A Matter of Size: Comparing IV and OLS estimates

Riccardo Ciacci*

February 25, 2021

Abstract

Sizeable differences between OLS and IV estimates might be interpreted in the literature as evidence that the instrument is not valid. Yet, to the best of my knowledge, this comparison is carried out using only the OLS coefficient as a benchmark but without taking into account any statistical measurement or information from the OLS regression. This paper establishes a framework where Oster (2017)'s methodology might be used to compare objectively OLS and IV estimates. This methodology offers evidence to support or discard validity of the instrument.

Keywords: instrumental variables, comparison IV and OLS estimates, instrument validity.

JEL codes: C26.

1 Introduction

Instrumental variables techniques are a fundamental method in the econometrics toolkit in order to solve both issues of endogeneity and measurement error (Angrist and Krueger 2001; Angrist and Pischke 2008). A common approach in econometrics is to interpret considerable differences between the size of Ordinary Least Squares (OLS) and Instrumental Variables (IV) coefficients as evidence against the validity of the instrument. The intuitive appeal of this strategy is that the OLS regression can be informative about the true effect the researcher wants to estimate. Nonetheless, in the literature, to the best of my knowledge, there is no formal methodology to compare these two estimates.

Oster (2017) makes use of information from the OLS regression - such as, inclusion of controls, size of variances and movement of R^2, etc.- to estimate a set of values where the true treatment effect should lie. The size of such set depends on how "informative" the observables are about the unobservables according to the researcher. Consequently, this methodology allows the researcher to compute a parameter to develop a formal bounding argument. This parameter is known as *coefficient of proportionality* and measures the relative size of the proportionality between selection on observables and unobservables.

*Universidad Pontificia Comillas, Department of Economics and Business Management, rciacci@icade.comillas.edu

This article argues that this methodology might be useful to make objective comparisons between IV and OLS estimates. To this end, taking into account that the IV estimates measure the effect only for the population whose choice of treatment is affected by the instrument, this paper suggests that larger (smaller) values of the coefficient of proportionality are evidence against (in favour of) the validity of the instrument.

The main contribution of this paper is to suggest a formal methodology through which scholars can compare IV and OLS estimates. First, this paper is related to several recent papers comparing the relative size of such two estimates without a formal methodology (see, among others, Alesina et al. (2013); Mejia and Restrepo (2013); Bhuller et al. (2020); Liu (2020). Second, this paper is also related to a branch of the literature that uses observables to asses the bias generated by unobservables in OLS settings (Murphy and Topel 1990; Altonji et al. 2005a,b, 2010; Oster 2017; De Luca et al. 2019).

The rest of the paper is divided in the following sections. Section 2 introduces the methodology intuitively and suggests how to implement it. Section 3 applies the methodology to different settings. Section 4 briefly explains how to use this methodology in Stata. Section 5 concludes.

2 Intuition

Let the population regression function (hereinafter, PRF) be given by the panel data regression:

$$y_{ih} = \alpha_1 + \beta_1 d_{ih} + \gamma w_{ih} + \theta_1 X_{ih} + \varepsilon_{1ih} \tag{1}$$

where, for each variable, ih denotes unit i at time h. d is the (scalar) treatment we are interested in, w is the vector of unobserved controls and X is the vector of observed controls. Given the nature of w we can only run

$$y_{ih} = \alpha_2 + \beta_2 d_{ih} + \theta_2 X_{ih} + \varepsilon_{2ih} \tag{2}$$

If we are in a setting where the assumptions exposed in Oster (2017) are plausible we can take into consideration the proportional selection relationship given by:

$$\delta \frac{Cov(d_{ih}, X_{ih})}{Var(X_{ih})} = \frac{Cov(d_{ih}, w_{ih})}{Var(w_{ih})} \tag{3}$$

which holds for some $\delta \neq 0$.

Now consider the simple univariate setting. We know that in the case where w is the only control, and it is omitted from the regression, the omitted variable bias is given by:[1]

$$\hat{\beta}_2 = \beta_1 + \gamma \frac{Cov(d_{ih}, w_{ih})}{Var(d_{ih})} \tag{4}$$

Recall that if the instrument is valid the IV coefficient consistently estimates β_1. There-

[1]The univariate setting is useful to develop intuition. Appendix Section A develops the multivariate case. Results are similar.

fore, given equation (3), we can compute how large δ needs to be to support the difference in size between the OLS estimates β_2 and the candidate to be the true effect β_1, the IV estimates. Plugging equation (3) into equation (4) and solving for δ we can easily get:

$$\delta = (\hat{\beta}_2 - \beta_1)\frac{Var(d_{ih})Var(X_{ih})}{\gamma Var(w_{ih})Cov(d_{ih}, X_{ih})} \tag{5}$$

A low coefficient of proportionality is evidence in favour of the IV estimates. Put it differently, a low δ means that not much selection on unobservables is needed to support that the true effect is the one estimated by the IV regression. Intuitively, δ tells us how large selection on unobservables, compared to observables, needs to be to support that the "true effect" has the size of the IV estimates. Since this analysis takes into account inclusion of controls, size of variances and movement of R^2, among other things, large values of δ might imply either that the instrument is not valid or that there are heterogeneous effects and the IV estimates are estimating them for a subpopulation (Angrist and Imbens 1995). In this circumstance (i.e. a large coefficient of proportionality) to distinguish between the two cases this methodology might be complemented using Masten and Poirier (2018) and carrying out an analysis, as in Bhuller et al. (2020), to explore whether in the data there is empirical evidence supporting the IV is estimating heterogeneous effects for a subpopulation.

In empirical settings in order to compute the identified sets it might be important to establish the sign of δ. In the simplified equation presented above (i.e. equation (5)) the sign of δ only depends on the sign we assume γ has, since variances are positive and we can estimate from the data all the other objects of the equation. This formula is a simplification drawn from the univariate setting, still the intuition is clear: giving the known sign of $(\hat{\beta}_2 - \beta_1)$ and $Cov(d_{ih}, X_{ih})$, we can guess the sign of δ depending on the sign of the effect we think the omitted variable has in the main regression. Once we have the sign we can compute the identified set for coefficients of proportionality with that sign to check how the bounds of the set vary as the coefficient of proportionality varies.

Lastly, a key input to estimate the identified set is the selection of the value that the R^2, statistic would take in a hypothetical full regression, a regression with both observables and unobservables (i.e. equation (1)). Oster (2017) denotes such value as R_{max}. To select R_{max} prior knowledge of the setting is crucial.[2] Namely, whether the researcher believes the full regression can explain the outcome variable completely. If this is the case R_{max} is set to 1.

Implementation:[3]

1. Set the sign of the coefficient of proportionality δ depending on the regression and the omitted variable the researcher wants to adress. Equation (5) can be helpful to this end.

[2] Oster (2017) discusses this issue in detail.

[3] This implementation is similar in spirit to the "Statements about δ" subsection discussed in Section 3.4 of Oster (2017). Unlike the afore-mentioned subsection, we do not have a suggested upper bound for δ. However, we would expect $|\delta| > 1$ since in our framework the IV regression is run since the OLS regression cannot pin down a causal relationship.

2. Set R_{max} according to knowledge of the setting.

3. Compute the exact value of the coefficient of proportionality δ, with the sign estimated above, that could explain the IV estimates.

4. This value measures the minimum amount of selection on unobservables (compared to observables) needed for the treatment effect to have the same size of the IV estimates (i.e. for the IV estimates to lie in the identified set). Consider/discuss whether this value is too large.

3 Empirical validation

This section makes use of different data sets to explore how to use the methodology introduced in the previous section. Comparing the OLS and IV estimates simply due to their relative size is not enough to assess whether the latter is a credible estimate of the true effect.

3.1 Observational data, example 1

In this section I use data from Ciacci (2018) where I estimate the effect of penalizing the purchase of prostitution on rape. Given selection into treatment in this setting, reverse causality and omitted variable bias are the main concerns connected to endogeneity of the treatment variable. Reverse causality arises from the concern that past values of rape could affect fines for sex purchase: prostitutes might prefer to locate in regions with low rapes. Omitted variable bias arises since I cannot control for variables that displace prostitutes. Such variables are negatively correlated with fines and positively with rape, leading OLS estimates to be downward biased.

To address these issues I use two instruments that exploit variation in flights to proxy access to sex tourism. The key identification assumption is that variation in the offering of intercontinental flights is independent of rape and fines for sex purchase patterns. In other words, the choice of flight companies to offer relatively more intercontinental flights does not depend on any reason connected to rape or fines for sex purchase. This seems plausible since to the best of my knowledge there is no evidence of flight companies that choose to offer more flights due to any reason connected to crime patterns.

My structural regression is:

$$log(1 + rape_{rmy}) = \beta fines_{rmy} + \alpha_r + \alpha_m + \alpha_y + \alpha_r * y + \gamma officers_{ry} + \varepsilon_{rmy} \qquad (6)$$

where r stands for region, m for month and y for year. The dependent variable is $log(1 + rape_{rmy})$ I use the variable in logs due to the dispersion of the distribution of rapes and $log(1 + y)$ since rape may take value 0, $fines_{rmy}$ is the number of fines for sex purchase issued by police officers in region r in month m and year y; α_r, α_m, α_y are respectively fixed effects for region, month and year; $\alpha_r * y$ is a region-year trend and the control variable $officers_{ry}$ is the number of police officers in region r in year y since police officers are hired regionally every year.

Therefore, as for equation (3) in this setting, d_{ih} is fines for sex purchase, X_{ih} is the number of police officers and w_{ih} are the omitted variables and past values of the outcome that displace prostitutes. In order to use Oster (2017)'s methodology the researcher needs to assess whether equation (3) is plausible in this setting. In this case this assumption boils down to whether we believe that selection on unobservables that displace prostitutes is proportional to the number of officers. To this extent, where there were higher past values of rape, currently there should be fewer prostitutes but more officers. Hence, the two appear to be oppositely related (i.e. negative coefficient of proportionality δ). Put it differently, in this setting it makes sense to suspect:

- Taking into account previous literature (see, inter alia, Farley and Barkan (1998); Farley et al. (2004)), unobservables (e.g. past values of rape) correlated with observables (e.g. officers) are positively correlated with the outcome variable (y following the nomenclature of 2), rape.
- The treatment variable: fines for sex purchase, d following the nomenclature of Section 2, is positively correlated with control variables, given results from the first-stage regression (available upon request).

Given the difference between the OLS and IV estimates, the sign of these two correlations imply the coefficient of proportionality δ is negative in this case.

Table 1 shows the coefficients estimated by Oster (2017)'s methodology setting a negative sign of the coefficient of proportionality δ and $R_{max} = 1$. The first ten rows of Table 1 show how large the coefficient of proportionality δ needs to be, column (1), to identify a coefficient of the size of column (2). In particular, for each value of δ the identified set is given by the coefficient estimated using Oster (2017)'s methodology and the OLS estimate.

The next-to-last row of Table 1 displays the OLS coefficient, whereas the last row shows the IV coefficient. Note that the latter is about 14 times larger than the former. At first sight this difference might appear considerable, however, taking into account Table 1, Oster (2017)'s methodology highlights that a negative coefficient of proportionality with size 1.16 is enough to identify a set that includes the IV estimates.[4] In other words, as long as selection on unobservables is slightly larger (i.e. 16 %) than selection on observables it is enough for the true treatment effect to have the size of the IV estimates.[5]

[4] The sign of the coefficient of proportionality δ is consistent with equation (5).

[5] This set could even seem conservative since officers are hired yearly per region. Hence, it is reasonable to expect that they might fluctuate relatively less compared to most crime variables.

Table 1: Comparison of OLS and IV estimates

(1)	(2)
δ	Coefficient
-1	0.01620
-2	0.03490
-3	0.07020
-4	0.11010
-5	0.12640
-10	0.15010
-20	0.15960
-50	0.16490
-100	0.16660
-1000	0.16810
Technique	Coefficient
OLS	0.00130
IV	0.01890

Notes: This table shows the bounds of the identified set using data from Ciacci (2018). Column (1) shows either the value of the coefficient of proportionality δ or the technique used. Column (2) shows the value of the associated estimated coefficient. For each coefficient of proportionality δ the bounds of the interval are given by the OLS estimates and the coefficient estimated with that coefficient of proportionality δ.

3.2 Observational data, example 2

This section applies the presented methodology to the seminal paper Acemoglu et al. (2001). The main threat to their OLS estimates is that there might be unobservables (mainly geographic and climatic features) positively correlated with economic growth and the treatment.[6] Given that IV estimates are greater than OLS estimates, there might be the concern that OLS regressions give rise to downward biased estimates. In this setting, a naive comparison between the two would highlight that the IV estimates are at most almost three times larger than OLS estimates. This setting is an example where even if the IV estimates are not much larger than the OLS estimates, it is necessary a large coefficient of proportionality δ to support IV estimates.

Table 2 considers the simplest IV regression with controls of Acemoglu et al. (2001).[7] In this regression model the authors regress the log of 1995 GDP per capita on average protection against expropriation risk between 1985–1995, instrumented with settler mortality, and control for continent dummies. Table 2 has the same format and assumptions (i.e. $R_{max} = 1$) of Table 1.

[6]For example consider temperate weather or natural resources. The paper also discusses a number of control variables which could affect economic growth but which might be affected by the treatment variable (such as current religion, diseases, etc.). This section is not meant to take into account these controls (*bad* controls).

[7]In Acemoglu et al. (2001) the output of this regression is displayed in column (7) of Table 4.

Further, in the setting of Acemoglu et al. (2001) it makes sense to suspect that:

- Taking into account previous literature (Gallup et al. 1999), unobservables (e.g. temperate weather) correlated with observables (e.g. latitude/geographic dummies) are positively correlated with the outcome variable (y following the nomenclature of Section 2), 1995 GDP per capita.

- Given results from the first-stage regression (Table 3 of Acemoglu et al. (2001)) and previous literature (Hall and Jones 1999), the treatment variable, d following the nomenclature of Section 2, is positively correlated with control variables.

Hence, δ has negative sign.[8] The last two rows of Table 2 show the OLS and IV estimates of this regression. Table 2 shows that, in spite of the IV estimates being less than three times larger than the OLS one, there is no coefficient of proportionality δ between -1 and -1000 that could rationalize the IV estimates.[9]

Table 2: Comparison of OLS and IV estimates

(1)	(2)
δ	Coefficient
-1	0.62030
-2	0.72640
-3	0.76790
-4	0.78880
-5	0.80130
-10	0.82630
-20	0.83870
-50	0.84620
-100	0.84870
-1,000	0.85100
Technique	Coefficient
OLS	0.42380
IV	0.98220

Notes: This table shows the bounds of the identified set using data from Acemoglu et al. (2001). In Acemoglu et al. (2001) the output of this regression is displayed in column (7) of Table 4. Column (1) shows either the value of the coefficient of proportionality δ or the technique used. Column (2) shows the value of the associated estimated coefficient. For each coefficient of proportionality δ the bounds of the interval are given by the OLS estimates and the coefficient estimated with that coefficient of proportionality δ.

There might be the concern that inclusion of controls could easily affect these results. To this extent, Table 3 considers the same regression model of Table 2 but where latitude

[8] This is a result using the IV and OLS estimates from Acemoglu et al. (2001), the two signs of the correlation discussed in this subsection and equation (5)

[9] Also in this case the sign of the coefficient of proportionality δ is consistent with equation (5). Yet, similar results hold even if δ is suspected to have positive sign. Tables are available upon request.

(i.e. a variable taking into account the distance from the equator scaled between 0 and 1) is included.[10] As with geographic binary variables, latitude is correlated with unobservables and positively correlated with the treatment variable.

Table 3 shows that inclusion of this control decreases slightly the OLS estimates and increases slightly the IV estimates. Yet, results are unchanged. There is no coefficient of proportionality δ between -1 and -1000 that could rationalise the IV coefficient.

Table 3: Comparison of OLS and IV estimates

(1)	(2)
δ	Coefficient
-1	0.62800
-2	0.72210
-3	0.75520
-4	0.77180
-5	0.78170
-10	0.80150
-20	0.81140
-50	0.81730
-100	0.81930
-1,000	0.82110
Technique	Coefficient
OLS	0.40130
IV	1.10710

Notes: This table shows the bounds of the identified set using data from Acemoglu et al. (2001). In Acemoglu et al. (2001) the output of this regression is displayed in column (8) of Table 4. Column (1) shows either the value of the coefficient of proportionality δ or the technique used. Column (2) shows the value of the associated estimated coefficient. For each coefficient of proportionality δ the bounds of the interval are given by the OLS estimates and the coefficient estimated with that coefficient of proportionality δ.

Lastly, Table 4 considers one of Acemoglu et al. (2001)'s most demanding specifications.[11] In this regression, the authors test the robustness of their results. Namely, they add a large set of controls to their regression model to check how their main coefficient changes. In this case both the OLS and the IV estimates decrease such as the identified sets of Tables 2 and 3 suggest. All the same, results do not change: no coefficient of proportionality δ between -1 and -1000 can rationalise the IV coefficient.

This analysis highlights that the IV estimates of these regressions were too large compared to the OLS estimates. This finding casts doubt on the IV estimation. However, the interpretation of these results is not straightforward, it requires prior knowledge of the setting we are analyzing. Specifically the researcher needs to determine whether the assumptions made to use this methodology are plausible. For example, it might be that

[10] This regression corresponds to column (8) of Table 4 in Acemoglu et al. (2001).

[11] The output of this regression is displayed in Table 6, column (9) of Acemoglu et al. (2001).

selection on observables is uninformative about selection on unobservables, yet, if this is the case there is no point in checking how the main coefficient changes after inclusion of controls. It is important to recall that, a further explanation that the methodology developed in this paper cannot discard, is that the effects are heterogeneous for the subpopulation affected by the instrument.

Table 4: Comparison of OLS and IV estimates

(1)	(2)
δ	Coefficient
-1	0.47760
-2	0.53820
-3	0.57150
-4	0.59130
-5	0.60420
-10	0.63200
-20	0.64690
-50	0.65610
-100	0.65920
-1,000	0.66200
Technique	Coefficient
OLS	0.37210
IV	0.71270

Notes: This table shows the bounds of the identified set using data from Acemoglu et al. (2001). In Acemoglu et al. (2001) the output of this regression is displayed in column (9) of Table 6. Column (1) shows either the value of the coefficient of proportionality δ or the technique used. Column (2) shows the value of the associated estimated coefficient. For each coefficient of proportionality δ the bounds of the interval are given by the OLS estimates and the coefficient estimated with that coefficient of proportionality δ.

3.3 Simulated data

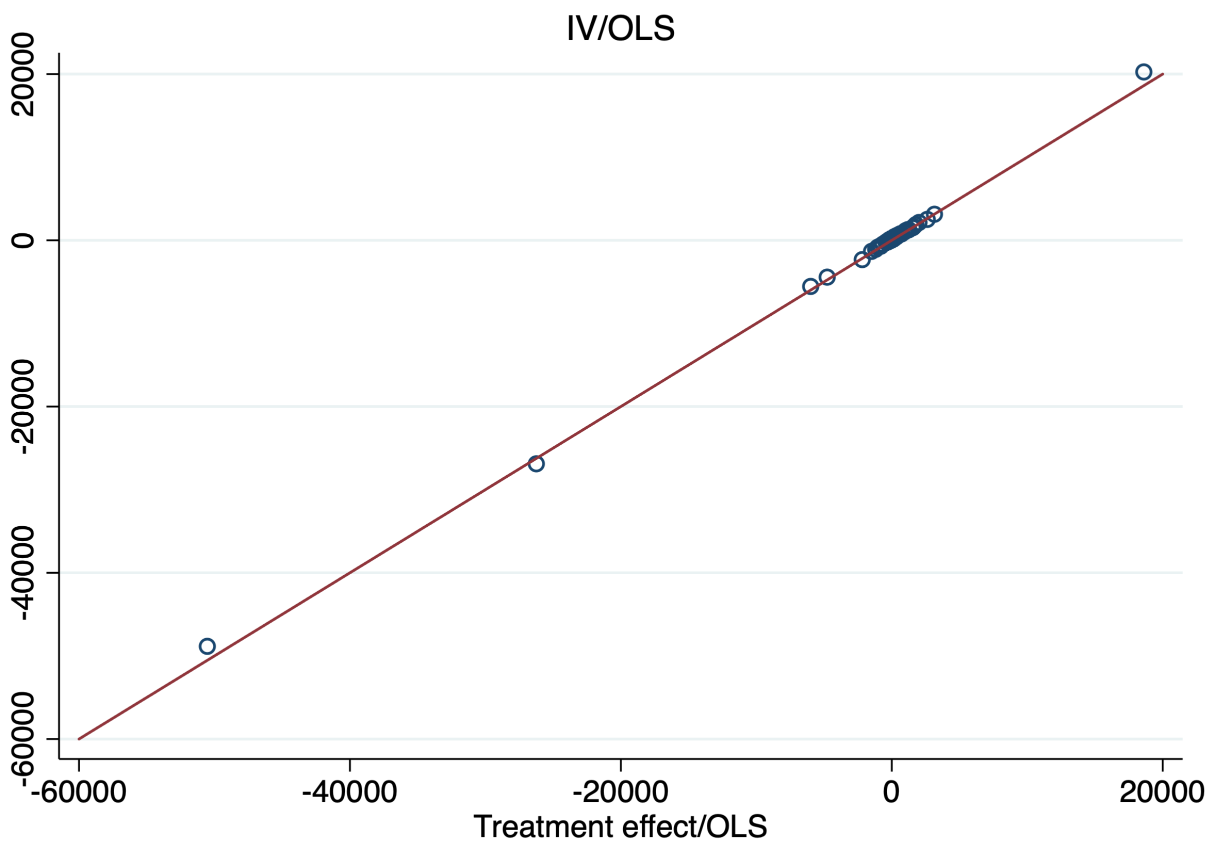
This section shows the results of simulating 10,000 samples with 10,000 observations each and specific features. In particular:

1. There is only one observable and one unobservable variable.
2. The unobservable is negatively correlated with the observable.
3. In absolute value the correlation between the unobservable and the treatment is always larger than the correlation between the observable and the treatment.
4. The PRF effect of the unobservable on the outcome is larger than the treatment effect.
5. The variance of the unobservable is larger than the one of the observable

The aim of this section is threefold. First, it investigates whether IV estimates are reliable even when they are considerably differente in size compared to the OLS ones. Second, it evaluates how accurately the procedure analyzed in this paper (i.e. in subsection "Implementation") computes δ. Third, it explores how the coefficients intervals estimated with the methodology developed in Oster (2017) perform in this IV setting.

Figure 1 shows the IV/OLS ratio is uninformative.[12] Such figure plots the ratio between the IV and OLS estimates on the vertical axis and the ratio between the Treatment effect and OLS estimates on the horizontal axis. The red line is the function given by the equality of the two axis. The figure shows data are distributed along this line, pointing that high values of the IV/OLS ratio are "justified" by high values of the Treatment effect/OLS ratio.

Figure 1: Comparison ratio IV/OLS vs Treatment effect/OLS



Notes: This figure plots the ratio between the IV and OLS estimates on the vertical axis and the ratio between the Treatment effect and OLS estimates on the horizontal axis. The red line is the function given by the equality of the two axis. Data are distributed on this line for any value of the ratios suggesting there is no difference between high and low values of the IV/OLS ratio.

[12]IV variables of the simulation fulfill the usual requirements (i.e. exogeneity, exclusion restriction and strong first stage).

Tables 5 and 6 respectively show descriptive statistics of δ and estimated δ (hereinafter, $\hat{\delta}$) for different values of $|\hat{\delta}|$. As it might be expected, comparing descriptive statistics suggests that the two are more likely to be similar if $|\hat{\delta}|$ is low. Put it differently, these results suggest that $\hat{\delta}$ is more likely to be informative about δ if $|\hat{\delta}|$ takes low values. To this extent, row 1 of Table 5 shows that out of 10,000 observations 3,593 have $|\hat{\delta}| < 1$, for these observations on average $\delta = 1.1$ while standard deviation, min and max are respectively 0.49; 0.18 and 2.45. Results of these last three statistics are stable as $|\hat{\delta}|$ increases. Likewise, comparing Tables 5 and 6 suggests that descriptive statistics between δ and $\hat{\delta}$ are similar for low values of $|\hat{\delta}|$.

There might be concerns about how likely is that the sign of the coefficient of proportionality δ is correctly estimated by $\hat{\delta}$. To this end, Table 7 shows the frequency and relative share of samples for which the sign of the coefficient of proportionality δ is correctly estimated (i.e. $sign(\hat{\delta}) = sign(\delta)$) as $\hat{\delta}$ increases. This table indicates, out of each category, for over 90% of the cases the sign of the coefficient is estimated correctly.

Table 5: Comparison of delta: estimated versus true (i.e. simulated)

Obs	Mean	Std. Dev.	Min	Max	$\|\hat{\delta}\| <$
3,593	1.1	0.49	0.18	2.45	1
6,753	1.06	0.49	0.18	2.45	2
8,198	1.05	0.49	0.18	2.45	3
8,945	1.04	0.49	0.18	2.45	4
9,295	1.03	0.49	0.18	2.45	5
9,762	1.01	0.49	0.18	2.45	10
10,000	1	0.49	0.18	2.45	100

Notes: This table shows descriptive statistics of δ for different values of $\hat{\delta}$.

Table 6: Comparison of delta: estimated versus true (i.e. simulated)

Obs	Mean	Std. Dev.	Min	Max	$\|\hat{\delta}\| <$
3,593	0.82	0.58	-1	1	1
6,753	1.37	0.73	-1	2	2
8,198	1.57	0.8	-1	3	3
8,945	1.73	0.93	-1	4	4
9,295	1.83	1.05	-1	5	5
9,762	2.06	1.48	-1	10	10
10,000	2.27	1.98	-1	10.9	100

Notes: This table shows descriptive statistics of $\hat{\delta}$ as it increases.

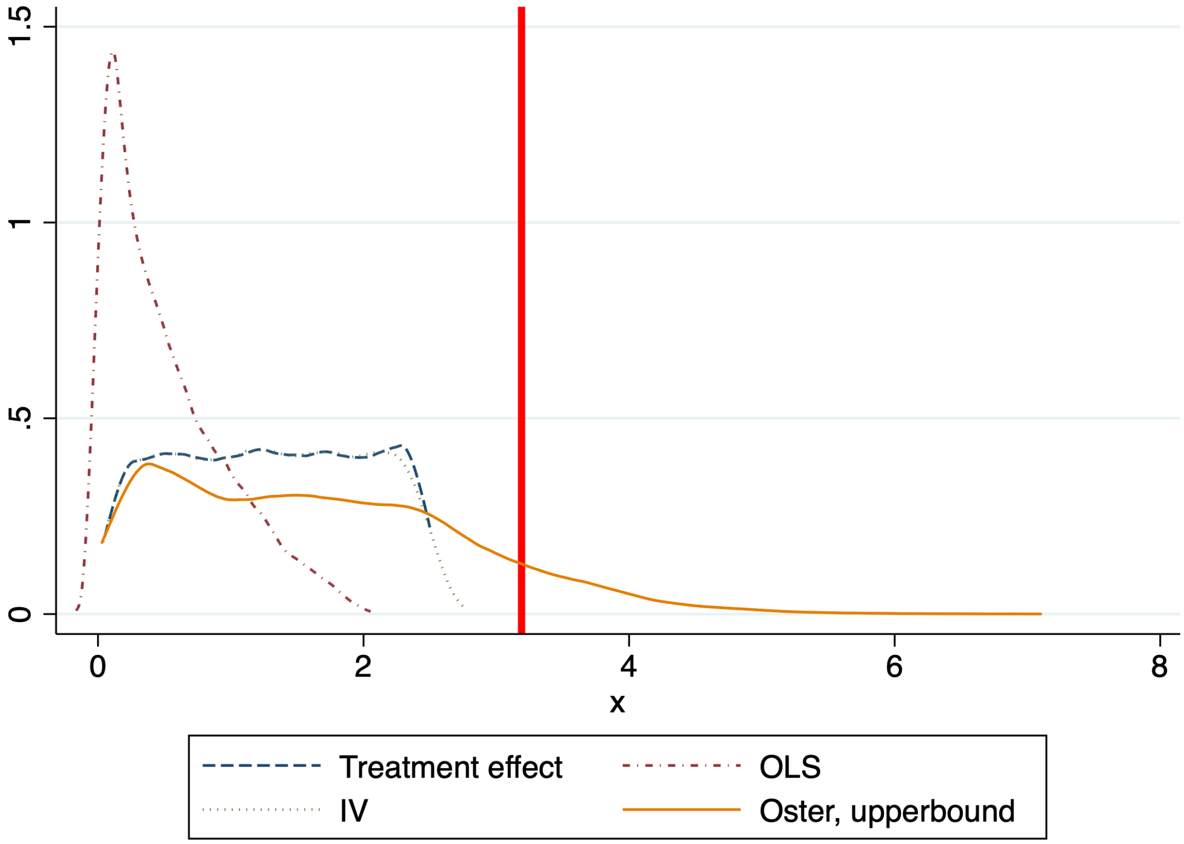
Table 7: Frequency

Frequency	Total	Percentage	$\|\hat{\delta}\| <$
3,261	3,593	0.91	1
6,421	6,753	0.95	2
7,866	8,198	0.96	3
8,613	8,945	0.96	4
8,963	9,295	0.96	5
9,430	9,762	0.97	10
9,668	10,000	0.97	100

Notes: This table shows the frequency and relative share of samples for which the sign of the coefficient of proportionality δ is correctly estimated (i.e. $sign(\hat{\delta}) = sign(\delta)$) as $\hat{\delta}$ increases.

Figure 2 plots the distribution of the treatment effect and the three estimators considered in this paper: OLS, IV and Oster for the samples in which the sign of the coefficient of proportionality δ is correctly estimated (i.e. $sign(\hat{\delta}) = sign(\delta)$), 9,668 samples out of 10,000. Given the structure of the samples the OLS and Oster estimators are respectively a lower- and upper- bound of the Treatment effect. The vertical line cuts the distribution of Oster estimators at the 90% percentile. This figure supports that in this setting the Oster estimator is a reasonable upper-bound of the Treatment effect. Indeed, the Oster estimator is closer to the Treatment effect, then the OLS estimator, for 90% of the samples considered.

Figure 2: Distribution of the treatment effect and estimators



Notes: This figure plots the distribution of the treatment effect and the three estimators considered in this paper: OLS, IV and Oster for the samples in which the sign of the coefficient of proportionality δ is correctly estimated (i.e. 9,668 samples out of 10,000). Given the structure of the samples the OLS and Oster estimators are respectively a lower- and upper- bound of the Treatment effect. This figure indicates in this setting the Oster estimator is more accurate than the OLS estimator. In particular, this is the case for 90% of the samples.

4 Stata code

This section discusses how to compute estimates as the ones in the Section 3 using STATA. Download Oster (2017) STATA command typing:

ssc install psacalc

Run the structural regression of interest. Afterwards, to compute the bound of the identified set for a certain value of δ type:

psacalc beta *treatment*, delta(δ)

where *treatment* is the name of the treatment variable and δ ranges from 1 (-1) to 1000 (-1000) depending on the sign suggested by the observational setting. Using the options: mcontrol (*unrelated controls*) and rmax($Rmax$) we can respectively add control variables, that the observational setting suggests are unrelated to unobservables, and set the maximum value of R. If no value is specified for $Rmax$, it is set to the default value of 1. In this way, psacalc computes the value of the β_{PRF}, the coefficient of the PRF, that would arise with a coefficient of proportionality of δ. By setting different values of δ we can build tables as those in Section 3.

To compute the exact value of δ leading to the IV estimate we replace the option delta(δ) with the option beta(β_{IV}) where β_{IV} is the value of the IV estimate. Specifically, we would type:

psacalc delta *treatment*, beta(β_{IV})

In this case, psacalc computes the value of the coefficient of proportionality δ that can explain that the coefficient of the PRF equals β_{IV}. However, it is worth noting that in this case the command assumes by default that δ is positive which might not make sense with the observational setting.

5 Concluding remarks

Considerable differences between the size of OLS and IV coefficients are likely seen as evidence against the validity of the instrument. The intuition supporting this viewpoint is that the OLS regression might be informative about the true effect the researcher wants to estimate. However, to my knowledge, in the literature there is no formal methodology to compare these two estimates.

To the best of my knowledge this is one of the first papers to suggest an objective criterion to compare IV and OLS estimates. For this purpose, this article adapts Oster (2017) setting to an IV framework. Furthermore, the analysis presented in this study disentangles in which setting this methodology might be used and suggests a way to implement it. Finally, this manuscript presents two observational examples and a simulated one to evaluate empirically its implementation.

This paper suggests that low values of the coefficient of proportionality offer supportive evidence that IV estimates are not *too large* with respect to OLS. The main limit of the methodology presented in this paper is that high values of such a coefficient might either cast doubts on instrument validity or merely highlight that effects are heterogeneous (Angrist and Imbens 1995; Masten and Poirier 2018). To tackle this issue, this study suggests to complement the usage of this methodology with different analyses depending on the specifical setting.

The analysis carried out in this paper indicates that further research on criteria to objectively compare IV and OLS estimates is needed. All in all, this paper sets the ground for an objeticve criterion to compare IV and OLS estimates.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American economic review 91*(5), 1369–1401.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Altonji, J. G., T. Conley, T. E. Elder, and C. R. Taber (2010). Methods for using selection on observed variables to address selection on unobserved variables.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005a). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human resources 40*(4), 791–821.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005b). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy 113*(1), 151–184.

Angrist, J. and G. Imbens (1995). Identification and estimation of local average treatment effects.

Angrist, J. D. and A. B. Krueger (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives 15*(4), 69–85.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020). Incarceration, recidivism, and employment. *Journal of Political Economy 128*(4), 1269–1324.

Ciacci, R. (2018). Banning the Purchase of Prostitution increases Rape: Evidence from Sweden.

De Luca, G., J. R. Magnus, and F. Peracchi (2019). Comments on Unobservable Selection and Coefficient Stability: Theory and Evidence and Poorly Measured Confounders are More Useful on the Left Than on the Right. *Journal of Business & Economic Statistics 37*(2), 217–222.

Farley, M. and H. Barkan (1998). Prostitution, violence, and posttraumatic stress disorder. *Women & health 27*(3), 37–49.

Farley, M., A. Cotton, J. Lynne, S. Zumbeck, F. Spiwak, M. E. Reyes, D. Alvarez, and U. Sezgin (2004). Prostitution and trafficking in nine countries: An update on violence and posttraumatic stress disorder. *Journal of trauma practice 2*(3-4), 33–74.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999). Geography and economic development. *International regional science review 22*(2), 179–232.

Hall, R. E. and C. I. Jones (1999). Why do some countries produce so much more output per worker than others? *The quarterly journal of economics 114*(1), 83–116.

Liu, S. (2020). Incarceration of african american men and the impacts on women and children. *Available at SSRN 3601259*.

Masten, M. A. and A. Poirier (2018). Salvaging falsified instrumental variable models. *arXiv preprint arXiv:1812.11598*.

Mejia, D. and P. Restrepo (2013). Bushes and bullets: illegal cocaine markets and violence in colombia. *Documento CEDE* (2013-53).

Murphy, K. M. and R. H. Topel (1990). Efficiency wages reconsidered: Theory and evidence. In *Advances in the Theory and Measurement of Unemployment*, pp. 204–240. Springer.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 1–18.

Appendix

A Multivariate case

Let the PRF be equation (1). However, given the nature of w we can only run equation (2).

Denote $\tilde{d}$ as the residual of a regression of d on X (namely, $d_{ih} = \tau_0 + \tau_1 X_{ih} + \tilde{d}_{ih}$). Then:

$$\hat{\beta}_2 = \beta_1 + \gamma \frac{Cov(d_{ih}, w_{ih}) - \tau_1 Cov(X_{ih}, w_{ih})}{Var(\tilde{d}_{ih})} \tag{A.1}$$

We can rearrange equation (A.1) using equation (3), assuming observables and unobservables are orthogonal (as in Oster (2017)), it is straightforward to get:

$$\delta = (\hat{\beta}_2 - \beta_1) \frac{Var(\tilde{d}_{ih}) Var(X_{ih})}{\gamma Var(w_{ih}) Cov(d_{ih}, X_{ih})} \tag{A.2}$$

which is similar to equation (5) of the univariate case.