Instrumental Variable Estimation with Imperfect Instruments

January 2018

1 Introduction

The analysis of causal effects, rather than simple forecasting, is widespread in the social sciences and used from policy evaluation to financial economics. A common issue when trying to isolate these effects is that the independent variables are endogenous to the regression equation which leads in general to biased regression results. Through an instrumental variable approach, researchers are nevertheless able to consistently estimate the causal effect of the variable of interest, given that a "good" instrument is available. A good instrument should fulfill the relevance and exogeneity conditions, that is, it should be correlated with the endogenous variable but not with the error term of the regression equation. The relevance condition is easily verifiable with data and many theoretic and applied papers investigate the effect of weak instruments. Exogeneity, however, is in many cases an untestable, but crucial assumption for IV estimators to yield consistent results. If the exogeneity assumption does not hold and we have weak instruments the IV estimator might even exacerbate the bias of OLS results. Tests for overidentification exist, for regressions where multiple instruments are available, but are not able to confirm the exogeneity of the instruments, only to reject non-exogeneity. Hence, researchers have to rely on economic theory and intuition if they judge on the validity of an instrument. Nevo and Rosen 2012 offer a new tool that deals with cases where one can cast doubt on the exogeneity assumption. They propose a method where the complete exogeneity assumption is replaced by two different hypotheses: (i) The correlation between the instrument and the error term has the same sign as the correlation between the endogenous regressor and the error term; (ii) The instrument is less correlated with the error term than the endogenous regressor. While these hypotheses are also not verifiable with data, they seem less restrictive than imposing complete exogeneity of the instrument. Under these assumptions, the method of Nevo and Rosen 2012 is able to provide valid estimates at the price of having only interval -, and not the usual point identified estimators.

In section 2 we will present the setup and the identification strategy of the authors. Unlike the original paper we concentrate on the linear case only since we only apply the researchers' method in linear IV models and to avoid simply rephrasing the original paper. Thereafter, in section 3 we derive the estimators of the bonds for the simple linear model with one endogenous regressor and one imperfect instrument, and extend the model to additional regressors (section 4) and multiple imperfect instruments (section 5). The derived bounds can be used to asses what would happen if the exogeneity assumption fails. Using these bounds, we will investigate in section 6 how this new method changes the results of estimating the returns to education, replicating one of the most renowned papers in this area by Card 1993.

2 General Setup

2.1 Assumptions

Nevo and Rosen 2012 derive the results for the general case, however since the researchers only use the linear model we also assume the model of the form

$$Y = X\beta + W\delta + U \text{ with } \mathbf{E}[U] = 0.$$

where Y is the outcome variable, X is at least one endogenous covariate, W are additional covariates that are either exogenous or can be consistently estimated by IV, and U is an unobserved additively separable mean zero error. There exist Z^w valid instruments which may include exogenous regressors of W.

The researchers want to inspect the case when point identification with GMM is not possible, because in this case additional imperfect instruments could not improve the estimation. Therefore, they restrict the analysis to the case when the number of valid instruments Z^w is equal to the number of covariates of W. Lastly, there are also imperfect instruments Z with dimension $n \times k$.

In the following, the authors will make use of the the following five assumptions. They are all introduced at once, but A5 is only needed for linear models.

Assumption A1 (random sampling): $(y_i, x_i, w_i, z_i, z_i^w, \epsilon_i)', i = 1, ..., n$

A random sample (Y, X, W, Z, Z^w) is is drawn from population $(\Omega, \mathscr{F}, \mathbb{P})$ where subscript i denotes observations and subscript j denotes individual elements of these vectors.

Assumption A2 (Z^w exogenous, X endogenous): $\mathbf{E}(Z^{w'}U) = 0, \mathbf{E}(XU) \neq 0$

Assumption A3 (same direction of correlation of X and U and Z and U)

$$\rho_{xu}\rho_{z_ju} \geq 0, j = 1, ...k_z$$

Assumption A4 (instruments Z "less endogenous" than X)

$$|\rho_{xu}| \geq |\rho_{z_ju}|, j = 1, ...k_z$$

Assumption A5 (rank and order): $\text{rank}(\mathbf{E}[(Z, Z^w)'(Z, Z^w)] = k_z + k_w$, $\text{rank}(\mathbf{E}[(X, Z^w)'(X, Z^w)] = k_w + 1$, and $\text{rank}(\mathbf{E}[(Z, Z^w)'(X, W)]) = k_w + 1, k_z \geq 1$

It it important to note that A3 and A4, serve to replace the stronger exogeneity assumption $\rho_{z_ju} = 0$.

2.2 Identification

In the following we show how Nevo and Rosen 2012 construct a weighted average of Z_j and X which is uncorrelated with the error term U and obtain a new estimator. This new estimator together with the IV estimator and OLS estimator will be useful for establishing bounds with regards to the value of the true parameter β.

To begin, the researchers define λ_j^* as follows:

$$\lambda_j^* = \rho_{z_ju}/\rho_{xu}$$

Here, λ_j^* expresses the relative correlation of the regressor x with the error term u and the correlation of instrument Z_j with the error term u. This identity is always unknown. However, using λ_j^* it is possible to define the function

$$V_j(\lambda) \equiv \sigma_x Z_j - \lambda_j \sigma_{z_j} X$$

. The function $V_j(\cdot)$ is the weighted average of Z_j and X which is by definition uncorrelated with the error term U.

λ_j^* and $V_j(\lambda)$ have two important features:

First, under A3 and A4 it is obvious that λ_j^* cannot be bigger than 1 or smaller than 0.

$$\lambda_j^* \in [0, 1], j = 1, ..., k_z \tag{2.1}$$

Second, $V_j(\lambda)$ satisfies the moment condition

$$\mathbf{E}[(Y - (X\beta + W\delta)) \cdot V_j(\lambda_j^*)] = 0 \tag{2.2}$$

A third restriction, which formalizes that valid instruments are exogenous, is needed to derive the identification results:

$$\mathbf{E}[Z_j^{w'}(Y - (X\beta + W\delta)) \cdot V_j(\lambda_j^*)] = 0 \tag{2.3}$$

The decisive idea of the paper is to use the fact that λ_j^* is bounded between 0 and 1 and derive the estimator assuming λ_j^* takes the maximum value of 1. Hence, by including λ which captures the correlation between the u and X and Z and by setting λ at its maximum value 1, the authors have been able to establish an additional moment condition, which can be used for estimation. Nevo and Rosen 2012 derive with this idea Proposition 1 and Corollary 1 which state that even in the general model there is a convex solution. Since we only concentrate on the linear case and the explicit expressions for the linear case will be derived in the following, we do not state Proposition 1 and Corollary 1 explicitly.

3 The Simple Linear Model with One Imperfect Instrument

The subsequent analysis applies for the case with one endogenous regressor X and one imperfect instrument Z. Additional exogenous regressors will be added in section 4 and the case with multiple imperfect instruments will be investigated in section 5. In a first step, in section 3.1 using A1-A3 a simple characterization of the bounds will be derived. Then, in section 3.2 the bounds will be improved by adding A4.

3.1 Characterizations of bounds

To begin, the researchers define β^{OLS} and β^{IV} as probability limits of standard OLS and IV for β. Hence,

$$\beta^{OLS} = \frac{\sigma_{xy}}{\sigma_x^2} = \beta + \frac{\sigma_{xu}}{\sigma_x^2} \tag{3.1}$$

$$\beta_z^{IV} = \frac{\sigma_{xy}}{\sigma_{xz}} = \beta + \frac{\sigma_{zu}}{\sigma_{xz}} \tag{3.2}$$

and their estimates are $\hat{\beta}^{OLS}$ and $\hat{\beta}_z^{IV}$, respectively. It can easily be seen that:

Lemma 1: If $\sigma_{xz} < 0$ then β lies between β^{OLS} and β_z^{IV}. β^{OLS} is the upper bound if $\sigma_{xu} > 0$, while β^{OLS} is the lower bound if $\sigma_{xu} < 0$. If instead $\sigma_{xz} \geq 0$, then if $\sigma_{xu} > 0$, $\beta \leq \min(\beta^{OLS}, \beta_z^{IV})$ while if $\sigma_{xu} < 0, \beta \geq \max(\beta^{OLS}, \beta_z^{IV})$

Let A1-A3 hold. If $\sigma_{xz} < 0$

$$B_{\text{L1}}^* = \begin{cases} \left[\beta_z^{IV}, \beta^{OLS}\right] & \text{if } \sigma_{xu} > 0 \\ \left[\beta^{OLS}, \beta_z^{IV}\right] & \text{if } \sigma_{xu} < 0 \end{cases}$$

If $\sigma_{xz} \geq 0$

$$B_{\text{L1}}^* = \begin{cases} \left(-\infty, \min[\beta^{OLS}, \beta_z^{IV}]\right] & \text{if } \sigma_{xu} > 0 \\ \left[\max[\beta^{OLS}, \beta_z^{IV}], -\infty\right) & \text{if } \sigma_{xu} < 0 \end{cases}$$

So Lemma 1 offers a finite two sided bound if $\sigma_{xz} < 0$ and that the bound is one-sided if $\sigma_{xz} > 0$. Especially, if $\sigma_{xz} < 0$ holds, the results might be economically helpful because the estimation offers bounds on both sides. Without implying 4, it is even possible to determine the direction of the estimator when $\sigma_{xu} < 0$ which is shown in the following and will ultimately result in Lemma 2.

First, subtracting equation 3.2 from 3.1 yields

$$\beta^{OLS} - \beta_z^{IV} = \frac{\sigma_{xu}}{\sigma_x^2} - \frac{\sigma_{zu}}{\sigma_{xz}} \tag{3.3}$$

Also

$$\beta^{OLS} - \beta = \frac{\sigma_{xu}}{\sigma_x^2} \tag{3.4}$$

holds. It shall be shown that $\text{sgn}(\beta^{OLS} - \beta_z^{IV}) = \text{sgn}(\beta^{OLS} - \beta)$. Reformulating this gives $(\beta^{OLS} - \beta_z^{IV})(\beta^{OLS} - \beta)$ and plugging in equation 3.3 as well as 3.4 yields

$$\frac{\sigma_{xu}}{\sigma_x^2}\left(\frac{\sigma_{xu}\sigma_{xz} - \sigma_x^2\sigma_{zu}}{\sigma_x^2\sigma_{xz}}\right) > 0 \tag{3.5}$$

Here, a case distinction is necessary.

Case 1: $\sigma_{xz} < 0$. A3 states that σ_{xu} and σ_{zu} have the same sign. Since σ_x^2 is always positive, and σ_{xz} is negative in case 1, $\sigma_{xu}\sigma_{xz} - \sigma_x^2\sigma_{zu}$ will always have the opposite sign as σ_{xu} (and σ_{zu}). Therefore, $\sigma_{xu}\sigma_{xz} - \sigma_x^2\sigma_{zu}/\sigma_{xz}$ has the same sign and plugging this observation in equation 3.5 proves that $(\beta^{OLS} - \beta_z^{IV})(\beta^{OLS} - \beta) > 0$.

Case 2: $\sigma_{xz} > 0$ also needs the assumption $\lambda^* < \rho_{xz}$. Transforming equation 2.1 yields $\sigma_{zu} = \lambda^*\sigma_{xu}\sigma_z/\sigma_x$. Plugging this into equation 3.3 gives

$$\beta^{OLS} - \beta_z^{IV} = \frac{\sigma_{xu}}{\sigma_x}\left(\frac{1}{\sigma_x} - \frac{\sigma_z\lambda^*}{\sigma_{xz}}\right). \tag{3.6}$$

Since σ_x is by definition always positive, σ_{xu} has the same sign as $\beta^{OLS} - \beta_z^{IV}$ if and only if $1/\sigma_x - (\lambda^*\sigma_z/\sigma_{xz}) > 0$. Transforming this equation gives $\sigma_{xz} - \sigma_x\sigma_z\lambda^* > 0$ or transformed $\rho_{xz} > \lambda^*$.

Combining both cases yield Lemma 2:

Lemma 2: A1-A3 hold. If $\sigma_{xu} < 0$ or $\rho_{xz} > \lambda^*$ then $\text{sgn}(\beta^{OLS} - \beta_z^{IV}) = \text{sgn}(\beta^{OLS} - \beta)$

Expressed verbally, Lemma 2 states that even if the IV is not valid, the IV estimate can correct the OLS estimation in the right direction. The intuition of Lemma 2 is as follows: When λ^* is small it means that the correlation between instrument and error term is small in comparison to the correlation between the endogenous covariate and the error term. If this is the case, the IV is almost exogenous because the bias is small. Accordingly, the IV corrects in the right direction. Unfortunately, only the first condition can be verified, because λ^* is unknown.

3.2 Imposing A4 (instrument is "less endogenous" than z)

Imposing now A4 will give sharper identification results, which are derived now and will be summarized in proposition 2.

Supposing $\rho_{xu} > 0$ gives with A4 that $\rho_{xu} \geq \rho_{zu} \geq 0$. Plugging in yields

$$\begin{aligned} \frac{\sigma_{xu}}{\sigma_x\sigma_u} &\geq \frac{\sigma_{zu}}{\sigma_z\sigma_u} \\ \Leftrightarrow \sigma_{xu}\sigma_z &\geq \sigma_{zu}\sigma_x \\ \Leftrightarrow (\sigma_{xy} - \sigma_x^2\beta)\sigma_z &\geq (\sigma_{zy} - \sigma_{xz}\beta)\sigma_x \\ \Leftrightarrow \frac{\sigma_z\sigma_{xy}}{\sigma_z(\sigma_z\sigma_x - \sigma_{xz})} &\geq \beta \end{aligned} \tag{3.7}$$

The expression on the left side of the equation is defined as $\beta_{v(1)}^{IV}$ so that

$$\beta_{v(1)}^{IV} \geq \beta \tag{3.8}$$

It is important to note the difference between β_z^{IV} and $\beta_{v(1)}^{IV}$. The former denotes the IV estimator, whereas the latter is the estimator derived from the above-defined function $v(\lambda)$ with $\lambda = 1$. $\beta_{v(1)}^{IV}$ can be expressed as a weighted function of β_z^{IV} and β^{OLS}:

$$\beta_{v(1)}^{IV} = \gamma\beta^{OLS} + (1-\gamma)\beta_z^{IV} \tag{3.9}$$

with $\gamma \equiv \sigma_z\sigma_x/(\sigma_z\sigma_z - \sigma_{xz}) = 1/(1-\rho_{xz})$

If instead $\rho_{xu} < 0$ would be true, the derived result would be inverted, meaning $\beta_{v(1)}^{IV} \leq \beta$. Investigating the different possible values for σ_{xz} and σ_{xu} it is possible to derive Proposition 2:

Proposition 2: A1-A4 hold. If $\sigma_{xz} < 0$, then

$$B_{\text{P2}}^* = \begin{cases} \left[\beta_z^{IV}, \beta_{v(1)}^{IV}\right]] & \text{if } \sigma_{xu} > 0 \\ \left[\beta_{v(1)}^{IV}, \beta_z^{IV}\right] & \text{if } \sigma_{xu} < 0 \end{cases}$$

with $\beta^{IV_{v(1)}} \equiv \frac{\sigma_z\sigma_{xy}}{\sigma_z(\sigma_z\sigma_x - \sigma xz)}$

If $\sigma_{xz} > 0$, then

$$B_{\text{P2}}^* = \begin{cases} \left(-\infty, \min(\beta_z^{IV}, \beta_{v(1)}^{IV})\right] & \text{if } \sigma_{xu} > 0 \\ \left[\max(\beta_z^{IV}, \beta_{v(1)}^{IV}), -\infty\right) & \text{if } \sigma_{xu} < 0 \end{cases}$$

The bounds are sharp.

So how did A4 improve the results? To asses this question it is necessary to compare Lemma 1 and Proposition 2. When $\sigma_{xz} > 0$ the bounds are still one sided. This result might surprise since λ^* is bounded from both sides between 0 and 1. An intuition for this finding can be seen when $\lambda^* = \rho_{xz}$. In this case, $\beta^{IV_{v(1)}}$ would not be defined and therefore β would not be binded. However, using A4 improved the bounds slightly because β^{OLS} is replaced by $\beta_{v(1)}^{IV}$.

In contrast, if $\sigma_{xz} < 0$ the bounds are largely improved. The bounds are also two sided without A4, but they are narrower in Proposition 2. To asses, to which degree the bounds are improved, the researchers derive Corollary 2.

Corollary 2: If $\sigma_{xx} < 0$, then

$$\frac{\beta^{IV}_{v(1)} - \beta^{IV}_{z}}{\beta^{OLS} - \beta^{IV}_{z}} = \frac{1}{1 - \rho_{xz}}$$

The corollary implies that the greater the magnitude of the correlation between x and z, the tighter is the bound that is achieved by adding A4. The maximal improvement can be obtained when $\rho_{xz} = 1$. In this case, the size of the bound is halved.

Propostion 2 also reveals the key finding of the paper: It is possible that if the IV variable is imperfect that β^{OLS} is closer to β than β^{IV}. This is even true when the instrument Z is less correlated with the error term U than the endogenous regressor X. Nevo and Rosen 2012 illustrate this finding for the case when $\sigma_{xz} < 0$ and $\sigma_{xu} < 0$. Then $\beta \in \beta^{IV}_{v(1)}, \beta^{IV}_{z}$ and as equation 3.9 reveals, $\beta^{OLS} < \beta^{IV}_{v(1)}$ and β^{OLS} is closer to β than β^{IV}_{z}. Also, reformulating equation 3.9 yields $\beta^{OLS} - \beta^{IV}_{v(1)} = \frac{\sigma_{xz}}{\sigma_x \sigma_z}(\beta^{IV}_{v(1)} - \beta^{IV}_{z})$ as an expression for the distance between β^{OLS} and $\beta^{IV}_{v(1)}$. Combining these three information shows that $\beta^{IV}_{v(1)}$ is closer to β^{OLS} than is β^{IV}_{z}. Hence, in this case if β is closer to $\beta^{IV}_{v(1)}$ it is possible that β is closer to β^{OLS} than to β^{IV}_{z}. Therefore, the IV estimator might over-correct and β^{OLS} is a better estimator than β^{IV}.

4 Additional Regressors

The authors extend the model to also include other regressors W:

$$Y = X\beta + W\delta + U \tag{4.1}$$

It is assumed that there exist a matrix of valid instruments Z_w for the additional regressors which has the same dimension as W.

To facilitate the analysis the authors apply sequential regression. By multiplying the estimation equation with a so-called annihilator matrix the authors will be able to modify the equation in such a way that it only depends on one regressor. In order to do so they first define the residuals of an IV regression of Y and X on W.

$$\tilde{X} = X - WE[Z^w W]^{-1}\mathbf{E}[Z^w X] \tag{4.2}$$

$$\tilde{Y} = Y - WE[Z^w W]^{-1}\mathbf{E}[Z^w Y] \tag{4.3}$$

Subtracting each side of the original equation by $\tilde{Y} = Y - W\mathbf{E}[Z^w W]^{-1}\mathbf{E}[Z^w Y]$ and replacing $Y = X\beta + W\delta + U$ on the left-hand side of the equation yields the following simplified estimation equation:

$$\tilde{Y} = \tilde{X}\beta + U \tag{4.4}$$

The resulting estimator of this equation is the following: $\beta = \sigma_{x\tilde{y}}\sigma_{x\tilde{x}}$

This facilitates the following analysis significantly. Proposition 2 can now simply be restated with regard to the above-specified estimator, which leads to the the following Proposition:

Proposition 3: If A1, A3 and A4 hold and there exists an exogenous instrument for the additional regressor Z^w and assume that $\rho_{xu} \geq 0$. If $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} > 0$, then B^* is a closed interval if $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} \leq 0$ B^* is an open interval. The estimators are defined as follows, with each factor of $\beta^{IV}_{v(1)}$being multiplied by -1.

$$\beta^{IV}_{z} = \frac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}} \tag{4.5}$$

$$\beta^{IV}_{v(1)} = \frac{\sigma_{v(1)\tilde{y}}}{\sigma_{v(1)\tilde{x}}} = \frac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z} \tag{4.6}$$

The identification sets for B_* are now given by:

$$B^* = [\beta^{IV}_{v(1)}, \beta^{IV}_{z}] \text{ if } \sigma_{z\tilde{x}} < 0 \tag{4.7}$$

$$B^* = [\beta^{IV}_{z}, \beta^{IV}_{v(1)}] \text{ if } \sigma_{z\tilde{x}} > 0 \tag{4.8}$$

The proof of this proposition is analogously to the proof of proposition 2. If B^* can only be defined as an open interval since $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} \leq 0$ then

$$B^* = [\max\{\beta_z^{IV}, \beta_{v(1)}^{IV}\}, \infty) \text{ if } \sigma_{z\tilde{x}} < 0 \tag{4.9}$$

$$B^* = [-\infty, \min\{\beta_z^{IV}, \beta_{v(1)}^{IV}\}) \text{ if } \sigma_{z\tilde{x}} > 0 \tag{4.10}$$

4.1 Bounds on other coefficients

Recalling that the estimators of additional regressors δ are also biased if β is biased the established bounds will be useful to obtain estimates of δ. Since we get no precise estimate of β we can only obtain bounds for the effects of the additional regressors as well. Define $Y^* = Y - X\beta$ to get

$$Y^* = W\delta + U \tag{4.11}$$

Since we make the assumption that the instruments Z_w are exogenous we obtain the following estimator for δ

$$\delta = \mathbf{E}[Z^w w]^{-1}\mathbf{E}[Z^w Y^*]\mathbf{E}[Z^w W_j], \tag{4.12}$$

$$\delta = \mathbf{E}[Z^w w]^{-1}\mathbf{E}[Z^w Y^*] \tag{4.13}$$

Analogously the estimation on one single component of δ can be defined as:

$$\delta_j = \mathbf{E}[Z^w \tilde{W}_j]^{-1}\mathbf{E}[Z_j^w \tilde{Y}^*] \tag{4.14}$$

With

$$\tilde{W}_j = W_j - W_{-j}\mathbf{E}[Z_{-j}^w W_{-j}]^{-1}\mathbf{E}[Z^w W_j], \tag{4.15}$$

$$\tilde{Y}^* = Y^* - W_{-j}\mathbf{E}[Z_{-j}^w W_{-j}]^{-1}\mathbf{E}[Z^w Y^*], \tag{4.16}$$

These transformations lead to Proposition 4:

Proposition 4: If the conditions for proposition 3 hold one can just replace Y^* in the estimate for δ_j by $Y^* = Y - X\beta$. When inserting the maximum and minimum value of the identification set B^* for β one obtains the following estimates for δ_j

$$\delta_{j0} = \mathbf{E}[Z_j^w \tilde{W}_j]^{-1}\mathbf{E}[Z_j^w \tilde{Y}] - \mathbf{E}[Z_j^w \tilde{W}_j]^{-1}\mathbf{E}[Z_j^w \tilde{X}]\beta_L \tag{4.17}$$

$$\delta_{j1} = \mathbf{E}[Z_j^w \tilde{W}_j]^{-1}\mathbf{E}[Z_j^w \tilde{Y}] - \mathbf{E}[Z_j^w \tilde{W}_j]^{-1}\mathbf{E}[Z_j^w \tilde{X}]\beta_U \tag{4.18}$$

It is important to note that $\tilde{y}$ is now defined differently as before:

$$\tilde{y} = Y - W_{-j}\mathbf{E}[Z_{-j}^{w'} W_{-j}]^{-1}\mathbf{E}[Z_{-j}^{w'} Y] \tag{4.19}$$

5 Multiple Imperfect Instruments

The authors consider the case when multiple imperfect instruments are available for the exogenous regressor X. The availability of such instruments can yield more precise bounds on β and δ. The authors consider that a number of k_z instruments Z_r (r=1,...,k_z) are available.

Proposition 5: Assume A1-A4. Then the identified set for $B^* = [\max_r \beta_{l,r}, \min_r \beta_{u,r}]$ and the identification region for δ_j, each j=1,..k_w is $D_j^* = [\max_r \delta_{l,r}^j, \min_r \delta_{u,r}^j]$. These bounds are sharp.

In order to establish the sharpness of the bounds we first need to derive Corollary 3 to Proposition 5, which states

Corollary 3 If $\sigma_{xz} < 0$, then any $\lambda^* \in [0, 1]$ is feasible. If $\sigma_{xz} > 0$, then if $\beta_{v(1)}^{IV} > \beta_z^{IV}$, $\lambda^* \in (\rho_{xz}, 1]$, while if $\sigma_{xz} > 0$ and $\beta_{v(1)}^{IV} < \beta_z^{IV}$, then $\lambda^* \in [0, \rho_{xz})$

In the following we will show that this corollary holds if two-sided bounds for β can be established. To do so we rearrange λ^* in the following way:

$$\lambda^* = \frac{\rho_{zju}}{\rho_{xu}} = \frac{\sigma_x \sigma_{\tilde{y}z} - \beta\sigma_x\sigma_{\tilde{x}z}}{\sigma_z\sigma_{z\tilde{y}} - \beta\sigma_z\sigma_{\tilde{x}x}} \tag{5.1}$$

which allows us to take the derivative of λ^* with respect to β:

$$\frac{d\lambda^*}{d\beta} = \sigma_x\sigma_z\frac{\sigma_{\tilde{x}\tilde{x}}\sigma_{\tilde{y}z} - \sigma\tilde{x}z\sigma_{x\tilde{y}}}{\sigma_z\sigma_{\tilde{x}y} - \beta\sigma_z\sigma_{\tilde{x}\tilde{x}}} \tag{5.2}$$

We can now see that λ^* is increasing in β in the direction of $\sigma_{\tilde{x}\tilde{x}}\sigma_{\tilde{y}z} - \sigma\tilde{x}z\sigma_{x\tilde{y}}$ When we can derive two-sided bounds for β we know from Proposition 3 that: $(\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z)\sigma_{z\tilde{x}} > 0$ and the bounds on β are $\frac{\sigma_{z\tilde{y}}}{\sigma_{z\tilde{x}}}$ for $\lambda^* = 0$ and $\frac{\sigma_{z\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_z}{\sigma_{z\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_z}$ for $\lambda^* = 1$

One can now apply Proposition 3 to each endogenous Instrument Z_j individually, which implies that β falls into the bounds obtained for each instrument. We can now easily show that these bounds are sharp, meaning that every β within these bounds is feasible. From Corollary 3 we plug λ^* into the equations obtained above such that there exists a λ^*_j for each j:

$$\beta = \frac{\sigma_{zj\tilde{y}}\sigma_x - \sigma_{x\tilde{y}}\sigma_{zj}\lambda^*_j}{\sigma_{zj\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{zj}\lambda^*_j} \tag{5.3}$$

By transforming this equation yields $\lambda^*_j = \frac{\rho_{zu}}{\rho_{xu}}$. We know that $\lambda \in [0,1]$ from Corollary 3, so A3 and A4 are true and the value obtained for β is actually feasible. This proposition implies that if each of the available imperfect instruments provides upper and lower bounds for β and δ then the identification set lies between the highest value for the lower bound and lowest value for the upper bound of all r estimated boundaries, since all estimations provide valid results. Thus, the use of multiple instruments can lead to significant tightening of the estimated bounds.

The authors now consider the case under which $(\sigma_{zx}\sigma_x - \sigma_{xx}\sigma_z)\sigma_{zx}<0$ and every instrument just provides a one-sided bound for β. If one of the instruments Z_1 and Z_2 appears to be more valid and relevant one might use this additional assumption to construct a weighted average of both instruments: $\omega(\gamma) = \gamma Z_2 - (1-\gamma)Z_1$

Proposition 6: Assume A1-A3 hold for both Z_1 and Z_2 and in addition assume that there exists some $\gamma^* \in [0,1]$ such that $\sigma_{\omega(\gamma^*)u} \geq 0$ and $(\sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)})\sigma_{\omega(\gamma^*)\tilde{x}} \geq 0$. This yields the following estimates for β

$$\beta^{IV}_{\omega(\gamma^*)} \leq \beta \leq \min\{\beta^{IV}_{Z1}, \beta^{IV}_{Z2}, \beta^{OLS}\} \tag{5.4}$$

Hence, if one can construct a weighted average of instruments such that this function fulfills the assumptions of proposition 2 and then runs an estimation with it, two sided bounds for the coefficients β and δ can be obtained. The requirements for the existence of such a weighting function are specified in Lemma 3. If $(\sigma_{\omega(\gamma^*)\tilde{x}}\sigma_x - \sigma_{x\tilde{x}}\sigma_{\omega(\gamma^*)})\sigma_{\omega(\gamma^*)\tilde{x}} \geq 0$ holds it suffices to check that $\sigma_{z_1\tilde{y}}\sigma_{\tilde{x}z_2} < \sigma_{z_2\tilde{y}}\sigma_{\tilde{x}z_1}$.This stems from ensuring that A3 and A4 hold. First, analogue to A3 the weighted sum of σ_{z_1u} and σ_{z_2u} needs to have the same direction as the correlation between X and the error term σ_{xu}. Second analogue to A4 the weighted sum of the correlation between instruments and the regressor X $\sigma_{\tilde{x}z_1}$ and $\sigma_{\tilde{x}z_2}$ is negative. If these conditions holds there exists an optimum weight γ^* such that two-sided bounds for β can be estimated.

If A4 (instruments are less endogenous than the error term) holds for both instruments it is possible to further tighten the bounds:

$$\beta^{IV}_{\omega(\gamma^*)} \leq \beta \leq \min\{\beta^{IV}_{Z1}, \beta^{IV}_{Z2}, \beta^{IV}_{v1(1)}, \beta^{IV}_{v1(2)}, \beta^{IV}_{v*(1)}\} \tag{5.5}$$

The intuition behind this assumption is straightforward. If we can impose A1-A4, we simply derive the same estimators as in the case with one available instrument for both instruments Z_1 and Z_2. However, now an additional moment condition can be constructed using the function $V^*(1) = \sigma_x\omega(\gamma^*) - \sigma_{\omega(\gamma^*)}X$. This additional moment condition is based on the notion of treating the weighting function of both instruments $\omega(\gamma^*)$ as a new instrument, which fulfills A1-A4. Since all of the estimators are obtained from valid data-generating processes we can now take the minimum as our upper bound for the estimate of β.

6 Application: Back to the Roots - The Effect of Education on Wages

6.1 Card's findings

We decided to apply the above described procedure to obtain valid bounds to the estimation performed in Card 1993. In this well-known paper, Card estimates the return to schooling in terms of future income, instrumenting schooling by geographical proximity of the child to college. Card 1993 obtains his data from the National Longitudinal Survey of Young Men (NLYSM), which was initiated in 1966. The dataset contains information about labor market outcomes of surveyed individuals and a variety of information about their respective background. The information used for estimating the labor market outcomes is obtained from the 1976 wave. The instruments used to approximate the level of education obtained are dummy variables indicating whether the child grew up in an area close to an accredited 4 year college. Card 1993 argues that this instrumental variable provides an exogenous source of variation in schooling, since individuals who grew up closer to a college have a higher level of schooling even when controlling for predicted level of education (based on race, age regional and family background characteristics). Thus, after controlling for all those factors distance to college appears to be an exogenous, random determinant of the level of schooling. Running an OLS estimation Card 1993 obtains estimates of an average increase of wages by 7.3 percent for one additional year of schooling, using college distance as an instrument he estimates a 13.2 percent increase.

Yet, it is important to bear in mind that the IV estimates need to be interpreted as the local average treatment effect, the effect of education on wages for those individuals who went to college since they lived in close proximity, but would not have gone to college if they would have lived far away. The results in Card 1993 suggest that the OLS estimate of returns to schooling is downward biased (which suggests a negative correlation between the unobserved error term and schooling). The findings in his paper contradict the hypothesis that the education variable in the wage regression is subject to the so-called "ability bias". This refers to the idea that higher capable students have higher levels of schooling and higher wages, yet not due to their education, but due to their "ability".

6.2 Estimation assuming the instrument is not perfect

The findings of Card 1993 rest on the assumption that his instrument, proximity to a college, is valid. The relevance condition is easily verified, contrary to the exogeneity condition. As already suggested in the original paper of Card 1993, the instrument might fail the latter condition. For example parents who care more about the education of their children might move to an area, which is closer to a college. We will elaborate on this later. Thus, it seems plausible that the instrument used by Card 1993 indeed imperfect and the procedure described above could provide a remedy. In the following application we will first replicate the estimation of the paper by Card 1993 using the regression specification of Verbeek 2004. Analogue to both, we include the following control variables in the OLS and IV estimation: experience, squared experience, being black, living in the south and a dummy for living in the South and a Metropolitan area (smsa) which yields the following regression equation:

$$Wage_i = \beta_0 + Educ_i\beta_1 + Exp_i\beta_2 + Exp_i^2\beta_3 + black_i\beta_4 + South_i\beta_5 + Smsa_i\beta_6 \tag{6.1}$$

For our IV regression we estimate the endogenous variable schooling by a dummy variable, indicating whether the individual lived near a college. Since experience and experience squared are endogenous with respect to schooling we approximate both variables by age and age squared respectively. Our replication of the OLS and IV regressions yield the same results as obtained by Verbeek 2004.

In order to apply the estimation procedure as detailed in the paper by Nevo and Rosen 2012 we first recode the dummy variables indicating proximity to a 4-year and 2-year college respectively.[1] In contrast to the specification in Card 1993 our dummy variables are now 1, if the individual did not live near a college and 0 if it did. Recalling the estimation procedure for Imperfect Instruments we will now verify the assumptions, which are crucial for our estimation procedure to work.

1. $\sigma_{xz} < 0$: According to Lemma 1 the instrument and the endogenous regressor need to be negatively correlated. This can easily be verified since Card 1993 shows that education decreases with the distance to college.

2. sgn(ρ_{xu})= sgn(ρ_{zu}): According to A3 the correlation of regressors and instruments with the unobserved variable need to have the same direction. In a first step, based on the findings of Card 1993 we assume that the sign of both correlations is negative. The reasoning for this

[1] Note that this is possible, as we also take into account the correlation between the IIV and the error term, which we assume to be negative in a first step. In general though, one must be careful when recoding variables.

is as follows: as outlined above the correlation between the error term and the instrument might be negative since the decision of the parents where to locate might depend on their prospects and ambitions for their child. Thus, parents who are particularly interested in their child's education level might be more ambitious and transfer this ambition to their children, which will make them obtain more schooling and be more ambitious in their job as well. Hence, there will be a negative correlation between the distance to college and the unobserved variable[2], $\rho_{zu} < 0$. Further, in this first step we assume that the correlation between schooling and unobserved variable is negative as well, $\rho_{xu} < 0$. If we would assume the instruments to be perfect, a negative correlation between education and unobserved variable is actually what is suggested by the estimation of Card 1993 and Verbeek 2004, who find a much lower coefficient for the OLS estimation than for the IV estimation.

   One reason why there could exist such a correlation would be a sort of "reverse ability bias". Abilities, which are later rewarded in the labor market, might have a negative impact on the schooling decision for some individuals. It is arguable that individuals, who are more conformist tend to do better in school, but might be less innovative and self-assertive, skills which will later be rewarded in the labor market. This would lead to a negative correlation between schooling and the unobserved effect $\rho_{xu} < 0$.

   Another possibility which is brought forward to explain the fact that the IV estimation increases the schooling coefficient is measurement error. Indeed, if measurement error in the schooling variable is present, the OLS estimator will underestimate the true coefficient of schooling. Yet, it seems rather simplistic to ad-hoc assume the existence of a measurement error when the IV estimate produces larger estimates than OLS.[3]

3. Analogue to A4 we will also impose the assumption that $\rho_{xu} > \rho_{zu}$, the instruments are less endogenous than x. This assumption seems plausible since the correlation between error term and instrumental variable according to the above-described channel might be fairly small. Also, it seems more plausible that the distance to a college is less correlated with the many determinants of wage which are located in the error term than education itself.

The estimates for this regression are presented in Table 1, together with the baseline OLS and IV estimates. The bound estimates suggest that the beta coefficient is indeed slightly higher than the OLS estimate, yet, the IV estimate of Card 1993 present the upper bound of the identification interval (when focusing on the estimates not their confidence interval, the small numerical difference seen in Table 1 is probably due do small sample bias which arises when differencing out the effect of the other variables). This suggests that considering geographic proximity as "perfect" instrument, would lead to estimates which are at least on the upper bound of possibilities, even if we assume that OLS produces downward biased results. A note of caution should be added to the OLS and IV standard errors. As it turns out, Verbeek 2004 did not use robust standard errors for his example, we opted to go with that convention as we only use the OLS and IV estimates for illustrative purpose and the coefficients are still unbiased in the case of heteroscedasticity.

6.3 Estimation with two imperfect instruments

An argument which is frequently brought forward when estimating the effect of education on wages is the so called "ability bias". Individuals with a higher ability tend to have more schooling but at the same time, ability positively influences earnings because the market values ability (for example to perform complex analytical tasks) rather than education itself. This would suggest that there is a positive correlation between the error term and education, suggesting that OLS overestimates the effects of education on wages. This would contradict the findings of Card 1993, if we assume that college proximity is a valid instrument. As ability cannot be measured, some authors include IQ as a crude proxy for this. IQ has a rather high correlation with the education choice ($\rho = 0.499$). And when we run the same OLS regression as before but include IQ, the coefficient for education is indeed lower than in the previous specification (the same is true for the "Knowledge of the World of Work score" (KWW) variable which Card used in his work to proxy for the ability bias as it was more widely available than IQ scores). This could be an indicator for such an ability bias.

[2] As the "will" of parents to send their kid to college decreases with said distance.

[3] This is not to say that measurement error is not a valid possibility. Rather, measurement error should be anticipated before the estimations are run. Unfortunately, many papers still rely on measurement error as an ad-hoc explanation for larger IV estimates (see for example Alfaro et al. 2004).

Table 1: Regression under hypothesis that $\rho_{xu} < 0$

	Dependent variable:		
	Log of wage		
	OLS	IV	IIV
	(1)	(2)	(3)
education	0.074^{***}	0.133^{***}	[0.079, 0.132]
	(0.004)	(0.051)	(0.067, 0.178)
experience	0.084^{***}	0.056^{**}	[0.056, 0.076]
	(0.007)	(0.026)	(0.041, 0.090)
experience^2	-0.002^{***}	-0.001	[-0.002, -0.001]
	(0.0003)	(0.001)	(-0.003, -0.000)
black	-0.190^{***}	-0.103	[-0.185, -0.103]
	(0.018)	(0.077)	(-0.189, -0.089)
smsa	0.161^{***}	0.108^{**}	[0.108, 0.159]
	(0.016)	(0.050)	(0.095, 0.171)
south	-0.125^{***}	-0.098^{***}	[-0.124, -0.098]
	(0.015)	(0.029)	(-0.136, -0.085)
Constant	4.734^{***}	4.066^{***}	[4.066, 4.722]
	(0.068)	(0.608)	(4.002, 4.781)
Observations	3,010	3,010	3,010
R^2	0.291	0.176	
Adjusted R^2	0.289	0.175	
Residual Std. error (df = 3003)	0.374		
F Statistic	204.932^{***} (df = 6; 3003)		
Std. Errors/CI in parentheses			*p<0.1; **p<0.05; ***p<0.01

Yet, both measures are arguably incomplete to capture all of the ability bias and IV estimates are still necessary.

As argued above, the instrument brought forward by Card 1993 might fail the exogeneity condition and would therefore constitute an IIV. A positive correlation between education and the error term, would suggest that a single instrument only provides one sided bounds. But the use of two instruments allows to get two sided bounds nevertheless. As outlined in the paper we also construct an additional moment consisting of two imperfect instruments. It is important to bear in mind, that in order to satisfy A3, we now need to use instruments, which are positively correlated with the error term. Since we assumed that the instruments we used to estimate our first regression are negatively correlated with the error term, we now rely again on the original dummy variables from Card 1993, these dummies take the opposite values of the dummies constructed for our regression and consequently, $\rho_{z1u} > 0$ and $\rho_{z2u} > 0$. We use the weight, suggested by the researcher. Distance to a 2-year college is weighted with a value of γ of 0.4., distance to a 4-year college with $1 - \gamma = 0.6$.

This specification implicitly entails the assumption that distance to a 4-year college is a more relevant instrument than distance to a 2-year college, which is confirmed by our data. The weighted function is now negatively correlated with X, which allows us to construct the two sided bounds. As explained above, we now assume the existence of an ability bias, which translates to $\sigma_{xu} > 0$. Reestimating our equation with these two additional instruments we achieve two sided bounds on our education variable.

Results of the OLS regression including IQ, which we argue provides support for the ability bias hypothesis, and the IIV estimation with two instruments are provided in Table 2. Note that there are fewer observations for the OLS regression including IQ because the IQ variable was not reported for all observations. Again, to ensure comparability to the previous results, the OLS estimation does not report robust standard errors. Again, we rely on A4 (instrument is less endogenous than X) and provide tests for the necessary conditions to use two IIVs in our separate code file, as they are more involved than for a single instrument.

Table 2: Regressions under hypothesis that $\rho_{xu} > 0$ (ability bias)

	Dependent variable:	
	Log of wage	
	OLS	IIV Two instruments
	(1)	(2)
education	0.069^{***}	[0.041, 0.067]
	(0.005)	(-0.02, 0.073)
experience	0.094^{***}	[0.080, 0.090]
	(0.010)	(0.066, 0.104)
experience^2	-0.003^{***}	[-0.003, -0.002]
	(0.0005)	(-0.003, -0.001)
black	-0.136^{***}	[-0.238, -0.200]
	(0.026)	(-0.253, -0.186)
smsa	0.153^{***}	[0.168, 0.192]
	(0.019)	(0.157, 0.204)
south	-0.079^{***}	[-0.141, -0.129]
	(0.018)	(-0.153, -0.117)
IQ	0.003^{***}	
	(0.001)	
Constant	4.483^{***}	[4.84, 5.152]
	(0.104)	(4.781, 5.212)
Observations	2,061	3,010
R^2	0.226	
Adjusted R^2	0.223	
Residual Std. error	0.368 (df = 2053)	
F Statistic	85.493^{***} (df = 7; 2053)	
Std. errors/CI in parentheses		*p<0.1; **p<0.05; ***p<0.01

6.4 Interpretation of results

How can our results be interpreted? The results from our estimation with only one imperfect instrument are not very precise[4]. Furthermore, although there is some reason to believe that there might be a negative correlation between unobserved variable and schooling decision (which would be in line with the results from Card 1993), this seems in contrast with at least a part of economic theory. What can be taken away from this application, is that IV provides estimates which are at the upper bound of the possible values for the true effect of education on wage, even if we assume a negative correlation between the error term and education in the standard wage regression.

However, the results from the application with two imperfect instruments provide some different and interesting insights. The bounds achieved are now by far tighter than before and also yield estimates, which have confidence intervals lower than the OLS estimates of Card 1993. Extending the analysis to the inclusion of imperfect instruments actually provides evidence for an ability bias. As Nevo and Rosen 2012 argue before, imperfect instruments with $\sigma_{zu} \neq 0$ will severely bias the IV estimator in particular if the correlation between X and Z is small, which is the case with at least the "near two year college" variable ($\rho_{xz} = 0.047$). Thus, taking into account our arguments that college proximity might be an IIV, one can conclude, that the IV estimation performed by Card 1993 might be subject to severe bias. Extending the analysis to consider the case of imperfect instruments provides support for Cards consideration that OLS overestimates the effect of education.

Yet, if this term paper has shown one thing, it is that IV estimates should be considered with a fair amount of caution. If the IV is correlated with the error term[5], IV produces biased results. The bound estimate procedure proposed in Nevo and Rosen, can provide more credibility to the results at the price of having only interval identified results. Yet, even these intervals are sensitive to the (unmeasurable) assumptions made about the correlation between the variables and the error term. Arguments based on theory will therefore continue to be an important part in any paper using the IV approach.

References

Alfaro, Laura et al. (2004). "FDI and economic growth: the role of local financial markets". In: *Journal of international economics* 64.1, pp. 89–112.

Card, David (1993). *Using geographic variation in college proximity to estimate the return to schooling.* Tech. rep. National Bureau of Economic Research.

Nevo, Aviv and Adam M. Rosen (Aug. 2012). "Identification With Imperfect Instruments". In: *The Review of Economics and Statistics* 94.3, pp. 659–671.

Verbeek, Marno (2004). *A guide to modern econometrics.* John Wiley & Sons.

[4]The bounds are quite large and so are the confidence intervals.

[5]Which can be argued for many papers, see for example the slides of Anett John' and Bruno Crepon' course on microeconometric evaluation of public policies at ENSAE.