, if they judge themselves to speak English well and are 21.4 percentage points more likely to stay on if the were born overseas and have poor English skills. This may reflect different attitudes to education, but may also reflect the relatively poor prospects these teenagers may face in the high unemployment youth labour market. Teenagers who attend a government school are 17 percentage points more likely to leave school early than their counterparts attending a Catholic school and are 12.5 percentage points more likely to leave than teenagers at private schools.

Parents' characteristics are important for explaining the decision to leave school early. Teenagers with fathers who have higher status jobs are less likely to leave high school early. Teenagers from single parent families are significantly less likely to complete high school. Parents with degree qualifications are much more likely to have children who complete high school: degree qualified fathers and mothers decrease the chances of leaving by 5.5 and 9.1 percentage points respectively.

There are initial indications that location has an influence on teenage decision making, as the section of state and state variables are jointly significant. If, however, the neighbourhood has some

| | Individual Effects | | Reduced Form | | Structural Form | | Sample |
|---|---|---|---|---|---|---|---|
| | Coeff. | t–stat | Coeff. | t–stat | Coeff. | t–stat | Average |
| **Neighbourhood** | | | | | | | |
| Average personal income | – | – | 0.000 | 0.39 | 0.000 | −0.43 | 5601.48 |
| Proportion with grad qual | – | – | 0.001 | 0.31 | 0.015 | 2.28 | 13.48 |
| Proportion with trade qual | – | – | 0.014 | 3.02 | 0.016 | 3.41 | 13.64 |
| Unemployment rate | – | – | 0.005 | 1.52 | 0.001 | 0.23 | 11.35 |
| Endogenous effect | – | – | – | – | 0.021 | 2.36 | 30.07 |
| **Personal Background** | | | | | | | |
| Male* | 0.080 | 5.59 | 0.081 | 5.65 | 0.080 | 5.60 | 0.50 |
| Age | 0.093 | 6.48 | 0.092 | 6.43 | 0.092 | 6.40 | 17.11 |
| Number of Siblings | 0.009 | 1.77 | 0.009 | 1.85 | 0.008 | 1.73 | 2.02 |
| English not $1^{st}$ language | | | | | | | |
| English good* | −0.080 | −3.12 | −0.072 | −2.76 | −0.072 | −2.75 | 0.10 |
| English poor* | −0.137 | −2.09 | −0.129 | −1.94 | −0.126 | −1.89 | 0.01 |
| Born overseas* | −0.077 | −3.15 | −0.076 | −3.12 | −0.075 | −3.08 | 0.12 |
| School | | | | | | | |
| Catholic* | −0.125 | −7.32 | −0.120 | −6.96 | −0.119 | −6.94 | 0.23 |
| Other non–government* | −0.170 | −7.10 | −0.161 | −6.48 | −0.160 | −6.46 | 0.09 |
| **Parent's Characteristics** | | | | | | | |
| Fathers occ. status @14 | −0.002 | −3.51 | −0.002 | −3.48 | −0.002 | −3.47 | 29.72 |
| Mother's occ. status @14 | 0.000 | −0.26 | 0.000 | −0.29 | 0.000 | −0.27 | 19.56 |
| Father not emp @14* | −0.017 | −0.50 | −0.016 | −0.45 | −0.012 | −0.36 | 0.05 |
| Mother not emp @14* | 0.004 | 0.20 | 0.004 | 0.18 | 0.003 | 0.14 | 0.36 |
| Father not present @14* | 0.059 | 2.26 | 0.060 | 2.30 | 0.061 | 2.32 | 0.16 |
| Mother not present @14* | 0.443 | 10.84 | 0.445 | 10.82 | 0.444 | 10.79 | 0.05 |
| Father has | | | | | | | |
| degree* | −0.055 | −2.17 | −0.052 | −2.04 | −0.052 | 2.01 | 0.17 |
| trade qualifications* | 0.020 | 0.95 | 0.016 | 0.79 | 0.017 | 0.84 | 0.17 |
| other post-secondary* | −0.039 | −1.55 | 0.045 | −1.78 | −0.046 | −1.83 | 0.10 |
| Mother has | | | | | | | |
| degree* | −0.091 | −3.56 | −0.083 | −3.22 | −0.087 | −3.39 | 0.13 |
| trade qualifications* | 0.019 | 0.52 | 0.017 | 0.49 | 0.018 | 0.49 | 0.04 |
| other post-secondary* | −0.025 | −1.19 | −0.025 | −1.17 | −0.026 | −1.22 | 0.15 |
| **Section of state** | | | | | | | |
| Other capital city* | 0.014 | 0.70 | −0.005 | −0.22 | −0.001 | −0.07 | 0.17 |
| Rural area* | 0.045 | 1.03 | 0.025 | 0.57 | 0.028 | 0.63 | 0.03 |
| Country town* | 0.026 | 0.99 | 0.008 | 0.29 | 0.010 | 0.39 | 0.08 |
| **State** | | | | | | | |
| Victoria* | −0.025 | −1.33 | 0.001 | 0.05 | −0.004 | −0.17 | 0.24 |
| South Australia* | −0.048 | −1.90 | −0.037 | −1.32 | −0.053 | −1.86 | 0.10 |
| Western Australia* | −0.008 | −0.32 | −0.008 | −0.29 | −0.010 | −0.33 | 0.11 |
| Queensland* | −0.072 | −3.35 | −0.045 | −1.79 | −0.048 | −1.89 | 0.17 |
| Tasmania* | −0.005 | 0.09 | 0.043 | 0.79 | 0.015 | 0.27 | 0.02 |
| ACT* | −0.119 | −3.43 | −0.087 | −2.19 | −0.095 | −2.40 | 0.04 |
| Log Likelihood | -2256.80 | | -2249.20 | | -2246.41 | | |
| Pseudo $R^2$ | 0.159 | | 0.162 | | 0.163 | | |
| Test overall significance | 854.28 | $\sim \chi^2(34)$ | 869.44 | $\sim \chi^2(38)$ | 874.99 | $\sim \chi^2(39)$ | |

Note that time dummies have been included in estimation but are not reported.

**Table 1**. Neighbourhood effects regression results

influence on a teenager's school leaving decision, this will not be captured fully by the variables we have included, and will instead enter the error term. In figure 1, we plot the actual and predicted probabilities of leaving school early against the proportion of the neighbourhood with graduate qualifications. We are interested in whether the difference between these two probabilities varies systematically across neighbourhoods.

To construct this heuristic measure of spatial correlation, we assign individuals to neighbourhoods and then smooth across neighbourhoods ranked according to proportion of neighbourhood with graduate qualifications. The non–parametric smoothing procedure we use is loess (Cleveland (1993. Visualising data)). Loess applies a weighting scheme to individuals on either side of the target individual so that more similar individuals receive higher weights[8]. The final step in the procedure is to use these weights to estimate a weighted least squares regression, centred on each individual in turn, of the outcome variable, eg the observed decision of whether or not to leave school early, on the variable defining the ranking of the individuals. The smoothed outcome for the centre individual is then calculated as the predicted value of the outcome from the regression.

The underlying rationale, is that neighbourhoods with similar education compositions should show similar dropout behaviour. This will be true, whether or not exogenous neighbourhood effects matter, providing that neighbourhoods with similar proportions of degree holders have similar compositions with respect to other characteristics. This procedure allows us to maintain the full sample as we can estimate actual neighbourhood dropout rates for large and small neighbourhoods when only the zero–one dropout decision is observed. We loose some information on endogenous social effects however, because neighbourhoods with unusually high dropout rates only form part of a weighted average when calculating the predicted and actual dropout rates. Thus, to the extent that endogenous neighbourhood effects are important, the figure will underestimate the systematic nature of the errors. We return to this issue later.

Figure 1 shows that a systematic difference between the smoothed actual and predicted probabilities of leaving school early does exist: the individual effects model is under–predicting the probability of being an early school leaver in less educated neighbourhoods and over–predicting this probability for high education neighbourhoods. Although family background explains a large amount of the absolute difference in the probability of being an early school leaver, it is not the whole story[9].

[8]The weights are derived from a tricube function:

$$T(u) = \begin{cases} (1-|u|^3)^3 & \text{for } |u| < 1 \\ 0 & \text{otherwise} \end{cases}$$

[9]We have tried to account for the unusual shape of the tails by allowing for a non–linear effect of parental background. Children living in the lowest (highest) ranked neighbourhoods are, presumably, more likely to have both parents with very low (high) education levels. To see whether it is this effect driving the tails, we interact the parent secondary dummies and the parent degree dummies. However, standard chi–tests show that both variables are insignificant. Further, including both additional dummies, does not change the shape of the predicted dropout probabilities (perhaps unsurprising, given that both variables are insignificant).

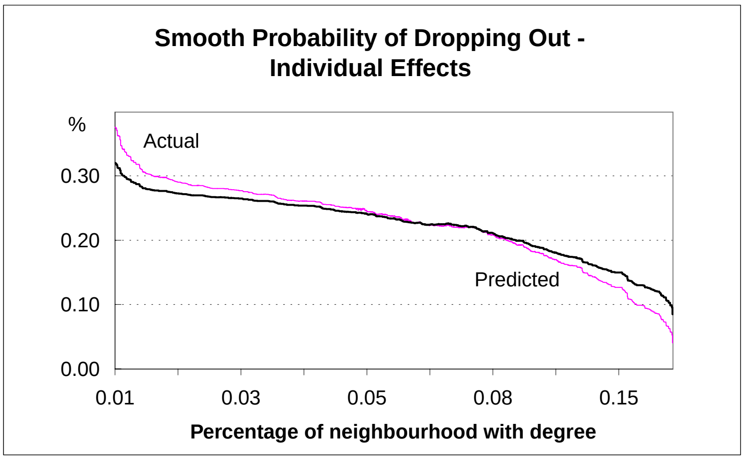


**Figure 1**. Individual effects - predicted versus actual dropout behaviour

### *4.3 Reduced form*

If we take expectations of equation 4.1, conditional on the individual's neighbourhood we obtain as a linear function of . Substituting this out we obtain equation 4.3.

$$y_i^* = \frac{\alpha}{(1-\delta)} + z_i'\beta + \frac{E(z|x_i)'(\gamma - \delta\beta)}{(1-\delta)} + \varepsilon_i. \tag{2}$$

Given personal characteristics and family background variables, we can test for the presence of neighbourhood effects by including average neighbourhood characteristics in the standard probit framework. We will not be able to separately identify the coefficients $\gamma$ or $\delta$, and therefore, we cannot distinguish between endogenous and exogenous neighbourhood effects in this reduced form specification. The results from estimating equation 4.3 are presented in columns 3 and 4 of 1.

In general, the size and significance of the marginal effects of personal characteristics and family background variables do not change noticeably. The variables which are most affected by the presence of the neighbourhood variables are, unsurprisingly, the section–of–state and state variables. The neighbourhood variables indicate that an individual is more likely to leave school early if the proportion of people in the neighbourhood with vocational qualifications is higher, and to a lesser extent, if the neighbourhood unemployment rate is higher.

Figure 2 compares the smoothed actual probability of leaving school early with the smoothed probability of leaving school early predicted by the reduced form estimation. Including neighbourhood effects has reduced the systematic error between the predicted and actual probabilities of leaving school early, although the reduced form model is still under–predicting the probability of leaving school early for the low education neighbourhoods. Thus, there appears to be some support for the Crane (1997) hypothesis of epidemic effects in low status neighbourhoods.

There are several possible explanations for the importance of the proportion of the neighbourhood with vocational qualifications. The most likely explanation, is that it captures the extent and usefulness of the job information network available to a teenager contemplating leaving school early. To the extent that vocationally trained adults are aware of jobs that offer opportunities to early

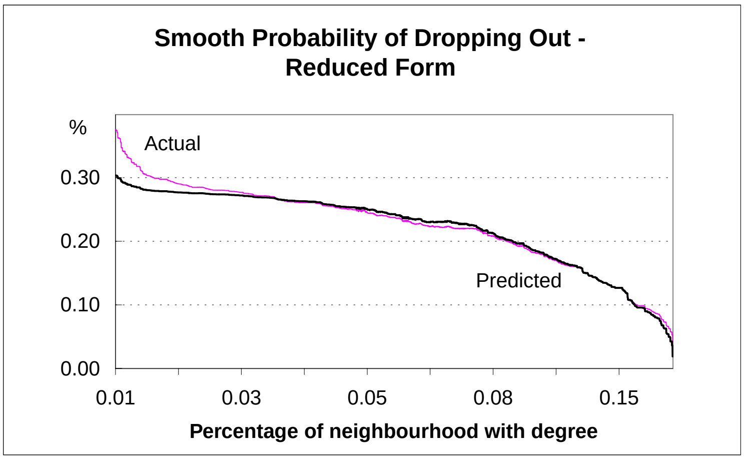


**Figure 2**. Reduced form - predicted versus actual dropout behaviour

school leavers, teenagers with access to this network will have higher expected benefits of leaving school early than teenagers in neighbourhoods without a high proportion of vocationally trained adults.

Another related possibility is that the proportion of vocationally trained adults represents the level of local labour demand, and therefore the probability of an early school leaver securing a job. For this effect to be operating, it would be necessary to argue that the proportion of vocationally trained adults is a better proxy for the local demand for unskilled labour than the local unemployment rate. This is not an unreasonable hypothesis, however, the distinction between these two channels cannot be resolved in the current context, especially given potential multicollinearity problems (see below).

The final possibility is that there is a role model effect, similar to the model presented by Struefert (1991). In this model the probability of leaving school early increases as the number of high earning, highly educated role models in the neighbourhood decreases. This model is based on the underlying assumption that the returns to completing high school and undertaking further education are higher than they are for leaving school before the completion of Year 12. This is true in Australia Gregory and Vella (1996). However, Dockery and Norris (1996) present evidence that suggests the returns to completing an apprenticeship are also high in Australia. If a teenager's information set includes a large number of adults receiving relatively high returns on their vocational training, and there are relatively few adults to demonstrate the returns to graduate education, this will naturally bias them towards leaving school early to find an apprenticeship.

### *4.4 Structural form*

Although we have established that neighbourhood effects appear to be present, we cannot determine whether these are exogenous or endogenous effects. If we had sufficient observations in each neighbourhood, we could estimate equation 4.1 using sample estimates of $E(y^*|x_i)$. Due to the small number of observations per neighbourhood, however, we must use information from individuals in 'similar' neighbourhoods to calculate a sample estimate, $\hat{E}(y^*|x_i)$. Again, we can smooth across neighbourhoods to construct this estimate. Because $\hat{E}(y^*|x_i)$ is calculated from the sample, it is potentially correlated with the error for each individual in that neighbourhood[10]. This is a side effect of the very feedback structure that we are trying to capture.

We try to solve this problem by finding suitable instruments for $\hat{E}(y^*|x_i)$. As always, a good instrument should be correlated with the average probability of dropout in the neighbourhood, but should not affect the individual's decision to dropout. We choose the average number of siblings in the neighbourhood as an instrument, because it should be correlated with the average probability of peers dropout, but should not have an effect on the individual's decision beyond this. Again, the small number of individuals in each neighbourhood suggests that we use a smoothed sample average of the number of siblings in the neighbourhood.

The results of estimating equation 4.1, instrumenting $\hat{E}(y^*|x_i)$ with the smoothed average number of siblings are presented in columns 5 and 6 of table 1. The estimated marginal effects of the variables capturing personal characteristics and family background do not change noticeably. The proportion of the neighbourhood with trade qualifications has remained positive and significant and the marginal effect is comparable to that estimated in the reduced form specification. The variable included to capture the endogenous effects, $\hat{E}(y^*|x_i)$ is positive and significant.

However, the estimated marginal effects and significance of the other neighbourhood composition variables change markedly. The local unemployment rate changes from being marginally significant to being insignificant, and the marginal effect of the proportion of the neighbourhood with graduate qualifications, which had a perverse sign in the reduced form regression, has become larger and significant. To some extent this was to be expected if the reduced form parameters were capturing both neighbourhood effects[11].

Again, we can provide a heuristic check on whether endogenous effects matter, by looking at the difference between actual smoothed probabilities of leaving school and the smoothed probabilities predicted from the structural form model. Figure 3 provides further support for the Crane (1997) hypothesis of epidemic effects in low status neighbourhoods because the inherently non–linear nature of the endogenous neighbourhood variable has improved the ability of the model to predict the probability of teenagers leaving school early in low status neighbourhoods. Although this specification has reduced the difference between actual and predicted smoothed probabilities at the low end of the education distribution, these differences have increased slightly at the upper end of the distribution relative to the reduced form specification.

In summary, it appears that significant neighbourhood effects influence a teenager's decision of whether or not to complete high school. Neighbourhood composition affects this decision through the proportion of the neighbourhood with vocational qualifications. There is also some evidence for the presence of endogenous neighbourhood effects. Although multicollinearity problems make it

[10] This is independent of the fact that we have smoothed over neighbourhoods, although smoothing lowers the correlation between the endogenous variable and the error term.

[11] When estimating the structural form excluding the proportion of the neighbourhood with graduate qualifications and the local unemployment rate, the endogenous effect variable is estimated to have a marginal effect of 0.4 of a percentage point with a t–statistic of 1.23, and the proportion of the neighbourhood with vocational qualifications has a marginal effect of 0.8 percentage points and remains significant. This confirms the intuition that the significance of the proportion of graduate qualifications in the neighbourhood is spurious, but also makes it difficult to assess the importance of endogenous neighbourhood effects.

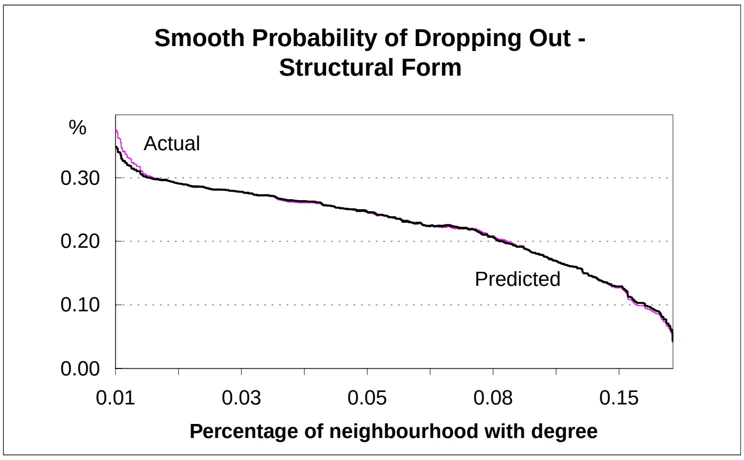


**Figure 3**. Structural form - predicted versus actual dropout behaviour

difficult to separate the two effects, the structural form model appears to be better at explaining actual school leaving behaviour over the whole distribution of neighbourhoods than the reduced form model. However, we note that these effects are dominated by personal characteristics and family structure. Before concluding, we briefly consider possible objections to our interpretation of the results as demonstrating the existence of neighbourhood effects.

### *4.5 Have we really found neighbourhood effects?*

One common objection to empirical analysis of neighbourhood effects is that the neighbourhood composition variables may just be picking up omitted individual level variables such as parents' attitudes. There are two responses to this objection. The first is that omitted background variables are more likely to be correlated with the large number of included background variables than with neighbourhood variables.

The second response is that the mechanisms by which such effects are supposed to occur is difficult to specify. Our interpretation can only be affected by an omitted variable, which is positively correlated with the proportion of the neighbourhood with vocational training, as a negatively correlated variable would induce negative bias which serves to strengthen our case. It is difficult to imagine an omitted individual level variable which increases the probability of a teenager leaving school early, and is more highly correlated with the proportion of vocationally trained adults in the neighbourhood than with any individual level variables.

The omitted variable problem is further complicated by the possibility of endogenous sorting. This will arise if there are omitted variables, such as school quality, which directly affect the probability of leaving school early, but also have an indirect effect on neighbourhood composition through the location decisions made by families on the basis of school quality. Thus, the omitted variable will be correlated with the neighbourhood composition variables, which we have treated as exogenous. Again, there are two possible responses.

The first is that we would expect an omitted variable which causes families to sort, to be more correlated with the endogenous neighbourhood effect, which we have instrumented for. Second, to the extent that the unobserved variable affecting the location decision of families is more highly correlated with the proportion of the neighbourhood with graduate qualifications than with the proportion of the neighbourhood with vocational training, we would expect the positive bias to be greater for this variable. This effect is not apparent in the results presented in table 1.

## 5. Conclusions

This chapter examines the factors that affect a teenager's decision to leave school early. In particular, we consider whether higher rates of early school leaving in some neighbourhoods is the result of 'clustering' of families with characteristics which discourage school completion, or whether the neighbourhood has an independent effect. We find that, although personal characteristics and family background variables explain much of the distribution of early school leaving behaviour across neighbourhoods, these variables are not enough.

We find evidence of significant exogenous neighbourhood effects. Specifically, we find that a larger proportion of vocationally trained adults in the neighbourhood increases the probability of a teenager leaving school early, even when the qualifications of each parent have been controlled for. We suggest that the most plausible explanation for the presence of this effect is that this variable is a proxy for the extent and usefulness of local job information networks or local labour market characteristics, which may affect the balance of costs and benefits to these teenagers of staying on at school. We also find some evidence for the presence of endogenous neighbourhood effects, which arise when the schooling decisions of other teenagers in the neighbourhood affect an individual's decision.

Distinguishing between endogenous and exogenous neighbourhood effects is important, because the policy implications of these two types of effects are quite different. Theoretical results suggest that endogenous feedback mechanisms, such as endogenous neighbourhood effects can lead to multiple equilibria even for initially identical neighbourhoods. One–off expenditures that reduce the rate of dropout may have long run benefits if the endogenous feedback mechanism pushes the neighbourhood to a new equilibrium. In contrast, policies that attempt to affect school decisions by changing neighbourhood composition may have to be ongoing if endogenous sorting in future periods pushes the neighbourhood configuration back to its old equilibrium.

It is also important to bear in mind that while the analysis in this chapter argues that neighbourhood effects influence teenage education decisions, we have only suggested possible mechanisms through which exogenous neighbourhood effects operate. Further research is necessary to identify the exact channel through which these socio–economic effects operate.

## 6. Appendix A: Variable definitions

The following variables take the value 1 when the characteristic is present and 0 otherwise;

- Personal characteristics: male, born overseas;
- Parent's characteristics: parent not employed when the respondent was 14, parent not present in the household when the respondent was 14;

- Parent's education: has a degree, has a trade qualification, has other post school qualifications (omitted category: parent has completed high school or less);
- Section of State: other city, rural area, country town (omitted category: capital city); and
- State: Victoria, South Australia, Western Australia, Queensland, Tasmania, ACT (omitted category: New South Wales; Northern Territory dropped due to too few observations).

The following variables are count variables:

- age, number of siblings.

The language proficiency variables are defined as:

- 'English good' takes the value one if the respondent does not have English as a first language, but regards their proficiency as 'very good' or 'good'; and
- 'English poor' takes the value one if the respondent does not have English as a first language and regards their proficiency as 'fair', 'poor' or 'very poor'.
- The omitted category contains respondents who speak English as their first language.

The 'School' variable is defined as:

- 'Catholic' takes the value one if the respondent is currently studying at a Catholic high school, or when last school before they left was a Catholic high school;
- 'Other non–government' takes the value one if the respondent is currently studying at a non–government, non–Catholic high school, or when last school before they left was an 'other non–government' school.
- The omitted category contains respondents whose current or last school was a government high school.

Parent's occupational status is measured as the socio–economic status of the respondent's parent when the respondent was 14. If the parent was not present in the household or was not employed the index is set to zero.

The neighbourhood variables are defined as:

- 'average personal income' is the average personal income of the respondent's postcode;
- 'proportion with grad qual' is the proportion of the respondent's postcode who recorded having a higher degree, a degree, a graduate diploma, or an undergraduate diploma.
- 'proportion with trade qual' is the proportion of the respondent's postcode who recorded having skilled vocational training or basic vocational training;
- 'unemployment rate' is the unemployment rate of the respondent's postcode.
- The omitted education category is the proportion of the respondent's postcode with high school education or less.

## References


Akerlof, George. 1997. Social distance and social decisions. *Econometrica* 65(5): 1005–1027.

Banerjee, Abhijit and Timothy Besley. 1990. Peer group externalities and learning incentives: A theory of nerd behaviour. Technical Report 68, John M.Olin Program for the Study of Economic Organisation and Public Policy.

Case, Anne and Lawrence Katz. 1991. The company you keep: The effects of family and neighbourhood on disadvantaged youth. Discussion Paper 3705, National Bureau for Economic Research.

Cleveland, William S. 1993. Visualising data. New Jersey: Hobart Press.

Cooper, Russell and Andrew John. 1988. Coordinating coordination failures in keynesian models. *Quarterly Journal of Economics* 103(3): 441–463.

Corcoran, Mary, Roger Gordan, Deborah Laren, and Gary Solon. 1992. The association between men's economic status and their family and community origins. *The Journal of Human Resources* 27(4): 575–601.

Crane, Jonathan. 1997. The epidemic theory of ghettos and neighbourhood effects on dropping out and teenage childbearing. *American Journal of Sociology* 96(5): 1226–1259.

Datcher, Linda. 1982. Effects of community and family background on achievement. *Review of Economics and Statistics* 64(1): 32–41.

Dockery, A.M. and K. Norris. 1996. The 'rewards' for apprenticeship training in australia. *Australian Bulletin of Labour* 22(2): 109–125.

Fernandez, Raquel and Richard Rogerson. 1988. Income distribution, communities and the quality of education. *Quarterly Journal of Economics* 111(1): 135–164.

Gregory, Robert and B. Hunter. 1995. The macro economy and the growth of ghettos in urban poverty in australia. Discussion Paper 325, Australian National Univesity.

Gregory, Robert and F. Vella. 1996. Selection bias and human capital investment: Estimating the rates of return to education of young males. *Labor Economics* 3(2).

Heath, Alex. 1998. Job search methods, neighbourhood effects and the youth labour market. Processed, Reserve Bank of Australia.

Hunter, B. 1996. Changes in the geographic dispersion of urban employment in australia: 1976-1991. Unpublished PhD manuscript, australian national university.

Jencks, Christopher and Susan Mayer. 1990. The social consequences of growing up in a poor neighbourhood. In L.E. Lynne Jr. and M.G.H. McGeary (eds.), *Inner City Poverty in the United States*. Washington D.C.: National Academy Press, Washington D.C., 111–186.

Manski, Charles. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60(3): 531–542.

Manski, Charles. 1995. *Identification problems in the social sciences*. Cambridge: Harvard University Press.

Meghir, Costas. 1997. The effect of school quality on educational attainment and wages. Processed, University College London.

Miller, Paul and Paul Volker. 1987. The youth labour market in australia. *Economic Record* 63(182): 203–19.

Montgomery, James. 1991. Social networks and labour market outcomes: Towards an economic analysis. *American Economic Review* 81(5): 1407–18.

Robertson, Donald and Jim Symons. 1996. Do peer groups matter? peer group versus schooling effects on academic achievement. Discussion Paper 311, Centre for Economic Performance.

Struefert, W. 1991. The effect of underclass isolation on schooling choice. Processed, University of Wisconsin.