Reanalysis of Empirical Research for Return to Schooling: a Weak Instrument Variable Problem

Zejun Jiang, Ke Lyu

School of Economics, University of Edinburgh, the United Kingdom

Keywords: Return to schooling, Quarter of birth, Instrumental variable, Endogeneity

Abstract: Previous researches indicated a positive causal effect of education on earnings. In these studies, a widely acknowledged point is the endogeneity of education which causes bias for estimate. One common method to deal with the endogenous problem is to introduce instrumental variables. Angrist and Krueger employ the quarter of birth as the instrument, however, subsequent researches point out that the quarter of birth is a weak instrument because of weak relationship with education leading to implausible results. We still use the quarter of birth as instrument and data of Angrist and Krueger's research in our supplementary work to focus on how to check the existence of weak instrument problem and then obtain the weak-identification-robust inference in return to schooling. The first finding is that the quarter of birth may not bring about weak instrument problem at the 30% bias level. The second is that we gain the credible confidence interval for the estimate of our interest under three weak-identification-robust test, namely the Conditional Likelihood Ratio (CLR) test, Anderson and Rubin (AR) test, and Kleibergen (K) test. Furthermore, we want to get data from a tracking survey of the same people to ensure that the return to schooling changes for individuals during the whole life.

1. Introduction

Over the past century, many scholars devoted to measure the causal effect between education attendance and earnings. They conclude that better-educated individuals could experience less unemployment and obtain higher earnings in labour market, therefore, a positive relationship is confirmed for the schooling and wages (Card, 2010) [8]. However, since the possibility of education affecting other factors could not be ruled out, there would be something unobserved which influence both education and earnings, namely endogeneity problem in econometrics (Hayashi, 2000)[11].

In the presence of endogeneity, part of studies concentrates on introducing instrumental variables which require validity and strength. The article written by Angrist and Krueger (1991) [2] uses the quarter of birth as an instrument because the quarter of birth makes an effect on education through compulsory education law. Unfortunately, because of the limited econometric methodology during that period, the quarter of birth, a weak instrument, has not been testified for its strength. Skeptics note that the quarter of birth is weakly related to education and most of them substitute the quarter of birth with other instruments (Bound, 1996) [4].

The aim of our paper is to survey the weak-identification-robust estimate refining results of the 1920s group in the research of Angrist and Krueger (1991) [2]. Firstly, the weak instrument test conducted by Stata command weakivtest illustrates comparably lower statistic value meaning that the model has weak instrument problems. Secondly, several weak-identification-robust tests are applied using command weakiv for interval estimation and more convincing conclusion.

2. Literature Review

The study of the association between education and labour market has been revival (Card, 2013) [9]. One reason might be the "return to education", especially in the resurge US labour market, Katz and Autor (1997) [3] implement a research of the growing discrepancy between more and less-educated workers. Another rational reason attributes to the increasing interest in the determinants of

  DOI: 10.25236/isetbm.2019.040

economic growth as well as the role of human capital in the development process (Topel, 1999) [25]. Eventually, the secondary and post-secondary school enrolment rates grew rapidly, which bring about new attention about the relative benefits (Card, 1993) [7]. Therefore, it is important and necessary to put forward the idea of study the association between individuals' educational attainment and earnings.

Endogeneity is the main problem effecting the measurement of "return to education". The most common way to solve the endogenous problem of explanatory variables, which always lead to the OLS shows a biased estimate, is to find instrumental variables (IV) estimation (Bound & Jaeger & Baker, 1995) [5]. Consequently, Angrist and Krueger (1991) [2] try and find an instrumental variable which is the date of birth measured by the quarter of birth. According to their findings, two main reasons of this choice, one is because the quarter of birth has its exogeneity, which not affect income directly: not correlated with ability, motivation, family connections, etc.

However, there are criticisms about Angrist and Krueger's article. Bound and Jaeger (1996) [4] argue that the relationship between the quarter of birth and earnings can hardly be explained as the quarter of birth is a weak instrument. Herman Van Dijk (2007) [10] employs different the quarter of birth only implies a maximum difference of one year, which illustrates a slight influence on return to education. As it shows, most of the criticisms are about the quarter of birth is a weak instrument, so that there exists bias when we estimate.

One common method to solve the weak instruments problem is to change another strong instrumental variable, such as parental education (Card, 1993) [7], college distance and college credit equivalents (Kane & Rouse,1993) [14], education background and college availability (Pons & Gonzalo,2001) [21], etc. Nevertheless, as the continuous development of statistics and econometrics, new statistical approaches to instrumental variables are evolving, like Conditional Likelihood Ratio (CLR) test (Moreira, 2003) [19] , Anderson and Rubin test (Anderson & Rubin,1949) [1], Kleibergen test (Kleibergen, 2005) [16]and Jann test (Jann, 2010) [13]. However, only a small group of researchers use these methods in the study of return to schooling.

3. Data

The data used in this paper is the same as what Angrist and Krueger used in their research (1991), which can be obtained on MIT's website1. But here, we only select one of their cohorts, the 1920s, as our data set meaning men who were born in 1920s in the US (only men with non-negative earnings are included in the dataset). We have referred the 1970 Census from the Public Use Samples of Basic Records. The sample selected whose age, race and weeks worked are allocated by the US Government are excluded, and only the men with positive earnings are included in our dataset to make it more meaningful for further analysis and discussion. The detailed data used is described in Table 1 below.

Table 1 Character of Dataset

Cohort	1920-1929
Source	1970 Census
Sample Size	247,199
Variables (Mean-Standard deviation)	
Log wage	5.1552
Education	11.4933
Age	44.7257

The birth of the men is detailed to the quarter of the birth, and the earnings are shown by the males' income per week, which is simply calculated by dividing their earnings within a year by the weeks they have worked throughout the year. Since the earning was originally described in discrete times, we transform it into continuous time by calculating the average value of the start and end points of the intervals, which are $100 in this dataset. With regard to the weeks worked, it is similar to the earnings, as this variable is transformed into the continuous type by dividing them into

[1] http://economics.mit.edu/faculty/angrist/data1/data/angkru1991

several intervals. In addition, some control variables are included in the choice of datasets. In the Census of 1970, 9 year-of-birth dummies, 8 region-of-residence dummies, SMSA status dummy, marital status dummy, age, age squared, and race choices are all included and selected by this dataset. These dummy variables are easily controlled and provide simple and direct intuition. For instance, the marital status dummy variable will be 1 if the male in the dataset has married with his wife currently, and the SMSA status dummy variable will be 1 if the male is working in an SMSA, and it would be 0 otherwise.

4. Economic Methods and Issues

4.1 Model Specification

Basic model for this paper could be written as

$$E_i = \Pi X_i + \Sigma_c \delta_c Q_{ic} + \Sigma_c \Sigma_j \theta_{jc} Y_{ic} Q_{\mathrm{ij}} + \epsilon_i \qquad (1)$$

$$\ln W_i = \rho X_i + \Sigma_c \xi_c Y_{ic} + \beta E_i + \mu_i \qquad (2)$$

where E_i is the total education period for ith individual, X_i is a vector of control variables which is the main difference among four models, Q_{ic} is a dummy variable for ith individual whose quarter of birth is j (j=1,2,3), Y_{ic} is a dummy variable for ith individual who is born in year c (c=1,2,3….10) and W_i is the weekly wage. β is the parameter of our interest and represents the return to schooling. μ_i is the error term. As for the dependent variable of the model, Card and Krueger (1992) [6] find little proof that the model to estimate the returns to education is non-linear. On top of that, Heckman and Polachek (1974) [12]show that the function of earnings can be approximated as a log-linear model, which indicates that the estimates attained by Angrist and Krueger's model are able to, to some extent, represent the average return to education based on their sample data. Thus, the paper accepts and employs the log-linearity characteristic to conduct estimates by using log weekly wage as the dependent variable to better approximate the models. Specifically, using the form of log wage can effectively eliminate the heteroscedasticity issue, as this form decreases the measurement units and transforms the absolute error into relative error, which is significantly lower than the absolute error.

In this model, E_i is endogenous because of omitted variables. Hayashi (2000) [11] argues, unobserved ability can affect wage which means this factor will occur in the error term μ_i, and this ability will also influence education. Therefore, the unobserved ability is one omitted variable in this case and bring about the endogenous problem which will bias the estimate. As the previous discussion, valid instruments should be introduced to resolve endogeneity, as well as the quarter of birth dummies and the product of the quarter of birth dummies and the year of birth dummies are unrelated with μ_i but correlated with E_i, so these dummies are chosen as excluded instruments. In addition, to partial out, the effect from other potential factors and thus get better estimates (Kough and Mark, 2015) [17]. In the 3.4 section, we find some variables also have a relationship with earnings and hence add them in this model as control variables, such as 9 year-of-birth dummies, 8 region-of-residence dummies, SMSA status dummy, marital status dummy, age, age squared, and race choices. In this paper, four models followed Angrist and Krueger (1991) [2] are applied to obtain estimates using different control variables which is listed in Table 2.

Table 2 Four Model Settings

	Model 1	Model 2	Model 3	Model 4
Endogenous variable(E_i)	Years of Education	Years of Education	Years of Education	Years of Education
Instrumental variables ($Q_{ic}, Y_{ic}Q_{\mathrm{ij}}$)	Quarter of Birth Year of Birth *Quarter of Birth	Quarter of Birth Year of Birth *Quarter of Birth	Quarter of Birth Year of Birth *Quarter of Birth	Quarter of Birth Year of Birth *Quarter of Birth
Control variables (X_i)	9 Year-of-Birth dummies	9 Year-of-Birth dummies, Age, Age-squared	9 Year-of-Birth dummies, Race, Married, SMSA, 8 Region-of-residence dummies	9 Year-of-Birth dummies, Age, Age-squared, Race, Married, SMSA, 8 Region-of-residence dummies

4.2 Identification Tests and Weak-Identification-Robust Tests

Two diagnostics tests using point estimate, Montiel-Pflueger robust weak instrument test (Montiel Olea and Pflueger, 2013) [20] and Cragg-Donald Wald test (Stock-Yogo, 2002) [23], are applied to test weak identification problem.

As for the Cragg-Donald Wald Statistic, the statistic is calculated as follows:

$$F = (Y' \ M_Z Y)^{-1/2} Y^{\perp'} \ P_{Z\perp} Y^{\perp} (Y' \ M_Z Y)^{-1/2} \frac{(T - K_1 - K_2)^2}{K_2}$$

where the matrix

$$Y^{\perp'} \ P_{Z\perp} Y^{\perp} = (M_X Y)' \ M_X Z((M_X Z)' \ M_X Z)^{-1} (M_X Z)' \ (M_X Y)$$

The minimum eigenvalue of F is the statistic value used for diagnostic tests. In this case, if there is one endogenous variable, the eigenvalue for F is the F-statistic mentioned by Staiger and Stock (1997) [22]. According to Stock and Yogo (2002) [23], if two instrumental variables are used for one endogenous variable, the F-statistic (at the significance level of 0.3) critical value is 4.31.

Another test is Montiel-Pflueger test which is robust for linear instrumental variables regression in the first stage. The Montiel Olea and Pflueger effective F statistic could be calculated as follows:

$$F = \frac{Y' \ P_Z Y}{tr(\widehat{W})}$$

where the matrix

$$\widehat{W}^2 = \frac{(Y' \ P_Z Y)' \ (Y' \ P_Z Y)}{T - K_1 - K_2}$$

The effective F statistic coincides with the Cragg-Donald Wald test, however, $\widehat{W}$ is estimated which generate difference for critical values.

Despite those diagnostics tests, weak-identification-robust tests are used for further estimate. Contrast to traditional diagnostic tests using point estimate, weak-identification-robust tests an estimate confidence intervals, thus, efficient estimate could be obtained despite weak instrument problem.

Anderson and Rubin (1949) [1] propose a joint test (AR) of structural parameters β. The two parts in null hypothesis could be discomposed in to the K test and J test respectively. K test is proposed by Kleibergen (2005) [16], as a Lagrange multiplier test, also called the score test. However, both the AR test and K test are now deprecated since Moreira (2003) proposes a Conditional Likelihood Ratio (CLR) test that dominates it. The CLR test has the same hypothesis as the K test but with better power and size properties under homoskedasticity (Moreira, 2003) [19]. In theory, either the AR test or the CLR test can be inverted to produce a confidence region for the parameter, but the AR test is much easier to work with.

5. Empirical Results

This section gives part of the empirical outcomes of weak-identification tests based on two different methods, namely Montiel-Pflueger F test and Cragg-Donald Wald test. Besides, some main weak-identification-robust inferences for the return to schooling are also shown and analysed according to results collected in Table 3.

Firstly, the part of weak-identification test in Table 3 reveals the results of Montiel-Pflueger F test (Montiel Olea & Pflueger, 2013) [13]and Cragg-Donald Wald test (Stock-Yogo, 2002) [23]. Both of Montiel-Pflueger F statistic and Cragg-Donald Wald statistic are lower than their critical values at the bias level of 5%, which implies that the weak instrument problems exist in all of the four models, thus TSLS estimates might have a bias. In addition, according to small statistic and p-value of the J test in each model, there is no over-identification instrument problem for these two cohorts. Secondly, under the existence of weak instrument problems, apart from finding a strong instrument, another method is to obtain robust estimates by using weak-identification-robust inference tests, such as Conditional likelihood ratio test (Moreira, 2003) [19], Anderson-Robin test (Anderson & Robin, 1949) [1], and Kleibergen K test (Kleibergen, 2005) [16], which are based on interval estimate.

Before talking about the robust estimates for the return to schooling in two cohorts, one interesting finding emerges in the interval estimate process. For men born in the 1920s, in Model 1 and Model 3, all of CLR test, K test and AR test reject the null hypothesis since ß=0 is not included in the confidence set at the significance level of 5%. However, tests of Model 2 and Model 4 give contradictory outcomes where the null hypothesis is failed to reject because the confidence interval for the estimate is infinite and thus includes ß=0. However, the main reason for infinite confidence set in CLR test, K test, and AR test comes from the distortion of weak instruments in testing equations, and she further gives an assertion that if weak problems exist, the confidence set for those three tests should be infinite. This argument indicates one probable speculation is that the weak-identification problem may not exist in Model 1 and 3.

Some evidence supports this speculation. When comparing statistics of three tests with their critical values at 30% bias level, the Cragg-Donald Wald test illustrates in Model 2 and 4 estimator is greater than 30% of the worst-case bias, while we could tolerate a bias which is up to 30% in Model 1 and 3. This means that one possible situation is that for Model 1 and 3, the estimate may not be interfered by the weak-instrument problem and thereby the results of TSLS of Model 1 and 3 are credible at the 30% bias level. In addition, the TSLS estimates in Model 1 and 3 are all included in the confidence sets of three weak-identification-robust tests no matter in 1920s cohort.

Table 3 Weakivtest and Weakiv Results for Men Born 1920-1929: 1970 Census

	Model 1	Model 2	Model 3	Model 4
TSLS ß(p-value)	0.0769*** (0.0004)	0.1303** (0.0334)	0.0669** (0.0151)	0.1007** (0.0334)
Confidence set	(.047, .106)	(.066, .196)	(.037, .096)	(.035, .166)
Weak Identification Test				
Montiel-Pflueger F statistic H0: Worst case bias>5% Worst case bias>30%	4.599 (22.351) (5.301)	1.085 (22.346) (5.326)	4.553 (22.351) (5.301)	1.025 (22.346) (5.326)
Cragg-Donald Wald statistic H0: TSLS bias>5%	4.681 (21.42)	1.00 (21.42)	4.59 (21.42)	1.00 (21.42)
TSLS bias>30%	(4.31)	(4.31)	(4.31)	(4.31)
Weak-Identification-Robust Inference				
Moreira CLR test H0: ß=0(p-value)	15.52*** (0.0005)	15.29 (0.1244)	12.30*** (0.0020)	9.15 (0.2920)
Confidence set	[.036, .115]	(-inf, +inf)	[.026, .105]	(-inf, +inf)
Kleibergen K test H0: ß=0(p-value)	10.96*** (0.0005)	1.42 (0.2332)	8.92*** (0.0028)	0.35 (0.5544)
Confidence set	(-inf,-1.806] U [.034, .117] U [1.298, +inf)	(-inf, +inf)	(-inf, -1.880]U [.025,.106]U [1.325, +inf)	(-inf, +inf)
Anderson-Robin test H0: ß=0 and E(ziεi)=0 (p-value)	51.54*** (0.0085)	36.84 (0.1225)	46.46*** (0.0281)	33.71 (0.2107)
Confidence set	[.025, .126]	(-inf, +inf)	[.009, .122]	(-inf, +inf)

***, **,* indicates that the coefficient is significantly different form 0 at the 1%, 5% and 10% significance levels, respectively.

AR test always has a wider confidence set than the CLR test and K test have. It may arise from that the actual estimator in the test equation of AR test is the product of the coefficient of instruments in the first stage and the difference between the estimate and the true value of ß (Kough and Mark, 2015) [17]. Therefore, if the weak identification problem exists, the coefficient of instruments will be small and bring about a wider range of possible value of estimate researchers will fail to reject.

In conclusion, the TSLS estimate for return to schooling is around 6%-7% for men born in 1920-1929. For confidence set from three weak-identification-robust tests, it is approximately [.04, .11] in the group. This confidence set is wider than that from the TSLS method, since when weak instruments only provide little or no information into the estimation process, a robust confidence set should be large or even possibly infinite to describle sampling uncertainty observed in data (Anna Mikusheva, 2007) [18].

6. Conclusion

This paper has used two common methods to test the strength of instruments deployed in the TSLS regressions under four models in three age-cohorts. One evidence is that weak instrument problems do exist in our four models for each cohort when we set the worst case bias level at 5%. But if we can tolerate the bias up to 30% worst-case bias level, we conclude that the estimate for return to schooling is credible in Model 1 and Model 3.

A robust confidence set should be large or even possibly infinite to describle sampling uncertainty observed in data when weak instruments only provide little or no information into the estimation process (Anna Mikusheva, 2007) [18]. And combined with that the bias problems may occur on relevant estimators at the 5% worst-case bias level when instrumental variables including the quarter of birth and the product of the quarter of birth and the year of birth, are weakly correlated with the year of education, the causality resulted from the TSLS regression should be treated with caution since they may be biased. In order to get a robust estimate for return to schooling under the existence of a weak-identification problem, it is necessary to try different weak-identification-robust methods for better results. This paper mainly applies three tests which are based on interval estimate rather than point estimate utilized by TSLS regression. The weak-identification-roust confidence sets obtained in the CLR test, K test, and AR test is consistent with the estimate got from the TSLS in each cohort, which illustrates that the outcome is plausible to a certain degree in Model 1 and Model 3, although in Model 2 and 4 the confidence set is infinite due to weak instrument. Some other researches have shown results for the possible estimate of the return to schooling. They get similar outcomes to what has shown in this paper. For example, Staiger and Stock (1997) [22] used the same data for men born in the 1930s and 1940s from the 1980 Census, and Kane and Rouse (1995) [15] focused on people born around 1960. Their results were almost the same as what we have obtained in this paper, which also affords evidence that the quarter of birth may not have a serious bias problem.

References

[1] Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. The Annals of Mathematical Statistics, 20(1), 46-63.

[2] Angrist, J. D. & Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? The Quarterly Journal of Economics, 106(4), 979-1014.

[3] Autor, D. H., Katz, L. F., & Krueger, A. B. (1997). Computing inequality: have computers changed the labor market?. Social Science Electronic Publishing, 113(4), 1169-1213.

[4] Bound, J., & Jaeger, D. A. (1996). On the Validity of Season of Birth as an Instrument in Wage Equations: A Comment on Angrist & Krueger's “Does Compulsory School Attendance Affect Schooling and Earnings?” (No. w5835). National Bureau of Economic Research.

[5] Bound, J., Jaeger, D. & Baker, R. M. (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. J. Am. Statist. Ass., 90, 443-450.

[6] Card, D. & Krueger, A. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. Journal of Political Economy, 100(1), 1-40.

[7] Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Nber Working Papers, 1127--1160.

[8] Card, D. (2010). Earnings, schooling and social capital of cooperative managers. Annals of Public & Cooperative Economics, 77(1), 1-20.

[9] Card, D. (2013). Estimating the returns to schooling: implications from a dynamic discrete choice model ☆. Labour Economics, 20(1), 92-105.

[10] Chesher, A., Dhaene, G., & Dijk, H. V. (2007). Endogeneity, instruments and identification. Journal of Econometrics, 139(1), 1-3.

[11] Hayashi, F. (2000). Econometrics. Princeton and Woodstock; Princeton University Press.

[12] Heckman, J. & Polacheck, S. (1974). Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship. Journal of the American Statistical Association. 69, 350-354.

[13] Jann, B., & Jann, K. (2010). Immunoelectrophoretic patterns of extracts from all escherichia coli o and k antigen test strains correlation with pathogenicity. Acta Pathologica Et Microbiologica Scandinavica, 78B (2), 263-264.

[14] Kane, T. J., & Rouse, C. E. (1993). Labor market returns to two-and four-year colleges: is a credit a credit and do degrees matter? (No. w4268). National Bureau of Economic Research.

[15] Kane, T. J., & Rouse, C. E. (1995). Labor-market returns to two- and four-year college. American Economic Review, 85(3), 600-614.

[16] Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. Econometrica, 73(4), 1103-1123.

[17] Kough, A. S., Paris, C. B., Behringer, D. C., & Iv, M. J. B. (2015). Modelling the spread and connectivity of waterborne marine pathogens: the case of pav1 in the caribbean. Ices Journal of Marine Science, 72(10), 403-4.

[18] Mikusheva, A. (2007). Uniform inference in autoregressive models. Econometrica, 78(5), 1411-1452.

[19] Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. Econometrica, 71(4), 1027-1048.

[20] Olea, M., & Luis, J. (2013). Essays on econometrics and decision theory. Technical Report of Ieice Prmu, 99(4), 99-104.

[21] Pons, E., & Gonzalo, M. T. (2001). Returns to schooling in spain: how reliable are iv estimates?. Working Papers (446).

[22] Staiger, D. & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. Econometrica, 65, 557-586.

[23] Stock, J. H., & Yogo, M. (2002). Testing for weak instruments in linear IV regression.

[24] Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. Journal of Business & Economic Statistics, 20(4), 518-529.

[25] Topel, R. (1999). Chapter 44 labor markets and economic growth. Handbook of Labor Economics, 3(3), 2943-2984.