Investigating the problems associated with using weak instruments for estimation of confounded causal effects

COMP3224 - Homework 5

Etienne Latif

April 2022

Abstract

Bound et al. (1995) lays forth a compelling paper describing the limitations of the uses of instrumental variables in estimation of causal effects when the instruments in question are weak [1]. We describe the mathematical derivations behind the findings that 2-stage least squares estimates for weak instruments are inconsistent when the instruments are not legitimate, and biased in finite samples when they are. We also present simulations to demonstrate these analytical results, finding that instrumental variable estimates become poor when we use weak instruments. Finally, we discuss the Angrist and Krueger (1991) case study [2], and give closing remarks on the implications of the paper. Ultimately, we find that instrumental variables estimation is likely to be far less useful than was originally believed, due to the strict conditions that need to be met for estimates based on the method to be unbiased or even consistent.

1 Introduction

In this paper we shall discuss the problems described by Bound et al. (1995) that arise when using 'weak instruments' for estimation of causal relationships between an endogenous treatment variable and a response [1], and come to a conclusion on the implications of such problems on the use of instrumental variables for causal inference, particularly in the field of economics. The Bound et al. paper demonstrates their analysis on a paper by Angrist and Krueger (1991) which uses instrumental variables to estimate the effect of education on wages [2], asserting that the results they found may in fact be biased. We'll begin by providing a brief overview of instrumental variables estimation and set up a canonical example system to base our inference on.

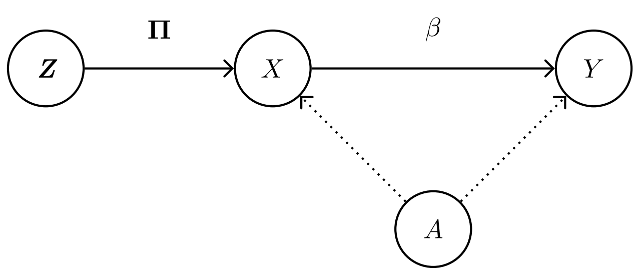


Figure 1: The canonical DAG for instrumental variable problems

Figure 1 exhibits a typical problem in which we may wish to estimate the causal effect of a treatment variable X on a response Y, but are confronted with a back-door pathway $X \leftarrow A \rightarrow Y$ via some set of unobserved confounding variables A. In this case, we call X an endogenous treatment variable and A a set of common exogenous variables to X and Y. The problem can be expressed by the following set of equations, based on the notation used by Bound et al.:

$$\begin{aligned} \boldsymbol{y} &= \beta \boldsymbol{x} + \boldsymbol{\varepsilon} \\ \boldsymbol{x} &= \mathbf{Z}\boldsymbol{\Pi} + \boldsymbol{\epsilon} \end{aligned} \tag{1}$$

Where $\boldsymbol{y}$, $\boldsymbol{\varepsilon}$, $\boldsymbol{x}$ and $\boldsymbol{\epsilon}$ are $N \times 1$ independent observation vectors from random variables Y, ε, X and ϵ respectively; β is a constant scalar value representing the causal effect of X on Y; $\mathbf{Z}$ is an $N \times K$ matrix where each row is an independent observation of the vector of instruments $\boldsymbol{Z} = (z_1, z_2, ..., z_K)$ and $\boldsymbol{\Pi}$ is a $K \times 1$ vector of constants. Note that we have assumed independence between $\boldsymbol{Z}$ and ϵ but some dependence between X and ε as a result of the endogeneity of X. Ideally, we would like that $\boldsymbol{Z}$ and ε are independent, implying that Y depends on $\boldsymbol{Z}$ only through X. We call the instruments in such a case 'legitimate'. We have not expanded these equations to include the common exogenous variables A as Bound et al. note that this will complicate the algebra while having no effect on the results we obtain.

Clearly β is our parameter of interest. One common way to estimate this value is with an ordinary least squares (OLS) estimator, given by $\hat{\beta}_{\text{ols}} = (\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{y}$. However, we shall demonstrate in the next section that the OLS estimator is biased and inconsistent under the conditions we have described. The estimator of interest is the 2-stage least squares estimator, which is obtained, intuitively, by two steps of OLS:

1. Obtain the OLS estimator from regression of the first stage $\hat{\boldsymbol{\Pi}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\boldsymbol{x}$ and take the fitted values $\hat{\boldsymbol{x}} = \mathbf{Z}\hat{\boldsymbol{\Pi}} = \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\boldsymbol{x} = \mathbf{P_Z}\boldsymbol{x}$, where $\mathbf{P_Z}$ is the projection matrix (see Appendix A.1).

2. Use the fitted values from the first stage estimation in place of $\boldsymbol{x}$ in the OLS estimator of the second stage $\hat{\beta}_{\text{iv}} = \left(\hat{\boldsymbol{x}}^T\hat{\boldsymbol{x}}\right)^{-1}\hat{\boldsymbol{x}}^T\boldsymbol{y} = \left(\boldsymbol{x}^T\mathbf{P}_{\mathbf{Z}}^T\mathbf{P_Z}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^T\mathbf{P}_{\mathbf{Z}}^T\boldsymbol{y} = \left(\boldsymbol{x}^T\mathbf{P_Z}\boldsymbol{x}\right)^{-1}\boldsymbol{x}^T\mathbf{P_Z}\boldsymbol{y}$ (simplifications result from the property of being idempotent, mentioned in Appendix A.1).

2 Inconsistency

2.1 Analytical results on the asymptotic behaviour of instrumental variables

We begin by demonstrating the conditions under which the ordinary least squares (OLS) estimation of this parameter is consistent:

$$\begin{aligned} \hat{\beta}_{\text{ols}} &= \frac{Cov(X,Y)}{Var(X)} \\ &= \frac{Cov(X,\beta X + \varepsilon)}{Var(X)} \\ &= \frac{Cov(X,\beta X)}{Var(X)} + \frac{Cov(X,\varepsilon)}{Var(X)} \\ &= \frac{\beta Var(X)}{Var(X)} + \frac{Cov(X,\varepsilon)}{Var(X)} \\ &= \beta + \frac{Cov(X,\varepsilon)}{Var(X)} \\ \implies \text{plim } \hat{\beta}_{\text{ols}} &= \beta + \frac{\sigma_{X,\varepsilon}}{\sigma_X^2} \end{aligned} \tag{2}$$

We then conclude that $\hat{\beta}_{\text{ols}}$ is a consistent estimator of β if and only if $\sigma_X^2 \neq 0$ and $\sigma_{X,\varepsilon} = 0$. However, under our assumption that X is endogenous, the second condition can not hold and therefore we conclude an OLS estimation to be inconsistent. As a consequence of this inconsistency, we shall turn our attention to the 2-stage least square (2SLS) estimator. The 2SLS estimator is obtained, intuitively, by two steps of OLS:

By the same logic as Equation 2 we can derive the conditions under which $\hat{\beta}_{iv}$ is consistent:

$$
\begin{aligned}
\hat{\beta}_{\text{iv}} &= \frac{Cov(\hat{X}, Y)}{Var(\hat{X})} \\
&= \frac{Cov(\hat{X}, \beta\hat{X} + \varepsilon)}{Var(\hat{x})} \\
&= \frac{Cov(\hat{X}, \beta\hat{X})}{Var(\hat{X})} + \frac{Cov(\hat{X}, \varepsilon)}{Var(\hat{X})} \\
&= \beta + \frac{Cov(\hat{X}, \varepsilon)}{Var(\hat{X})} \\
\implies \text{plim } \hat{\beta}_{\text{iv}} &= \beta + \frac{\sigma_{\hat{X},\varepsilon}}{\sigma^2_{\hat{X}}}
\end{aligned} \tag{3}
$$

We then find that $\hat{\beta}_{\text{iv}}$ is a consistent estimator of β if and only if $\sigma^2_{\hat{X}} \neq 0$ and $\sigma_{\hat{X},\varepsilon} = 0$. To give some intuition behind these statements, firstly note that $\sigma_{\hat{X},\varepsilon} = 0$ implies that $\hat{\beta}_{\text{iv}}$ is consistent only when there is no association between $\boldsymbol{Z}$ and ε (i.e. $\boldsymbol{Z}$ affects Y only through X), and we can note that if there is such an association, the inconsistency of $\hat{\beta}_{\text{iv}}$ increases as the association increases. Secondly, $\sigma^2_{\hat{X}} \neq 0$ implies an association between X and $\boldsymbol{Z}$, and as this association decreases, the inconsistency of $\hat{\beta}_{\text{iv}}$ increases when there is some association between $\boldsymbol{Z}$ and ε.

Bound et al. (1995) takes these inferences a step further, and presents the relative inconsistency of the 2SLS estimator to the OLS estimator by subtracting β from both sides of the probability limits in Equations 2 and 3 and dividing as follows:

$$
\begin{aligned}
\frac{\text{plim } \hat{\beta}_{\text{iv}} - \beta}{\text{plim } \hat{\beta}_{\text{ols}} - \beta} &= \frac{\sigma_{\hat{X},\varepsilon}/\sigma^2_{\hat{X}}}{\sigma_{X,\varepsilon}/\sigma^2_X} \\
&= \frac{\sigma_{\hat{X},\varepsilon}/\sigma_{X,\varepsilon}}{\sigma^2_{\hat{X}}/\sigma^2_X} \\
&= \frac{\sigma_{\hat{X},\varepsilon}/\sigma_{X,\varepsilon}}{R^2_{X,\boldsymbol{Z}}}
\end{aligned} \tag{4}
$$

We notice that the denominator is the population R^2 value for a regression of X on $\boldsymbol{Z}$. Here we will informally define weak instruments as instrumental variables with low values of $R^2_{X,\boldsymbol{Z}}$ (described in Appendix A.2). Recalling that $\sigma_{X,\varepsilon}$ is necessarily non-zero as X is endogenous, we can infer that when there is an association between $\boldsymbol{Z}$ and ε such that $\sigma_{\hat{X},\varepsilon} \neq 0$, the effect of this term on the inconsistency of the IV estimator is increased when the instrument is weak. Of particular concern is the notion that we may find large inconsistencies from only very small correlations between the instrument and error if the instrument is weak enough, as the value of $R^2_{X,\boldsymbol{Z}}$ will approach 0, severely restricting the scope of problems in which weak instruments can yield consistent estimates.

2.2 Empirical comparison of OLS and IV estimation with instruments of varied strength on large samples

The results we have obtained so far are demonstrated to be correct by simulation. 20 million data points were simulated from two systems in order to approximate the asymptotic results under large data sets. The first system is described by the following structural equations:

$$
\begin{aligned}
Z &= \varepsilon_Z \\
A &= 0.1Z + \varepsilon_A \\
X &= 2A + 5Z + \varepsilon_X \\
Y &= 2X + 2A + \varepsilon_Y
\end{aligned} \tag{5}
$$

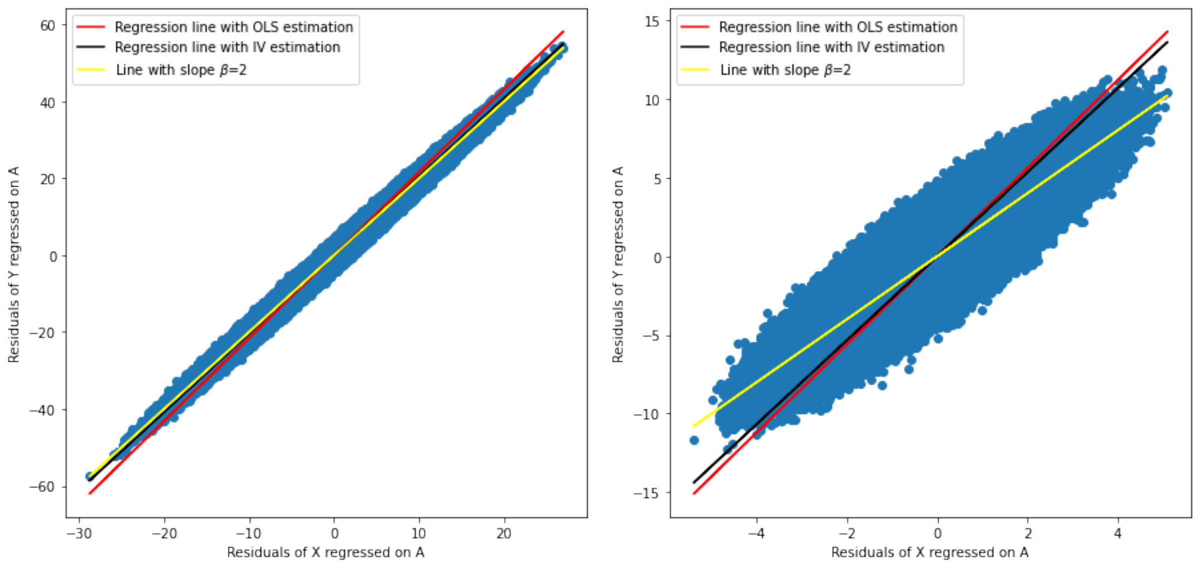


Figure 2: The results of OLS vs IV estimation on a large data set with a strong instrumental variable (left) and a weak instrumental variable (right).

Where each error term is given by a standard Gaussian. This system uses a single instrumental variable Z which has a weak association with the common exogenous variable A but a strong association with the endogenous variable X ($R^2_{X,Z} = 0.844$). Our parameter of interest is $\beta = 2$. Using the methods of OLS and 2SLS we obtained results of $\hat{\beta}_{\text{ols}} = 2.157$ and $\hat{\beta}_{\text{iv}} = 2.038$. Clearly we see that the IV estimator is more consistent than the OLS estimator.

The second system is given by similar set of structural equations to the first system, but with a much weaker association between the instrumental variable and the endogenous variable:

$$
\begin{aligned}
Z &= \varepsilon_Z \\
A &= 0.1Z + \varepsilon_A \\
X &= 2A + 0.1Z + \varepsilon_X \\
Y &= 2X + 2A + \varepsilon_Y
\end{aligned}
\tag{6}
$$

In this system the association between Z and X is much lower ($R^2_{X,Z} = 0.018$), and hence we consider Z a weak instrument. Again we find an inconsistent estimate with OLS of $\hat{\beta}_{\text{ols}} = 2.798$, but this time we see that the IV estimate is also largely inconsistent, with $\hat{\beta}_{\text{iv}} = 2.667$. This fits our analytical conclusion that when there is an association between the instrumental variables and the common exogenous variables, even if this association is very small, the effect on the inconsistency of the IV estimation is amplified by weaker instruments.

The results on both systems have been displayed with partial regression plots in Figure 2. It is clear that the instrument that has strong association with the endogenous treatment variable gives us a significantly more consistent IV estimate, as the regression line lies far closer to the line with slope of the true $\beta = 2$.

3 Finite-sample bias

3.1 Analytical results on the bias of estimators with legitimate instruments in finite samples

In this section we inherit the general system from Equation 1 and work under the assumption of $E[\varepsilon, \boldsymbol{Z}] = 0$, from which it follows that $\hat{\beta}_{\text{iv}}$ is consistent. Bound et al. assert that under the condition of consistency $\hat{\beta}_{\text{iv}}$ is nonetheless biased in the direction of $E[\hat{\beta}_{\text{ols}}]$, and that this bias decreases as sample size and $R^2_{X,\boldsymbol{Z}}$ increase.

The paper proposes several methods of deriving approximations or exact values (under certain assumptions) of the bias of the IV estimator relative to the OLS estimator. The mathematical derivations of most of these results is beyond the scope of the paper, but some important results are noteworthy. The notion of a 'concentration parameter' is introduced, defined as $\tau^2 \equiv \mathbf{\Pi}^T \mathbf{Z}^T \mathbf{Z} \mathbf{\Pi} / \sigma_\epsilon^2$ [3]. We find that the population analogy of the F-statistic (see Appendix A.3) on the instruments in the first-stage OLS estimation is τ^2/K, and that the bias of the IV estimator relative to the OLS estimator is approximately $1/(1+\tau^2/K)$ [4]. We can find an approximate formula for this relative bias by first finding the bias of the OLS estimator by subtracting β from both sides of the penultimate line of Equation 2 and taking expectations:

$$E[\hat{\beta}_{\text{ols}} - \beta] = \frac{\sigma_{X,\varepsilon}}{\sigma_X^2} \tag{7}$$

Then our approximation for the bias of the IV estimator follows clearly [5]:

$$E[\hat{\beta}_{\text{iv}} - \beta] \approx \frac{\sigma_{X,\varepsilon}}{\sigma_X^2} \frac{1}{1+\tau^2/K} \tag{8}$$

From this equation we see that $E[\hat{\beta}_{\text{iv}} - \beta] \to 0$ as $\tau^2/K \to \infty$ and $E[\hat{\beta}_{\text{iv}} - \beta] \to \sigma_{X,\varepsilon}/\sigma_X^2$ as $\tau^2/K \to 0$. Intuitively this result confirms that the bias of the IV estimator shrinks as the first stage becomes stronger, and implies higher bias for weaker instruments (where we are now defining the strength of instruments by the size of their F-statistic).

3.2 Empirical comparison of finite-sample bias for OLS and IV estimation with legitimate instruments

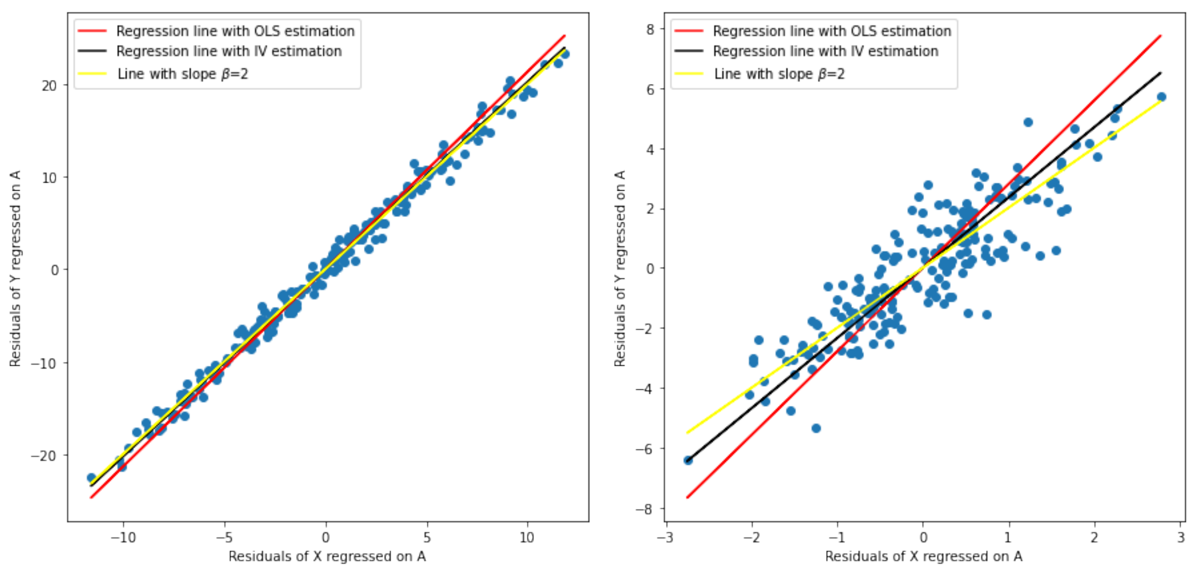


Figure 3: The results of OLS vs IV estimation on a small (finite) data set with a strong instrumental variable (left) and a weak instrumental variable (right).

Here we demonstrate how the bias of IV estimation varies as the association (and hence the F-statistic) between Z and X is changed. In these simulations only 200 data points were simulated so as to avoid the estimators tending to their probability limits. The first system is modified from Equation 5; the effect of Z on A is removed as follows:

$$
\begin{aligned}
Z &= \varepsilon_Z \\
A &= \varepsilon_A \\
X &= 2A + 5Z + \varepsilon_X \\
Y &= 2X + 2A + \varepsilon_Y
\end{aligned} \tag{9}
$$

Where all error terms are again standard Gaussian. Here the single instrumental variable Z is legitimate (as Y depends on Z only through X), and has a strong association with X (with an F-statistic of 1004.492 for regression of X on Z). Our parameter of interest is again $\beta = 2$. The calculated estimates were $\hat{\beta}_{\text{ols}} = 2.133$ and $\hat{\beta}_{\text{iv}} = 2.022$; the bias of a strong IV estimator is demonstrated to be very small even in small samples. Recalling that τ^2/K is analogous to the F statistic (though the F-statistic will tend to overestimate this value in any finite sample [1]), plugging the aforementioned F-statistic and $E[\hat{\beta}_{\text{ols}} - \beta] = 0.133$ into Equation 8 gives an approximate expected bias of $E[\hat{\beta}_{\text{iv}} - \beta] \approx 0.0001$, which fits our results assuming the F-statistic is overestimating.

For comparison, we modified the system from Equation 6 similarly, removing the effect of Z on A, resulting in the following structural equations:

$$
\begin{aligned}
Z &= \varepsilon_Z \\
A &= \varepsilon_A \\
X &= 2A + 0.1Z + \varepsilon_X \\
Y &= 2X + 2A + \varepsilon_Y
\end{aligned} \tag{10}
$$

In this system, Z, while still legitimate, is clearly a weak instrument, only giving an F-statistic of 1.060. The estimates on this system were $\hat{\beta}_{\text{ols}} = 2.787$ and $\hat{\beta}_{\text{iv}} = 2.341$. Analytically obtaining the expected bias as previously, we get $E[\hat{\beta}_{\text{iv}} - \beta] \approx 0.382$, which is well inline with the observed value of 0.341.

We have demonstrated through simulation the correctness of our assertion that the finite-sample bias of an IV estimator relative to the OLS estimator is inversely proportional to the strength of the association between the instrumental variables and the endogenous explanatory variable. Importantly, this means that weak instruments suffer larger finite-sample bias than good ones.

The results of the two simulations are presented with partial regression plots in Figure 3. For the IV estimate using the strong instrument, the regression line closely fits the line with slope of $\beta = 2$. In contrast, it is clear that the IV estimate using the weak instrument has a bias in the direction of $\hat{\beta}_{\text{ols}}$.

4 Case study: Angrist and Krueger (1991)

We shall here discuss Angrist and Krueger (1991) which uses quarter (of the year) of birth as an instrument for educational attainment to estimate wage equations. They used data from the US census on men who were born in the 1930s and 1940s [2]. Their assertion is that due to compulsory schooling laws requiring children to begin school in the calendar year that they turn six and remain in education until age sixteen, those born earlier in the year are required to begin school at an older age and are allowed to leave at an earlier point in their education, hence creating an association between quarter of birth and educational attainment. The argument follows then that since there is a confounded relationship between educational attainment and income (it is not difficult to think that some other variables such as race and location of residence may have an effect on both of these variables), the true causal effect of educational attainment on income can be recovered using an IV estimator with quarter of birth as the instrument. In this work Angrist and Krueger provided a feasible mechanism and compelling evidence that there was in fact a small effect of quarter of birth on men's educational attainment, and hence on income. However, despite their use of very large samples, we have demonstrated throughout this paper that IV estimates are likely to be inconsistent, and even suffer finite-sample bias when they are based on weak instruments. Bound et al. (1995) provides a strong argument against the validity of the conclusions that Angrist and Krueger made.

4.1 Inconsistency: the amplifying effect of a weak association between quarter of birth and educational attainments on inconsistency of IV estimators

Bound et al. first argue against the credibility of consistency in the IV estimator with quarter of birth as the instrument. As we deduced in Equation 4, any small association between the instrument and the common exogenous variables will result in large inconsistencies in the IV estimator, relative to the OLS estimator, when the instrument is weak. By the admission of Angrist and Krueger (1991) the association between quarter of birth and educational attainment is a small one; it is intuitively difficult to imagine any other case. The paper in fact reported that there was only an average of 0.1 decrease in years of educational attainment for men born in the first quarter of the year compared to those who weren't, for the years 1930-1939. As it is clear that quarter of birth is a weak instrument, we turn our attention to possible other exogenous variables on wages that may be affected by quarter of birth, to ascertain the legitimacy of the estimator (i.e. does income depend on quarter of birth only through educational attainment?).

Although it is not obvious how quarter of birth affects wages outside of educational attainment, Table V in Angrist and Krueger (1991) reports IV estimates with and without a number of control variables; namely race, urban status, marital status and region of residence. Bound et al. notes interestingly that when these controls were included, compared to when they were not, there was a 21% decrease in the IV estimate compared to only an 11% decrease in the OLS estimate. This significant decrease for the IV estimate, where fitted values of educational attainment from a regression on quarter of birth are being used in the second stage regression, implies an association between quarter of birth and the set of control variables. The example is given that black men are slightly more likely to be born in the first quarter of the year than the other quarters when compared to white men, and that black men also, on average, have lower income than white men. It appears that there are associations between quarter of birth and income that are not transmitted through educational attainment, making the instrument illegitimate.

Furthermore, Bound et al. demonstrate that an extremely large proportion of association between quarter of birth and income is a result of family income at the time of birth. They used the census data to calculate that mean log per capita family income among children aged 0-3 years old decreases by 0.0238 between the first quarter of the year and the other three quarters considered collectively. Based on a intergenerational correlation in income on the order of 0.4 [6], multiplying this value with 0.0238 gives a reasonable expectation of a 0.95% decrease in wages between men born in the first quarter of the year and those born in the other three quarters. Bound et al. note that the census data shows only a 1.1% decrease in wages between men born in the first quarter of the year and those born in the other three quarters; we find that family income at birth very plausibly accounts for a large amount of the association between quarter of birth and income.

As we have demonstrated that there are variables other than educational attainment which likely account for some of the association between quarter of birth and wages, and that quarter of birth is only weakly associated with educational attainment, then our results from Section 2 strongly imply that the IV estimates found by Angrist and Krueger (1991) are likely to be inconsistent.

4.2 Finite-sample bias: the risk of bias in IV estimates with weak instruments, even on large samples

Finally Bound et al. claim that using very large samples does not necessarily insulate an experiment from the risk of finite-sample bias when using weak instruments; they assert that the IV estimates obtained in Angrist and Krueger (1991) may be subject to such bias. Furthermore, they demonstrate that adding interactions as excluded variables, as they mention was common practice and was used by Angrist and Krueger, may increase this bias further.

Bound et al. reexamined the sample of men born between 1930 and 1939. We shall compare the results in the first two and last two columns of Table 1 [1]. The first two columns show the OLS and IV estimates and standard errors, as well as first stage F and partial R^2 statistics on the instrumental variables obtained when using age and age squared (in quarters of years) as control variables and quarter of birth as an instrumental variable. The last two columns show the same statistics but with the interaction between quarter of birth and year of birth as instrumental variables. The partial R^2 value increases by only 0.02 with 28 instrumental variables compared to only 3, causing a large drop in the value of the F statistic from 13.486 to 1.613. This small F statistic implies a strong likelihood of finite-sample bias in the IV estimate. There is also a minor drop in standard error from 0.033 to 0.029 implying a slight increase in precision when more weak instruments are used.

It is further stated that Angrist and Krueger (1991) used interactions between state of birth and quarter of birth to improve their IV estimates as compulsory schooling laws vary between states. It was found that by including these interactions, the standard error on the IV estimate dropped from 0.029 to 0.011; the precision of the estimates increased. However, the F statistic only increased from 1.613 before introducing these interactions to 1.869 when they were used. This increase in F statistic is clearly not large enough to reduce the possibility of finite-sample bias.

From these results, it is firstly clear that it is possible that the IV estimates obtained by Angrist and Krueger (1991) may be affected by finite-sample bias, despite the use of large samples, as a result of the use of weak instrumental variables. Secondly, increasing the number of weak instruments seems to improve the precision of IV estimates, but at the cost of an increase in bias. Importantly, we learn that large samples can not necessarily compensate for the effect of weak instruments, and recognise that this places a large importance on the need to validate the legitimacy of instrumental variables if useful inference about the effect of an endogenous treatment variable is to be drawn using them.

5 Discussions

In this paper we have discussed the limitations of instrumental variables for the estimation of causal effects. We find that it is essential that instrumental variables be strongly associated with the endogenous variable if they are to make accurate estimates, and in fact it seems that this detail has been overlooked in popular papers such as Angrist and Krueger (1991). Strikingly, we have shown that when the instrumental variables are even slightly illegitimate (i.e. Y depends on $\boldsymbol{Z}$ NOT only through X), the estimator based on these instrumental variables can still be largely inconsistent if the first stage regression is particularly weak. Clearly the strict constraints we have described on the appropriate usages of instrumental variable estimation imply that the range of these possible uses is far more limited than was originally believed; a situation in which an instrumental variable is truly legitimate and has a strong association with the endogenous variable seems unlikely to arise in the majority of real world cases, and in many cases such an instrument may not exist at all.

In theory we initially believed that instrumental variable estimation could be used to yield impressive results over ordinary least squares estimates. In the correct circumstances they appear quite capable of recovering the true causal effect between two variables despite the effect of confounding variables, and based on the empirical simulations we ran, it seems they should be favoured over OLS estimation in these cases (Figure 2 and Figure 3). However, as we have found that the conditions under which they perform well are so strict, we do not believe they can not be relied on as a standard tool to find the strength of causal effects. Although IV estimation did outperform OLS estimation in the simulations using weak instruments, they still showed significant inconsistency in the first experiment (Figure 2) and bias in the second (Figure 3). From the example of Angrist and Krueger (1991), we also find that the flaws of weak instruments can not be easily remedied by using large samples or increasing the number of weak instruments that are used; though these methods do have merit (for example, using more instruments seemed to yield greater precision), they were not sufficient fixes to the issue of bias in IV estimation. Rather than accept the "less" inconsistent or biased of the possible estimators, it would seem more desirable to continue the search for more appropriate methods of causal analysis in these cases.

Appendix A Statistical concepts

Here we give some brief background for the statistical terms introduced throughout the paper.

A.1 Projection matrix

The projection matrix is the matrix that maps the vector of observed responses to the vector of fitted values. It has many important properties, but two are of particular note. Let $\mathbf{P_Z}$ be the projection matrix. Then $\mathbf{P_Z}$ is:

1. Symmetric.
2. Idempotent: $\mathbf{P}_{\mathbf{Z}}^2 = \mathbf{P}_{\mathbf{Z}}$.

A.2 Coefficient of determination (R^2)

We use the coefficient of determination as a measure of explained variance in the dependent variable by the independent variables in a regression model. The value lies in $[0, 1]$ and can be interpreted as the proportion of variance that is explained by the model. To formalise this idea, we consider a sample of an independent variable $y_1, y_2, ..., y_n$ with sample mean $\bar{y}$, and corresponding fitted values $f_1(\boldsymbol{x}_1), f_2(\boldsymbol{x}_2), ..., f_n(\boldsymbol{x}_n)$ with independent variables $\boldsymbol{x}$. We define the residual sum of squares as $S_r = \sum_{i=1}^n (y_i - f_i(\boldsymbol{x}_i)^2)$ and the total sum of squares as $S_t = \sum_{i=1}^n (y_i - \bar{y})^2$. Then $R^2_{\boldsymbol{x},y} = S_r/S_t$.

A.3 F-statistic

The F-statistic is a statistic for the joint significance of independent variables in explaining the variance in the dependent variable. We can consider an unrestricted model in which we regress the dependent variable on our instrumental variables, and a restricted model in which our instrumental variables are omitted. For the unrestricted model we can denote the residual sum of squares S_u, and for the restricted model we use S_r. Assuming n observations, p_u variables in the unrestricted model and p_r variables in the restricted model, we obtain $F = \frac{(S_r/S_u)/(p_u - p_r)}{S_u/(n-p_u)}$. The greater the joint significance of the instrumental variables for explaining the variance in the dependent variable, the larger the value of F will be.

Appendix B Code

The code used to run the simulations in Sections 2.2 and 3.2 is given here for reproducibility.

In [ ]:

```python
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.feature_selection import f_regression
import matplotlib.pyplot as plt
```

2.2 Empirical comparison of OLS and IV estimation with instruments of varied strength on large samples

Good instrument

In [ ]:

```python
def simulate_good_instrument(n):
    # Simulate an instrument Z with a weak association with the common exogenous variable A but with a strong ass
    # the endogenous variable X.
    Z = np.random.normal(0, 1, n)
    A = np.random.normal(0, 1, n) + 0.1 * Z
    X = np.random.normal(0, 1, n) + 5 * Z + 2*A
    Y = np.random.normal(0, 1, n) + 2 * X + 2*A
    return(Z.reshape(-1,1), A.reshape(-1,1), X.reshape(-1,1), Y.reshape(-1,1))

(Z1, A1, X1, Y1) = simulate_good_instrument(20000000) # Large n to demonstrate asymptotic behaviour
```

In [ ]:

```python
lr1 = LinearRegression() # OLS regression
lr1.fit(X1, Y1)
b_ols_str = lr1.coef_[0][0] # OLS estimate for a strong instrument
print("OLS estimate: ", b_ols_str)
```

```
OLS estimate:  2.157278679956988
```

In [ ]:

```python
lr2 = LinearRegression() # First stage OLS regression
lr2.fit(Z1, X1)
X_hat1 = lr2.predict(Z1)
lr3 = LinearRegression() # Second stage OLS regression
lr3.fit(X_hat1, Y1)
R2_str = np.var(X_hat1) / np.var(X1)
b_iv_str = lr3.coef_[0][0] # IV estimate for a strong instrument
print("R^2: ", R2_str)
print("IV estimate: ", b_iv_str)
```

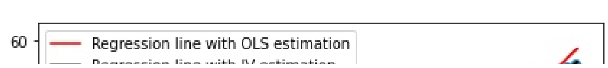
```
R^2:  0.8438520137652634
IV estimate:  2.038354958544784
```

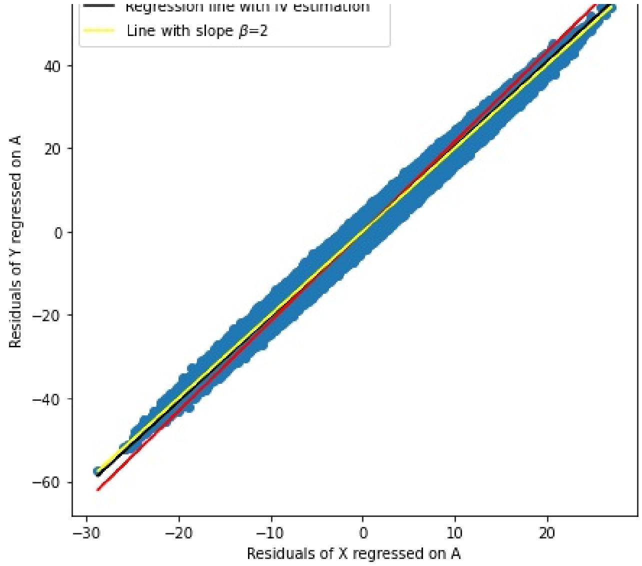
In [ ]:

```python
# Residuals from regression of Y against A
Y_lr = LinearRegression()
Y_lr.fit(A1, Y1)
Y_res = Y1 - Y_lr.predict(A1)

# Residuals from regression of X against A
X_lr = LinearRegression()
X_lr.fit(A1, X1)
X_res = X1 - X_lr.predict(A1)

# Plot the Y residuals against X residuals
fig, ax = plt.subplots(1,1, figsize=(7,7))
ax.scatter(X_res, Y_res)

# Plot the regression line obtained with OLS estimation
ax.plot(X_res, b_ols_str * X_res, color='r', label='Regression line with OLS estimation')

# Plot the regression line obtained with IV estimation
ax.plot(X_res, b_iv_str * X_res, color='k', label='Regression line with IV estimation')

# Plot the correct regression line
ax.plot(X_res, 2 * X_res, color='yellow', label=r'Line with slope $\beta$=2')


# Axes and legend
ax.set_xlabel('Residuals of X regressed on A')
ax.set_ylabel('Residuals of Y regressed on A')
ax.legend()
```


Out[ ]:

```
<matplotlib.legend.Legend at 0x1b6a6aebaf0>
```





Weak instrument

We now simulate the same system but with a much weaker association between Z and X, to demonstrate that IV estimation suffers significant inconsistency when using weak instruments.

In [ ]:

```
def simulate_weak_instrument(n):
    # Simulate an instrument Z with a weak association with the common exogenous variable A but with a strong ass
    # the endogenous variable X.
    Z = np.random.normal(0, 1, n)
    A = np.random.normal(0, 1, n) + 0.1 * Z
    X = np.random.normal(0, 1, n) + 0.1 * Z + 2*A
    Y = np.random.normal(0, 1, n) + 2 * X + 2*A
    return(Z.reshape(-1,1), A.reshape(-1,1), X.reshape(-1,1), Y.reshape(-1,1))

(Z2, A2, X2, Y2) = simulate_weak_instrument(20000000) # Large n to demonstrate asymptotic behaviour
```

In [ ]:

```
lr4 = LinearRegression() # OLS regression
lr4.fit(X2, Y2)
b_ols_wk = lr4.coef_[0][0]
print('OLS estimate', b_ols_wk)
```

```
OLS estimate 2.7979488582141525
```

In [ ]:

```
lr5 = LinearRegression() # First stage OLS regression
lr5.fit(Z2.reshape(-1, 1), X2)
X_hat2 = lr5.predict(Z2)
lr6 = LinearRegression() # Second stage OLS regression
lr6.fit(X_hat2, Y2)
R2_wk = np.var(X_hat2) / np.var(X2)
b_iv_wk = lr6.coef_[0][0] # IV estimate for a strong instrument
print("R^2: ", R2_wk)
print("IV estimate: ", b_iv_wk)
```

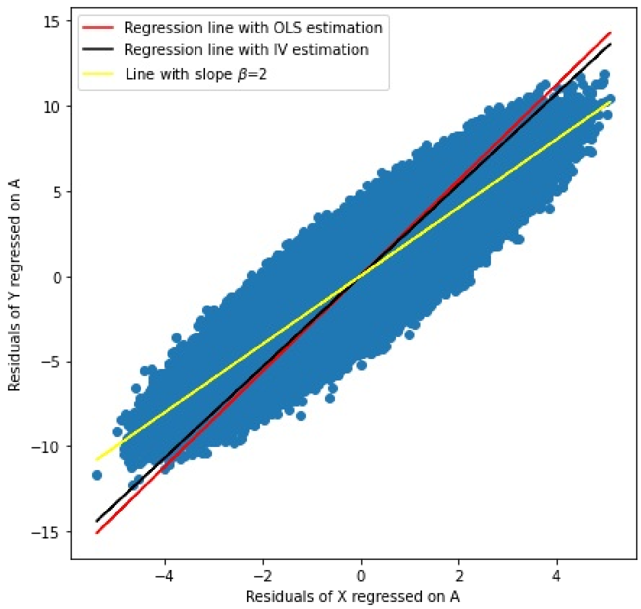
```
R^2:  0.017534897188273856
IV estimate:  2.6671610513261856
```

In [ ]:

```
# Residuals from regression of Y against A
Y_lr = LinearRegression()
Y_lr.fit(A2, Y2)
Y_res = Y2 - Y_lr.predict(A2)

# Residuals from regression of X against A
X_lr = LinearRegression()
X_lr.fit(A2, X2)
X_res = X2 - X_lr.predict(A2)

# Plot the Y residuals against X residuals
fig, ax = plt.subplots(1,1, figsize=(7,7))
ax.scatter(X_res, Y_res)
```

```
# Plot the regression line obtained with OLS estimation
ax.plot(X_res, b_ols_wk * X_res, color='r', label='Regression line with OLS estimation')

# Plot the regression line obtained with IV estimation
ax.plot(X_res, b_iv_wk * X_res, color='k', label='Regression line with IV estimation')

# Plot the correct regression line
ax.plot(X_res, 2 * X_res, color='yellow', label=r'Line with slope $\beta$=2')

# Axes and legend
ax.set_xlabel('Residuals of X regressed on A')
ax.set_ylabel('Residuals of Y regressed on A')
ax.legend()
```

Out[ ]: <matplotlib.legend.Legend at 0x1b6a6aeb0d0>



3.2 Empirical comparison of finite-sample bias for OLS and IV estimation with legitimate instruments

Good instrument

In [ ]:
```
def simulate_good_instrument2(n):
    # Simulate an instrument Z with no association with the common exogenous variable A and with a strong associa
    # the endogenous variable X.
    Z = np.random.normal(0, 1, n)
    A = np.random.normal(0, 1, n)
    X = np.random.normal(0, 1, n) + 5 * Z + 2*A
    Y = np.random.normal(0, 1, n) + 2 * X + 2*A
    return(Z.reshape(-1,1), A.reshape(-1,1), X.reshape(-1,1), Y.reshape(-1,1))
```

In [ ]:
```
# Simulate the data
np.random.seed(0)
(Z3, A3, X3, Y3) = simulate_good_instrument2(200) # Simulate only a small data set (200)
```

In [ ]:
```
lr7 = LinearRegression() # OLS regression
lr7.fit(X3, Y3)
b_ols_str_fin = lr7.coef_[0][0] # OLS estimate for a strong instrument
print("OLS estimate: ", b_ols_str_fin)
```

```
OLS estimate:  2.133425189970813
```

In [ ]:
```
lr8 = LinearRegression() # First stage OLS regression
lr8.fit(Z3, X3)
X_hat3 = lr8.predict(Z3)
lr9 = LinearRegression() # Second stage OLS regression
lr9.fit(X_hat3, Y3)
f_str,_ = f_regression(X3, Z3.ravel()) # F-value for the strong instrument
```

```python
b_iv_str_fin = lr9.coef_[0][0] # IV estimate for a strong instrument on finite sample
print("F statistic: ", f_str[0])
print("IV estimate: ", b_iv_str_fin)
```

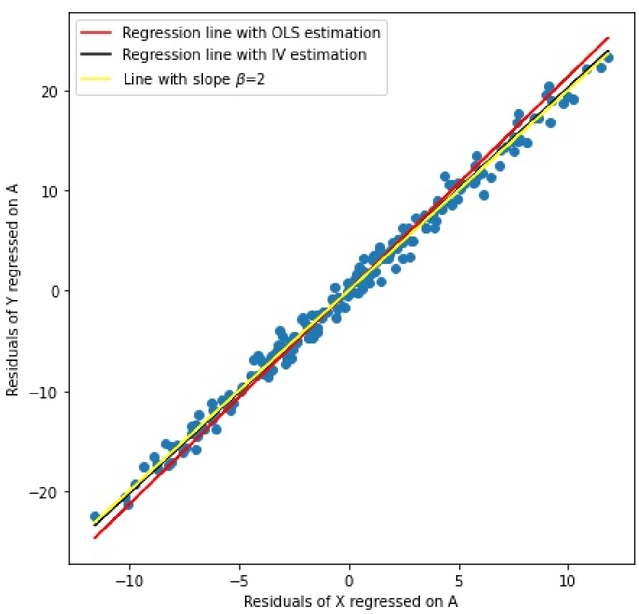
```
F statistic:  1004.4924228872461
IV estimate:  2.0226373162117097
```

In [ ]:

```python
# Residuals from regression of Y against A
Y_lr = LinearRegression()
Y_lr.fit(A3, Y3)
Y_res = Y3 - Y_lr.predict(A3)

# Residuals from regression of X against A
X_lr = LinearRegression()
X_lr.fit(A3, X3)
X_res = X3 - X_lr.predict(A3)

# Plot the Y residuals against X residuals
fig, ax = plt.subplots(1,1, figsize=(7,7))
ax.scatter(X_res, Y_res)

# Plot the regression line obtained with OLS estimation
ax.plot(X_res, b_ols_str_fin * X_res, color='r', label='Regression line with OLS estimation')

# Plot the regression line obtained with IV estimation
ax.plot(X_res, b_iv_str_fin * X_res, color='k', label='Regression line with IV estimation')

# Plot the correct regression line
ax.plot(X_res, 2 * X_res, color='yellow', label=r'Line with slope $\beta$=2')

# Axes and legend
ax.set_xlabel('Residuals of X regressed on A')
ax.set_ylabel('Residuals of Y regressed on A')
ax.legend()
```

Out[ ]:

```
<matplotlib.legend.Legend at 0x1b6a6996f70>
```



Weak instrument

In [ ]:

```python
def simulate_weak_instrument2(n):
    # Simulate an instrument Z with no association with the common exogenous variable A but a weak association wi
    # the endogenous variable X.
    Z = np.random.normal(0, 1, n)
    A = np.random.normal(0, 1, n)
    X = np.random.normal(0, 1, n) + 0.05 * Z + 2*A
    Y = np.random.normal(0, 1, n) + 2 * X + 2*A
    return(Z.reshape(-1,1), A.reshape(-1,1), X.reshape(-1,1), Y.reshape(-1,1))
```


In [ ]:

```python
# Simulate the data
(Z4, A4, X4, Y4) = simulate_weak_instrument2(200) # Simulate only a small data set (200)
```

In [ ]:

```
lr10 = LinearRegression() # OLS regression
lr10.fit(X4, Y4)
b_ols_wk_fin = lr10.coef_[0][0] # OLS estimate for a strong instrument with a finite sample
print("OLS estimate: ", b_ols_wk_fin)
```

```
OLS estimate:  2.787320391730289
```

In [ ]:

```
lr11 = LinearRegression() # First stage OLS regression
lr11.fit(Z4, X4)
X_hat4 = lr11.predict(Z4)
lr12 = LinearRegression() # Second stage OLS regression
lr12.fit(X_hat4, Y4)
f_wk,_ = f_regression(X4, Z4.ravel()) # F-value for the weak instrument
b_iv_wk_fin = lr12.coef_[0][0] # IV estimate with the weak instrument
print("F statistic: ", f_wk[0])
print("IV estimate: ", b_iv_wk_fin)
```

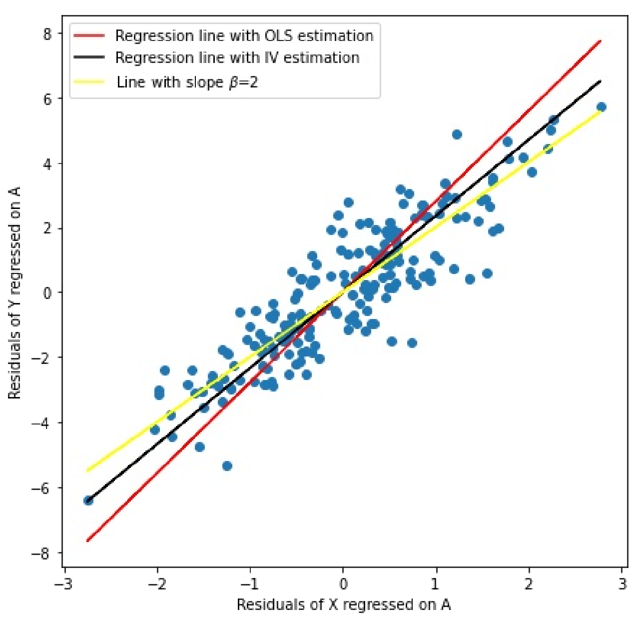
```
F statistic:  1.0603225918475694
IV estimate:  2.3413489359001467
```

In [ ]:

```
# Residuals from regression of Y against A
Y_lr = LinearRegression()
Y_lr.fit(A4, Y4)
Y_res = Y4 - Y_lr.predict(A4)

# Residuals from regression of X against A
X_lr = LinearRegression()
X_lr.fit(A4, X4)
X_res = X4 - X_lr.predict(A4)

# Plot the Y residuals against X residuals
fig, ax = plt.subplots(1,1, figsize=(7,7))
ax.scatter(X_res, Y_res)

# Plot the regression line obtained with OLS estimation
ax.plot(X_res, b_ols_wk_fin * X_res, color='r', label='Regression line with OLS estimation')

# Plot the regression line obtained with IV estimation
ax.plot(X_res, b_iv_wk_fin * X_res, color='k', label='Regression line with IV estimation')

# Plot the correct regression line
ax.plot(X_res, 2 * X_res, color='yellow', label=r'Line with slope $\beta$=2')

# Axes and legend
ax.set_xlabel('Residuals of X regressed on A')
ax.set_ylabel('Residuals of Y regressed on A')
ax.legend()
```

Out[ ]:

```
<matplotlib.legend.Legend at 0x1b6a691b460>
```



References

[1] J. Bound, D. A. Jaeger, and R. M. Baker, "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 443–450, 1995. [Online]. Available: https://doi.org/10.1080/01621459.1995.10476536

[2] J. D. Angrist and A. B. Krueger, "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, vol. 106, no. 4, pp. 979–1014, 1991. [Online]. Available: http://www.jstor.org/stable/2937954

[3] R. L. Basmann, "Remarks concerning the application of exact finite sample distribution functions of gcl estimators in econometric statistical inference," *Journal of the American Statistical Association*, vol. 58, no. 304, pp. 943–976, 1963. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10480680

[4] D. Staiger and J. H. Stock, "Instrumental variables regression with weak instruments," National Bureau of Economic Research, Working Paper 151, January 1994. [Online]. Available: http://www.nber.org/papers/t0151

[5] S. Cunningham, *Causal Inference: The Mixtape.* Yale University Press, 2021. [Online]. Available: http://www.jstor.org/stable/j.ctv1c29t27

[6] D. J. Zimmerman, "Regression toward mediocrity in economic stature," *The American Economic Review*, vol. 82, no. 3, pp. 409–429, 1992. [Online]. Available: http://www.jstor.org/stable/2117313