Neighbourhood effects in small neighbourhoods

Henry G. Overman

This is an electronic version of an Article published in Urban studies, 39 (1), pp. 117-130 © 2002 Routledge, part of the Taylor and Francis Group.
http://www.tandf.co.uk/journals/titles/00420980.asp

Neighbourhood Effects in Small Neighbourhoods

Henry G. Overman*
London School of Economics

22 September 1999

ABSTRACT: This paper uses data on a sample of Australian teenagers to test for neighbourhood effects on school dropout rates. The data allows us to test for neighbourhood effects at two different spatial scales. We find that educational composition of the larger neighbourhood can influence the dropout rate. We argue that this is most likely to reflect the structure of local labour market demand. We also find that low socio-economic status of the immediate neighbourhood has an adverse impact on dropout rate. This suggests that government policy may need to consider the socio-economic composition of quite small geographical areas if it considers interfering in the market to create greater income mixing within neighbourhoods.

Key words: Neighbourhood effects; Education decisions; Housing policy.
JEL classification: I20, J24, R12.

*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.

Correspondence address:

Henry G. Overman
Department of Geography and Environment
London School of Economics
Houghton St
London WC2A 2AE

1. Introduction

This paper considers the existence and the scale of neighbourhood effects on the drop out decisions of Australian teenagers. We deal with two related questions. First, does the concentration of poorer families in poor neighbourhoods amplify individual and family effects on drop out tendencies? Second, at what spatial scale might such effects occur? That is, do neighbourhood effects depend on the socio–economic composition of the immediate or the larger locality?

Put simply, neighbourhood effects occur when geographical location matters over and above personal characteristics. That is, when children from otherwise identical families, with identical abilities etc, show different drop out propensities as a function of the type of neighbourhood that they live in. There are various theories that suggest why location may affect socio–economic outcomes. Peer group effects may mean that children in worse neighbourhoods come under greater pressure from peers to drop out of school and engage in other activities. Alternatively, information mechanisms may be important, whereby children in worse neighbourhoods may be unable to correctly assess the returns to education by observing the adults around them. We return to other possible explanations below.

At an aggregate level, it is obvious that worse neighbourhoods have higher drop out rates. However, the fact that individual behaviour appears to be related to neighbourhood characteristics may result from the tendency of families with similar characteristics to live close to each other. Thus, identifying these effects is difficult, because people sort across geographical space according to characteristics that matter for socio–economic outcomes. Sorting across neighbourhoods leads to a *correlation* between neighbourhood characteristics and drop out rates. The neighbourhood effects that we want to capture are ones where neighbourhood economic and demographic characteristics *cause* changes in drop out behaviour. The problem is exacerbated by the fact that we do not know a–priori at what scale these neighbourhood effects may occur. To separate out the effects of sorting, we need information on both individual and neighbourhood characteristics. To analyse the spatial extent of neighbourhood effects we need information on the characteristics of both large neighbourhoods and the smaller neighbourhoods that make up those large neighbourhoods.

In this paper, 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 Australian Census to create a data set of individuals with information on personal characteristics, family background and immediate geographical environment. For the entire sample, we can place teenagers in geographical neighbourhoods that roughly correspond to school catchment areas. For a smaller sub–sample, we can identify where the family live within these larger neighbourhoods. We use census data on the socio–economic conditions in both the larger and the smaller neighbourhoods to test for the presence of neighbourhood effects. The data set has a number of key advantages. First, the sample is relatively recent (1989 to 1994). Most other empirical studies use data from far earlier time periods. Second, the neighbourhood data is from the Australian Census conducted in 1991. This means that neighbourhood variables do not have to be constructed from the sample, but are actual population values that reflect the 'true' socio–economic characteristics of the neighbourhood. In addition, these characteristics are measured relatively near the start of the sample period, thus reducing potential endogeneity problems[1].

A rapidly growing empirical literature has considered the existence of neighbourhood effects. **?** provide a detailed survey of the early literature. Most of this work uses US data. There are relatively few papers which deal with the existence of neighbourhood effects in a public school system. Likewise, there is very little work that looks directly at the issue of the scale at which neighbourhood effects matter.

Many US studies on education outcomes use the 1968 sample of the University of Michigan Panel Study of Income Dynamics (PSID) combined with the 1970 Census Fifth Count for Zip Codes[2]. 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 consists of a number of socio–economic indicators recorded by three or five digit zip code. **?** uses data from the Census Bureau's 1970 Public Use Microdata Samples to test for epidemic effects triggered after some critical level of social problems is reached. For this study, the neighbourhood variables were calculated from data on around 1500 nearest neighbour families. **?** study the influence of neighbourhoods on the outcome of youths in low income neighbourhoods in inner city Boston. These four frequently cited studies all use different neighbourhood definitions and provide little information on the size characteristics of the neighbourhood that they consider. Our data allows for two neighbourhood definitions which, although based on statistical data collection areas, actually correspond to neighbourhood concepts that may be considered important in determining drop out propensities. This allows us to explicitly consider the scale at which neighbourhood effects may occur.

The paper is structured as follows. Section 2 describes the data. In particular, we discuss how we link the Australian Youth Survey to the 1991 Australian Census data. We also consider neighbourhood definitions and variable availability. Section 3 sets out our empirical model, and explores the importance of neighbourhood effects for the school leaving decisions of Australian teenagers. Section 4 concludes.

2. Data and definitions

The Australian Youth Survey is compiled by the Australian Department of Employment, Education and Training[3]. 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.

The AYS provides detailed geographic information for all respondents. As well as providing information about which state the respondent lives in, and the section of state[4] the respondent lived in before they were fourteen years old, the AYS allows individuals to be located by their geographic neighbourhood in most years. In 1989 the information is recorded by 1986–defined

[1] These can arise if families sort across neighbourhoods in response to these characteristics.

[2] See for example **?** and **?**.

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

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

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 significantly 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.[5]

Although postcode and CD areas may not correspond exactly to some consistent notion of neighbourhood we still use them to define the neighbourhood of respondents. Postcode areas are viewed as defining some larger neighbourhood, CD areas as defining some smaller neighbourhood. Postcode areas often correspond closely to school catchment areas. CD areas are somewhat more arbitrary, but their small size means that they reflect the immediate geographical neighbourhood well. In addition, using these as our neighbourhood definitions means that we can get very detailed information on a whole range of socio–economic indicators at two different neighbourhood levels. Further, this data does not need to be constructed from the sample, but instead can come from population values obtained through the census. Finally, socio–economic indicators at both the large and small neighbourhood level are likely to be highly correlated with the same indicators for 'correctly' specified neighbourhoods.

We have information on a range of neighbourhood characteristics at both the CD and the postcode level from the 1991 census. This includes information about male and female educational attainment, household and personal income and labour force status. We also have a neighbourhood socio–economic status (SES) variable which was constructed from 1991 census data at the CD level by **?**. As the name suggests, this variable is constructed to provide an indicator of the socio–economic conditions in a neighbourhood as a function of a number of characteristics including income, labour force status and educational composition. It is thus a neighbourhood equivalent of the individual SES variable that we also have available. Using this socio–economic status index has one key advantage – it captures the combined impact of a variety of neighbourhood characteristics that tend to be highly collinear. When we try to include these variables separately, the collinearity leads to high standard errors on the individual coefficients. Using the socio–economic index helps reduce this multicollinearity problem.[6]

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 other members of their household. Unfortunately, parental income is not well measured. Child reported income figures are available, but the response

[5]Loosely, CDs might correspond to blocks and postcodes to wards.

[6]The socio–economic index is constructed using factor scores from principal components analysis. The index is based on several variables: the proportion of the population in Professional, Administrative and Clerical occupations; the proportion of very high income earners; the number of families per house; the proportion of families who own or are purchasing their own home; the percentage of population with various qualifications; and the number of households with more than three cars.

rate is relatively low and the quality of the data is questionable. There is, however, detailed information about the occupational status and education levels of both parents, variables which are likely to be good proxies for income, especially permanent income. These variables are also likely to provide information about parental attitudes to education. Information on other important variables are also available, including the number of siblings and the type of school attended.

The following analysis is restricted to major urban areas, for two reasons. The first is that the ABS introduces sampling error into small postcodes to ensure confidentiality[7]. Second, the type of neighbourhood effects that we may expect in an urban context differ from those that we may expect in a rural context, and our interest lies predominantly with the former.

The large childhood neighbourhood 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[8]. Small childhood neighbourhood is defined as the collection district where the individual was living when they were 16. Again, 17 year olds in the 1989 sample were also allocated their childhood CD code. Childhood CD code is only recorded for the first two waves of the sample. The data is missing for all subsequent waves. We are thus left with two samples – an unrestricted sample for whom all postcode data is available; and a restricted sub–sample for whom all CD data is available. Below, we show that the characteristics of the restricted sub–sample are representative of the total sample. In addition, our initial specifications which do not incorporate the CD data allow us to compare the results from the restricted and the unrestricted sample. These results suggest that the restricted sample is representative in terms of behaviour as well. We will return to this issue below.

Childhood postcodes are only defined if the children are living with one or both or 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. Thus, if children who drop out are more likely to be living away from home, then we under sample this group of respondents. We also exclude respondents who reported that they had spent most of their life until 14 overseas. In addition we exclude respondents that are married. Neither of these sample restrictions changes the results in any fundamental way.

2.1 How representative is the CD sub–sample?

When we want to consider the importance of small neighbourhood effects we need to restrict the sample, as CD information is only available at the start of this period. Table 1 compares the characteristics of this restricted sample with the unrestricted sample. The table gives mean values for a number of key variables. The table suggests that the restricted sample is representative of the

[7]Although the CD data that we have do not suffer from that problem, so we could have reconstructed accurate small postcode data from the weighted average of constituent CDs.

[8]Excluding these individuals does not change the results, although it does reduce the accuracy of some point estimates.

	Full sample	Restricted sample
Male	49.7%	52.4%
Number of siblings	2.0	2.0
English good	10.0%	10.1%
English poor	1.0%	1.0%
Catholic school	22.7%	23.1%
Other non-government	9.4%	10.2%
Father's occ. status	29.7	30.7
Mother's occ. status	19.6	19.3
Father not present	15.5%	14.7%
Mother not present	5.0%	4.7%
Father degree	16.6%	17.4%
Father trade qualifications	16.9%	16.5%
Mother degree	12.8%	13.0%
Mother trade qualifications	4.2%	4.2%
Neighbourhood personal income	5601.5	5629.01
Neighbourhood percentage vocational	13.6%	13.3%
Neighbourhood unemployment rate	11.3%	11.5%

Table 1. CD sub–sample characteristics

total sample. In Sections 3.1 and 3.2 will see that the sub–sample also appears to be representative in terms of drop out behaviour.

3. Empirical model and results

We want to estimate the effects of two different types of neighbourhood on education decisions. In particular, we will consider the decision on whether or not to complete high school – legally, the first free education decision available to Australian teenagers. In Section 3.1 we present the results of estimation that ignores neighbourhood effects. As outlined above, we do this for both the unrestricted sample and the restricted sample. In Section 3.2 we add neighbourhood variables and discuss the importance of both small and large neighbourhood effects for understanding teenage education decisions. Finally, in Section 3.3 we check the robustness of our small neighbourhood results using a fixed effects logit specification. Effectively, this specification uses large neighbourhood dummies, rather than specific characteristics, to capture the large neighbourhood effects.

3.1 Individual effects

We start by estimating the model assuming that there are no neighbourhood effects. Thus, our basic specification is:

$$y_i^* = \alpha + z_i'\beta + \varepsilon_i; \tag{1}$$

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; and ε_i is the (normally distributed) error term which contains all the unobserved factors which affect individual i's propensity to leave school before the final year.

Because the observed variable is the zero–one drop out decision, rather than the underlying probability we estimate a probit model. As always, the magnitude of the effects of each variable depend on where the probability is evaluated. We present the results in marginal effects form so the coefficients give the impact of a one unit change in the variable, 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 drop out if the individual has that characteristic rather than the omitted characteristic. Data appendix **??** gives definitions of variables and specifies the omitted categories for each group of dummy variables. The results for the individual effects specification for the restricted and the total sample are presented in the first two columns of Table 2. Column 1 gives full sample results, column 2 restricted sample results.

We briefly discuss the outcomes for the full sample, before considering the differences between the samples. For the full sample, males are 8 percentage points more likely to drop out than females. Teenagers with more brothers and sisters are more likely to leave school early. Teenagers without English as a first language are significantly more likely to complete high school than teenagers who are born overseas[9]. 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 other non–government schools. A number of parental characteristics are important for explaining the propensity to drop out. High occupational status for fathers has a positive effect on the probability of staying on, as does having a mother or a father with a degree. Teenagers from single parent families are much more likely to drop out – particularly if it is their mother who is not present in the household.

Turning now to the restricted sample, we see that the results are broadly comparable. Only two coefficients change sign, and both are insignificant in both the restricted and the full sample estimation. The standard errors of point estimates are increased in the smaller sample and some variables that were significant become insignificant. Most noticeable among these is that the father degree and English as a foreign language variables are no longer significant. However, all other background variables and parental characteristics remain significant with the same sign. The behaviour in the restricted sample would appear to be representative of the total sample.

3.2 Neighbourhood effects

We start by considering the inclusion of large neighbourhood effects. We then consider the inclusion of small neighbourhood effects. Introducing them in this order allows us to check that the smaller restricted sample is, again, representative in terms of both characteristics and behaviour. The equation that we estimate is now:

$$y_i^* = \alpha + z_i'\beta + E(z|X_i)'\gamma + \varepsilon_i; \qquad (2)$$

where $E(z|X_i)$ are the average characteristics of the individuals in the large neighbourhood. Remaining notation is as for Equation 1. The third and fourth columns of Table 2 show the

[9] Mainly British immigrants.

	Individual		Neighbourhoods			
	(1)	(2)	(3)	(4)	(5)	(6)
Large Neighbourhood						
Average personal income			0.010	−0.002	0.001	
Proportion trade qual.			0.013**	0.015**	0.026**	
Unemployment rate			0.005	0.008	0.409	
Small Neighbourhood						
Average personal income					0.005	0.004
Proportion trade qual.					−0.010**	0.010
SES					−0.008**	−0.006
Personal Background						
Male*	0.080**	0.103**	0.081**	0.102**	0.100**	0.406**
Age	0.093**	0.034	0.093**	0.034	0.035	0.280*
Number of siblings	0.009*	0.025**	0.009*	0.025**	0.025**	0.108**
English not first language						
English good*	−0.080**	−0.046	−0.072**	−0.038	−0.040	−0.377
English poor*	−0.137**	−0.160	−0.129*	−0.153	−0.149	−0.409
Born overseas*	−0.077**	−0.102**	−0.076**	−0.103**	−0.106**	−0.600**
School						
Catholic*	−0.125**	−0.156**	−0.120**	−0.149**	−0.147**	−0.776**
Other non–government*	−0.170**	−0.153**	−0.161**	−0.144**	−0.142**	−0.597**
Parent Characteristics						
Father's occ. status @14	−0.002**	−0.002**	−0.002**	−0.002**	−0.002**	−0.010**
Mother's occ. status @14	−0.0001	0.0002	−0.0002	−0.0002	−0.0002	−0.001
Father not emp @14*	−0.017	−0.030	−0.016	−0.028	−0.031	−0.144
Mother not emp @14*	0.004	−0.022	−0.003	−0.025	−0.230	−0.237
Father not present @14*	0.059**	0.072	0.061**	0.076*	0.072	0.192
Mother not present @14*	0.443**	0.427**	0.445**	0.427**	0.426**	1.999**
Father has:						
degree*	−0.055**	−0.016	−0.052**	−0.011	−0.010	−0.136
trade qualifications*	0.020	0.002	0.016	−0.003	0.002	−0.131
other post-secondary*	−0.039	−0.044	0.045	−0.051	−0.045	−0.220
Mother has:						
degree*	−0.091**	−0.146**	−0.083**	−0.138**	0.122**	−0.687**
trade qualifications*	0.019	−0.005	0.017	−0.004	−0.010	−0.004
other post-secondary*	−0.025	−0.034	−0.025	−0.036	−0.035	−0.298
State						
Victoria	−0.025	−0.074**	0.001	−0.037	−0.024	
South Australia	−0.048*	−0.130**	−0.037	−0.104**	−0.090*	
Western Australia	−0.008	−0.074	−0.008	−0.073	−0.046	
Queensland	−0.072**	−0.143**	−0.045*	−0.100**	−0.088*	
Tasmania	0.005	−0.093	0.043	−0.029	−0.020	
ACT	−0.119**	−0.251**	−0.087**	−0.218**	−0.198**	
Number of obs.	4401	1654	4401	1654	1654	1372
Overall significance	854.2	256.5	869.35	262.5	269.7	159.6
	$\sim \chi(34)$	$\sim \chi(32)$	$\sim \chi(37)$	$\sim \chi(35)$	$\sim \chi(37)$	$\sim \chi(26)$

Note that time dummies and section of state dummies have been included in the estimation but are not reported. (*) indicates significance at the 10% level; (**) indicates significant at the 5% level. Average personal income is in '000s of dollars. Column (1) reports the same results as column (1) in Table **??**

Table 2. Small & large neighbourhood regression results

results when we include a number of neighbourhood variables. Column 3 gives the full sample results, column 4 the restricted sample results.

Choosing which neighbourhood variables to include is not easy as, a priori, all of them may have an important influence on drop out effect. However, unsurprisingly, collinearity problems dominate when all of the possible neighbourhood variables are included. Our empirical approach was to start with a large number of neighbourhood variables and test down to a more parsimonious representation. Initially we included the following variables separately for male and females: proportion of neighbourhood with a higher degree; proportion of neighbourhood with a degree; proportion of neighbourhood with various types of diploma; proportion of neighbourhood with skilled vocational qualifications; proportion of neighbourhood with basic vocational qualifications; proportion of neighbourhood with no qualifications; neighbourhood unemployment rate; personal income. We then tested to see if we could combine the male and female variables – we could never reject the hypothesis that the coefficients were the same. Next, we combined the degree qualifications, the vocational qualifications and the other post–secondary qualifications. Again, we could never reject the hypothesis that the coefficients on the sets of variables were the same. We then dropped the no qualifications and the other post secondary qualifications variables which were consistently very insignificant. This left us with percentage of neighbourhood with a graduate qualification; percentage of neighbourhood with a vocational qualification; neighbourhood unemployment rate and average neighbourhood personal income. Both average neighbourhood personal income and percentage with graduate qualification were insignificant, but would appear to be highly collinear. In the end, we present results after dropping the graduate qualification variable. Results are comparable if we drop the neighbourhood income variable instead.

As can be seen from Table 2, only one of the neighbourhood variables is significant – proportion of neighbourhood with vocational qualifications is significant at the 1% level. Average neighbourhood personal income is insignificant and neighbourhood unemployment rate is (just) insignificant. The results are somewhat surprising. Our personal prior was that neighbourhood effects would operate through concentrations of either high educated or low educated adults, or through income. The fact that they appear to work through the proportion of adults with vocational qualifications has two interesting interpretations.

First, this could reflect the importance of job networks as emphasised by, for example, **?**. Young people in these neighbourhoods have access to a larger social network of people that can get them in to jobs where schooling qualifications are not necessarily required. Informational effects may play an additional role – when young people assess the returns to formal education they may use people in their own neighbourhoods to inform that decision. A high proportion of vocationally qualified individuals earning a living from jobs that de–emphasise formal learning may lead to young people forming different opinions about the value of that formal education. The second interpretation is a more classical local labour market interpretation. High concentrations of vocationally qualified adults may indicate neighbourhoods with local labour markets where unskilled job opportunities are more readily available. Given the local nature of labour markets, it may be attractive for children to drop out in neighbourhoods where there are greater job opportunities for unskilled labour. These two channels may obviously interact – school drop outs may find it easier to get connected in to the local labour market when they know a high proportion of adults who work in that market. Notice that the coefficient on neighbourhood unemployment, although (marginally) insignificant, points to somewhat more 'negative' neighbourhood effects. Neighbourhoods with high unemployment rates tend to see higher drop out. This effect is presumably not a result of teenagers dropping out to work in the local labour market (where unemployment is high), but reflects negative feedbacks whereby, for example, a culture of high unemployment leads to high drop out rates and even higher local unemployment. We return to this issue below.

Before introducing small neighbourhood effects variables, we can again compare the results from the restricted sample to those from the unrestricted. From Table 2, column 4, we see that the neighbourhood effects, for significant variables, are almost identical. The differences between the individual, family and state effects remain as before. This suggests that, with the exceptions mentioned in Section 3.1, the restricted sample behaviour is representative of the overall sample, particularly when it comes to neighbourhood effects. Table 1 reinforces this impression. We see that in terms of neighbourhood characteristics, the restricted sample is highly representative.

We now introduce small neighbourhood effects for the restricted sample. Thus, the equation we now estimate is:

$$y_i^* = \alpha + z_i'\beta + E(z|X_i)'\gamma + E(z|x_i)'\delta + \varepsilon_i; \tag{3}$$

where $E(z|x_i)$ are the average characteristics of the individuals in the small neighbourhood. Remaining notation is as for Equation 2. Again, we test down from a much broader specification. This time, the process is helped because we have a neighbourhood SES variable, which captures income, occupation and employment characteristics allowing us to use this variable to avoid some of our earlier multicollinearity problems. After testing down, we are left with small neighbourhood variables that are very similar to the large neighbourhood variables. We have CD SES rather than personal income or graduate qualifications, percentage of the CD with vocational qualifications and the unemployment rate of the CD. Even in this parsimonious representation, the unemployment rate remains highly insignificant – so we drop this variable which leaves us with the specification reported in column 5 of Table 2.

The results are interesting, and highly informative with respect to the interpretations of the possible neighbourhood effects that we outlined above. First, notice that the significance and the sign of the coefficients on the large neighbourhood variables are unchanged. Second, both small neighbourhood SES and small neighbourhood proportion vocational are significant and have a negative effect on school drop out rate. Consider the negative effect of the proportion vocational education in the small neighbourhood. This suggests that the more classical local labour market interpretation may well be the correct one. A large neighbourhood with a high percentage of vocational educated adults proxies for high local demand for (complementary) unskilled labour. High local labour market demand for unskilled labour alters the incentives to drop out and drop out rates rise accordingly. However, conditional on that, a high concentration of vocational qualified adults in the smaller neighbourhood reduces the drop out rate. Informational networks would appear to play a small part in the effect of vocationally qualified adults on drop out. In fact, a high proportion of vocationally qualified adults in the small neighbourhood would appear to encourage students to stay on at school – possibly so that they can move on to more vocational training. At the same time, a low SES score in the small neighbourhood has a significant impact on school drop out rates. This suggests that there are small clusters of low SES families with much higher drop out rates than we would predict given family background and personal characteristics. Negative neighbourhood feedbacks would appear to occur at the small neighbourhood level acting through the socio–economic composition of that small neighbourhood.

3.3 Fixed effects estimation

In Section 3.2 we tested for the presence of small neighbourhood effects after conditioning on a number of large neighbourhood characteristics. A stronger test for the presence of small neighbourhood effects would involve conditioning out all of the variation in drop out probabilities that may possibly be due to large neighbourhood effects. In the specification in Section 3.2 neighbourhood effects work through neighbourhood average personal income, the proportion of adults with trade qualifications and the neighbourhood unemployment rate. In this section, we want to replace these variables with neighbourhood dummies, so that the dummies capture any difference in average drop out rates between large neighbourhoods, no matter what the cause. To do this, we would need to introduce individual neighbourhood effects to our specification which would condition out the average drop out propensity in the large neighbourhood, leaving small neighbourhood effects to explain the variation within those neighbourhoods.

For discrete dependent models, the choice between fixed and random effects models is somewhat constrained. Introducing fixed effects in to the standard probit specification is problematic. There is no feasible way to remove the heterogeneity from the nonlinear structure (by differencing for example) and with large numbers of cross–sectional units, estimation of the individual neighbourhood dummy coefficients is intractable. Some progress has been made on a probit specification incorporating random effects. However, as with standard linear formulations, we need to assume that the individual random effects are uncorrelated with the other regressors. If neighbourhood effects do occur at the small neighbourhood level, then this assumption is clearly unlikely to hold, precisely because each large neighbourhood is formed from a collection of small neighbourhoods.

Instead, we use a (Chamberlain) conditional (fixed effects) logit specification[10]. This specification represents the simplest way of conditioning out large neighbourhood heterogeneity. The basic idea is to consider the conditional likelihood function, where the likelihood for each set of neighbourhood observations is conditioned on the number of 1s in the neighbourhood. The fixed effects logit specification is:

$$\text{Prob}(y_{ij} = 1) = \frac{e^{\alpha_i + \beta' x_{ij}}}{1 + e^{\alpha_i + \beta' x_{ij}}}; \tag{4}$$

where α_i is the fixed effect for large neighbourhood i, and the x_{ij} are the characteristics of individuals that vary within large neighbourhoods. These characteristics include small neighbourhood characteristics as well as background variables.

[10] See **?**.

The conditional likelihood is

$$L^c = \prod_{i=1}^{N'} \text{Prob}\Big(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iN} = y_{iN} \Big| \sum_{n=1}^{N} y_{iN}\Big), \tag{5}$$

where N' is the number of neighbourhoods and N is the number of teenagers within each neighbourhood.

We must drop two types of neighbourhoods when implementing this procedure. The first are large neighbourhoods with only one observation. The second are large neighbourhoods where behaviour is uniform. That is, large neighbourhoods where everyone drops out, or where everyone stays on. Neighbourhoods with uniform behaviour do not contribute to the conditional likelihood function. We also drop variables that do not vary within groups[11]. See **?** for more details on implementing the conditional logit model. The restrictions leave us with a sample of 1372. The results for this fixed effects logit specification are reported in Column 6 of Table 2.

Unfortunately, parameter values are not directly comparable across the two different specifications. However, we can see that the sign and significance of variables remains unchanged, suggesting that our small neighbourhood results are robust to conditioning out all large neighbourhood effects whatever their cause. Most of the individual and family variables have the same sign and significance as the probit results suggesting that they are robust to alternative specifications of the large neighbourhood effects. All of the small neighbourhood variables are insignificant. However, the SES small neighbourhood variable is only just insignificant[12]. Given the reduction in sample size, and the small number of observations within each large neighbourhood, we were surprised that any of the small neighbourhood variables were close to being significant. The fixed effects logit results suggest that the trade qualification result is not robust. However, the small neighbourhood SES result is quite robust to a very general specification of the large neighbourhood effect. Living in an area where the immediate neighbourhood has low socioeconomic status has a negative effect on drop out propensities.

4. Conclusions

We have tested for the presence of both small and large neighbourhood effects on the drop out rate of Australian teenagers. Two neighbourhood effects appear to operate. The first works at the large neighbourhood level through the proportion of the adult population with vocational education. A high proportion of vocationally trained adults leads to a higher drop out rate. This would appear to be consistent with two possible mechanisms – one working through local labour market demand, the other through social networks. The results for the small neighbourhood variables suggest that the former is the most likely channel. A high proportion of adults with vocational qualifications in a small neighbourhood reduces drop out probability. This suggests that the high drop out rates associated with high concentrations of vocationally qualified adults reflect local labour market conditions. We have shown that the SES result for small neighbourhoods is quite robust to conditioning out all large neighbourhood effects.

[11]Specifically the state dummies.

[12]It is actually significant at the 11% level.

We have also found that the socio–economic status of small neighbourhoods matters for drop out rate. The channels through which this variable might operate are presumably those identified by **?**, **?** and others. Such channels include effects on the assessment of returns to education, the importance of social networks and the influence of peer–group pressure.

Our results here do not allow us to separate out the channels through which small neighbourhood socio–economic status influences drop out rates. Information on average neighbourhood drop out rates might allow us to do this, but such information is not available and only a very poor proxy can be constructed from the data given the number of observations in each small neighbourhood. Our results are clearer on the channel through which the structure of large neighbourhoods impact on drop out rates.

The policy implications of these results are interesting. First, the fact that large neighbourhood effects seem to operate through the structure of local labour market demand rather than through some other neighbourhood mechanism suggests that high drop out rates may sometimes be a rational response to perceived local labour market conditions. This suggests that local employers of large numbers of unskilled workers may need to play an important role if governments wish to reduce drop out rates in certain neighbourhoods. Second, the fact that small neighbourhood effects exist, and seem to operate through the socio–economic status of the neighbourhood suggests that government policies placing small clusters of low SES families in better neighbourhoods may have little significant impact on drop out rates. 'Forced' mixing through government housing programs may need to ensure that low SES families are well dispersed throughout more affluent neighbourhoods, rather than concentrated in 'sink' estates. Refining the policy implications will involve separating out the mechanisms through which the effects operate. This identification is left to further work.

Appendix A. AYS data appendix

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 whose last school 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 whose last school 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 or CD;
- 'proportion with grad qual' is the proportion of the respondent's postcode or CD 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 or CD who recorded having skilled vocational training or basic vocational training;
- 'unemployment rate' is the unemployment rate of the respondent's postcode or CD.
- the SES variable is a socio–economic status indicator based on factor–component analysis using a variety of socio–economic variables. It has a mean of 1.00 and a range of 0.714 to 1.42. See **?** for details of the construction.