Recreating David Card's "Using Geographic Variation In College Proximity To Estimate The Return To Schooling" 1993 Working Paper

Shane Ellard   *York University*

In this short paper I recreate the main results found by David Card in his 1993 working paper "Using Geographic Variation In College Proximity To Estimate The Return To Schooling".

The Economic Returns to Schooling

One persistent fact of labour markets is that more-educated workers earn more. However, a naïve comparison between the earnings of, for example, high school and college educated workers would yield a biased estimate of the return to schooling. Education levels are not randomly assigned. Rather, individuals make decisions about how much schooling they receive. If these decisions about education are correlated with an unobserved component of earnings, then a difference-in-means comparison will misstate the return to education. In this case, education is said to be endogenous to earnings and ordinary-least squares regression (OLS regression) will yield spurious results.

Instrumental variables are often used to address this issue. An instrument for educational attainment is some observable variable that affects the schooling decision but <u>does not</u> directly affect earnings. Variation in an instrument can therefore provide variation in schooling that is exogenous to earnings. It is that variation researchers exploit when using an instrument in so-called instrumental variable regression (IV regression).

In general valid instruments are hard to find. In his 1993 working paper[1], Card proposed using proximity to college as an instrument for educational attainment. It is easy to see why college proximity should affect the schooling decision. Living close to college gives students the option of living at home and thus lowers the cost of attending school. If it's also true that proximity to college does not directly affects earnings then it can be used as an instrument.

With this in mind Card drew data from the National Longitudinal Survey of Young Men (NLSYM)[2] to run series of basic OLS and IV specifications. The NLSYM is a longitudinal survey of young men in the U.S. and contains information on individuals' educational attainment, age, family background, and so on. Importantly, the data also includes dummy variables indicating if the individual lives near a 2-year and 4-year college. Using these data Card ran a series of basic OLS and IV specifications regressing log wages

---

[1] See Card 1993.
[2] Card has provided this data which is freely available through the Wooldridge data package.

on education, using proximity to college as an instrument for education in the latter case. In the coming sections I will replicate these main results.

Returns to Schooling using OLS

To begin his analysis Card estimated the returns to schooling using five OLS specifications, regressing log wages on education. These specifications are replicated in Table 1—all five specifications include a linear education term, an experience term, a race indicator, and dummy variables for residence in the south and in a metropolitan area (SMSA). The second adds indicator variables for the region of residence in 1966 and another for residence in a SMSA in 1966. The third adds measures of the mother's and father's education and the fourth adds an interaction between the parental education terms. The fifth specification adds indicator variables for family structure (e.g. single parent household).

All five of these specifications yield similar results for the returns on schooling. This can be quite easily seen by comparing the coefficients in the first row of Table 1. Regardless of the controls used, the basic OLS estimates imply a return to education of about 7.3 percent per year of schooling.

Despite the similarities across specifications one should still be skeptical of these results. Recall, we suspect that education is correlated with an <u>unobserved</u> part of earnings and that our estimate is therefore biased. In his 1993 paper Card outlined several reasons why one would worry this is the case. One reason is so-called "ability bias". If individuals have an unobserved characteristic "ability" that raises wages and also influences one to receive more schooling then the OLS estimate of the return to schooling will be biased upwards. Another potential issue that Card mentions is measurement error. Measurement error induces a negative correlation between the errors in earnings and schooling, this creates a downward bias in the OLS estimate of the return to schooling. Similarly, a downward bias occurs if returns to schooling are heterogeneous and if those with lower levels of education have higher returns to schooling[3].

Returns to Schooling using IV

Table 2 replicates the basic IV estimates of the return to education done by Card, using college proximity as an instrument for completed education. Specifically, a variable indicating whether one is near an accredited 4-year college is used as the instrument. Columns one and two report the coefficients for the first stage regressions (the effect of college proximity on years of schooling). Columns three and four contain the reduced form regressions (the effect of college proximity on log wages). Columns five and six report IV estimates of the return of schooling. The regressions in columns one, three and five exclude parental

---

[3]For a detailed discussion of these problems see Card's (1999) later work.

education and family structure controls whereas columns two, four and six include these controls.

As Card addressed, one problem is that if education is endogenous to earnings, then so is experience—as it is mechanically related to education. To address this, Card implemented the same IV specifications but used age as an instrument for experience. These specifications are replicated in Table 3. Like before, columns one, three and five exclude familial controls whereas columns two, four and six include them.

The two IV specifications shown in Table 2 and 3 yield similar results regardless of what controls are included or how experience is treated. Living near college had a strong positive effect on schooling, between .32 and .38 years, and a positive effect on wages, between 4.2 and 4.8 percent. Moreover, the IV specifications imply a return to education between 12.2 and 13.6 percent. As Card points out in his paper, the IV estimates on the return to schooling are about 50 percent higher than the corresponding OLS estimates.

To probe these findings Card implemented several alternative specifications—some of these specifications are replicated in Tables 4 and 5. Table 4 contains alternative OLS specifications and Table 5 contains the same alternative specifications but using IV for education. The first columns in each table contain the basic respective specifications from above (all the prior controls are included, and age is used as an instrument for experience). The second columns then control for "ability" as directly measured by the "Knowledge of World Work" (KWW) score. As seen in Table 4, in the OLS specification KWW is a significant predictor of earnings, and its inclusion reduces the estimated returns to schooling. However, when KWW is included in the IV specification the coefficient on KWW is small and statistically insignificant (see Table 5).

A criticism of this specification is that KWW is treated as an error-free measure of ability. To address this Card repeated the exercise but used IQ to instrument KWW. This specification is replicated in the third columns of Tables 4 and 5. As Card found, this lowers the IV estimate of the return to schooling but increased the standard errors of the coefficients of education and KWW score to where neither is statistically different from zero.

The final column in Table 5 uses an alternative measure of college proximity to instrument education. Specifically, indicators for proximity to a 4-year college and proximity to a 2-year college are both used as instruments. Using both indicators as instruments leads to a small reduction in the estimated returns to schooling, but nevertheless the implied return to education and standard errors are very similar to the previous IV estimations.

Altogether, the results in Tables 4 and 5 confirm our findings from the initial IV specifications. Regardless of what controls are used the IV estimates on the returns to schooling are 40 to 50 percent higher than the respective OLS estimates.

Is College Proximity a Valid Instrument?

As Card explains in his paper, for college proximity to serve as a valid instrument for education it must have no direct effect on earnings (that is, it must be exogenous to earnings). There are several reasons Card gives for why this may not be the case. One possibility is that families who place a large emphasis on education will choose to live closer to colleges. Their children might also have higher "ability" or may be more motivated to achieve labour market success. This would result in a positive correlation between college proximity and the unobserved part of the earnings equation, yielding an invalid instrument. A second possibility Card discusses is that the presence of a college could be associated with higher quality local primary and secondary schools. As school quality effects wages the omission of this variable would lead to an error in wages that is correlated to college proximity[4]. One last possibility Card mentions is that if there are only imperfect indicators for place of residence, and if those who grew up with a college nearby also live in higher-wage areas, then college proximity would be correlated with an unobserved geographic wage premium. Again, this would yield an invalid instrument.

Typically one can't simply test if an instrument is exogenous. However, in this case Card proposed an identification strategy. Recall, interpreting college proximity as a factor that lowers the cost of schooling it is reasonable that the effect should be largest for poor families. Prospective students from low-income families may greatly benefit by living near a 4-year college and reducing costs by living at home—whereas those from high-income families may be unaffected by this. Using this, Card proposed including college proximity in the earnings regression to test its direct effects and using an <u>interaction</u> between college proximity and an indicator for low parental education to instrument education.

This identification strategy is presented in Table 6 which includes the reduced form estimates and structural equation. Specifically, low family background is defined as neither parent graduating from high school. Then the interaction with this indicator variable and proximity to a 4-year college is used to instrument education. The coefficients in columns 1 and 2 verify that the effects of living near college are larger for those with low family background. Moreover, in column 3 we can see the estimated return to schooling is about the same as our prior IV estimates. It is also notable that in this regression the coefficient on college proximity itself is small and not significantly different from zero. Altogether, these results provide no evidence against the assumption that college proximity does not directly affect earnings and gives us more confidence that it is a valid instrument.

---

[4]See Card's earlier work with Krueger (1992).

Conclusion

This short paper recreated the main analysis done by Card in his 1993 working paper, where Card used college proximity as an instrument for education. The results from Tables 2 to 5 suggest that conventional OLS estimates of the return to schooling are downward-biased. In particular, the IV estimates suggest the return to schooling is about 12.2 to 13.6 percent whereas the OLS estimates suggest it is about 7.3 percent.

Like Card's findings, these results aren't changed by small changes in specification. Specifically, the inclusion of KWW scores and using proximity to 2-year and 4-year colleges as instruments did not significantly change the findings. Despite this, for college proximity to be a valid instrument we rely on the relatively strong assumption that college proximity has no direct effect on earnings. To test this I repeated the approach taken by Card: I included college proximity directly in the wage regression and used an interaction between poor family background and college proximity to instrument education. The results of this specification imply a return to schooling similar to the basic IV specifications. Also, the coefficient on the direct effect of college proximity was not significantly different from zero.

Like Card concluded, while these estimates are not very precise, they all imply that the return to schooling is relatively high for students with poor family background. As Card explains, these results also go against the conventional wisdom that individuals with higher education would earn above-average income regardless of education.

References

Card, D., & Krueger, A. B. (1992). Does school quality matter? Return to education and the characteristics of public schools in the United States. *Journal of political Economy*, 100(1), 1-40.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. *NBER working paper*, (w4483).

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, 3, 1801-1863

Table 1: OLS Regressions for Log Hourly Earnings

	(1)	(2)	(3)	(4)	(5)
Education	0.074	0.075	0.073	0.073	0.072
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Experience	0.084	0.085	0.085	0.085	0.085
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Experience Squared / 100	-0.224	-0.229	-0.230	-0.230	-0.233
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Race Indicator	-0.190	-0.199	-0.194	-0.195	-0.190
	(0.017)	(0.018)	(0.019)	(0.019)	(0.019)
Live in South	-0.125	-0.148	-0.146	-0.146	-0.148
	(0.015)	(0.028)	(0.028)	(0.026)	(0.026)
Live in SMSA	0.161	0.136	0.136	0.138	0.139
	(0.015)	(0.019)	(0.019)	(0.020)	(0.020)
Num.Obs.	3010	3010	3010	3010	3010
R2	0.291	0.300	0.301	0.302	0.304
R2 Adj.	0.289	0.296	0.297	0.296	0.298

Notes: Column 2 includes indicators for regions. Column 3 includes mother and father education levels. Column 4 includes interactions between these education terms. Column 5 includes indicators for family structure.

Table 2: IV Regressions for Log Hourly Earnings

	Education	Education (2)	Earnings	Earnings (2)	IV	IV (2)
Education (Fit)					0.132	0.136
					(0.054)	(0.056)
Near College	0.320	0.319	0.042	0.043		
	(0.085)	(0.081)	(0.018)	(0.017)		
Num.Obs.	3010	3010	3010	3010	3010	3010

Notes: Columns 1, 3, and 5 include controls for experience and indicators for race and regions. Columns 2, 4, and 6 includes controls for parents education and family structure. Age used as instrument for experience.

Table 3: IV Regressions for Log Hourly Earnings, Age as Instrument for Experience

	Education	Education (2)	Earnings	Earnings (2)	IV	IV (2)
Education (Fit)					0.122	0.130
					(0.046)	(0.050)
Near College	0.382	0.357	0.047	0.046		
	(0.114)	(0.104)	(0.018)	(0.018)		
Num.Obs.	3010	3010	3010	3010	3010	3010

Notes: Columns 1, 3, and 5 include controls for experience and indicators for race and regions. Columns 2, 4, and 6 includes controls for parents education and family structure. Age used as instrument for experience.

Table 4: Alternate OLS Specifications for Log Hourly Earnings

	Basic OLS Estimate	KWW	KWW with Instrument
Education	0.072	0.055	0.059
	(0.004)	(0.005)	(0.012)
Num.Obs.	3010	2963	2040

Notes: Column 1 includes controls for experience, race, regions, parental education and family structure. Age is used as an instrument for experience. Column 2 includes KWW score. Column 3 uses IQ as an instrument for KWW.

Table 5: Alternate IV Specfications for Log Hourly Earnings

	Basic IV Estimate	KWW	KWW with Instrument	Near 2-yr & 4-yr College
Education (Fit)	0.130	0.132	0.082	0.121
	(0.050)	(0.079)	(0.086)	(0.051)
Num.Obs.	3010	2963	2040	3010

Notes: Column 1 includes controls for experience, race, regions, parental education and family structure. Age is used as an instrument for experience. Column 2 includes KWW score. Column 3 uses IQ as an instrument for KWW score. Column 4 does not include KWW score, but uses proximity to 2-year and 4-year colleges as instruments for education.

Table 6: Interaction of Parent Education and College Proximity as Instrument for Education

	Education	Earnings	Structural Model
Education (Fit)			0.137
			(0.098)
Near College	0.190	0.023	-0.003
	(0.138)	(0.025)	(0.039)
Near College * Low Parental Education	0.337	0.046	
	(0.188)	(0.033)	
Num.Obs.	3010	3010	3010

Notes: All columns includes controls for experience, race, regions, parental education and family structure. Age is used as an instrument for experience. The interaction term between low parental education and near 4-year college is used as an instrument for education.