

# **Let's Call their Bluff: The Politics of Econometric Methodology**

by

**Jean-Paul Azam<sup>1</sup>**

## **1 Introduction**

Some regressions can be highly misleading. In particular non-causal regressions can put you in trouble. For example, the literature on “aid-ineffectiveness” is littered with papers whose wrong conclusions deeply influenced policy-makers. The classic one is Boone (1996), who showed that foreign aid seems to have no impact on economic growth, without controlling for endogeneity. This finding was endorsed by many academic authorities and several books got published to develop its policy implications. Arndt et al. (2017) overturned this result by showing that foreign aid has a beneficial impact on economic growth and many other outcomes, after controlling for endogeneity in a simple two-stage approach. But donors have been misled for twenty years. The next section illustrates this issue by reference to a paper by Azam & Berlinschi (2010) that brings out that foreign aid is probably disbursed for reducing the inflow of immigrants from low- and lower-middle income countries, although a superficial look at the data suggests the opposite. In this instance, proper instrumentation is crucial to dispel the wrong impression given by descriptive statistics.

The present short paper aims at offering a meaningful framework for understanding how instrumental variables may be used to evaluate policy effectiveness using historical data. It brings out the strength of an identification strategy aimed at revealing the unobservable preferences of the policy makers, based on preference proxies. This is a crucial step to make, as policy makers are not necessarily very proud of their true motivations and are often keen to

---

<sup>1</sup> Toulouse School of Economics, UT1- Capitole. This paper was presented as the NEPS Lecture at the International Institute for Social Studies, The Hague, June 25, 2019. Comments by participants are gratefully acknowledged, especially by Corine Bara, Raul Caruso, Annekatrin Deglow, Mario Ferrero, Jean-Pierre Marchand, Mansoob Murshed, Andrea Ruggeri and Marijke Verpoten. All remaining error's are the author's responsibility. Jean-Paul Azam acknowledges funding from ANR under grant ANR-17-EUR-0010 (Investissement d'Avenir program).

use smoke screens to hide them. It also shows the limitations of this method, due to the measurement error entailed by such a use of proxies as control functions. However, it emphasizes that the resulting attenuation bias is playing against the econometrician in such a way that it reinforces the confidence that one can put in the findings when the control function is actually significant. The strength of this approach is then illustrated by reference to a study of the Naxalite conflict in India showing that the tribal insurgents cannot be accused of having the initiative in the armed violence, despite the Federal Government's claims to the opposite. In fact, the state forces, including local police and occasional militia, have at least as much responsibility as the tribal people for the slaughter.

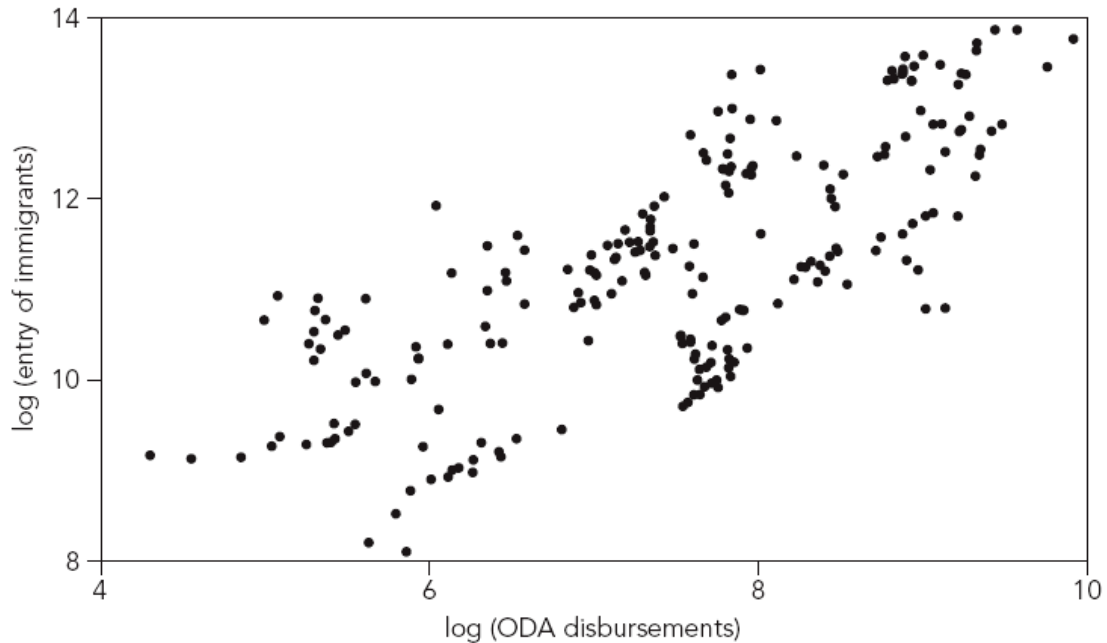
## **2 A Case for Econometric Citizen Oversight**

The key point in evaluating policy effectiveness from historical data is to control for the policy maker's endogenous responses. This requires that the econometrician seeks to uncover the policy maker's true preferences, which might be quite at variance with the proclaimed ones. For example, Azam & Berlinschi (2010) have thus shown that foreign aid is in fact allocated by OECD countries with a view to reduce immigration from low and lower-middle countries. When these findings were first presented at Nuffield College in June 2009, Adrian Wood, who had worked a long time at DFID, came at the coffee break to tell me: "Jean-Paul, it is true that the Foreign Office exerted a lot of pressure on our aid allocation with a view to reduce immigration, but we don't do it anymore". He obviously crossed his heart with his hand in saying that, but I am not sure he expected me to believe him.

### **2.1 Do we Buy Immigrants?**

Are OECD governments so keen to attract immigrants from poor countries that they bribe their governments to send more of them? Eyeballing figure 1 might suggest so, as we observe a positive correlation between the number of migrants entering OECD countries and the Official Development Assistance (ODA) disbursed by the host countries in this scatter diagram. Azam & Berlinschi (2010) have tackled this empirical puzzle and shown that this impression is undoubtedly misleading. Table 1 displays their most important findings. Column 1 shows the findings of a regression analysis that does not control for the endogeneity of foreign aid disbursements in response to possible surges in immigration flows from lower- and lower middle income countries. It shows that adding control variables helps to mitigate the misleading impression given by figure 1's scatter diagram. Adding the (lagged)

unemployment rate and the (lagged) share of social expenditures in GDP reduces the upward bias shown by the chart to insignificance. However, it leaves unexplained why rich countries are giving so much aid money to migrants' source countries.



**Source:** Azam & Berlinschi, 2010.

**Figure 1: The Positive Correlation between Foreign Aid and Immigration**

## 2.2 Revealing Donors' Hidden Agenda

Controlling for endogeneity solves this problem by showing that OECD countries give aid money *because* it helps to reduce immigration from low and lower middle income countries, among other things. This is shown at column 2 by the fact that (i) the negative impact of foreign aid disbursements has a highly significant impact on the flow of immigrants facing the donors, and (ii) the estimated endogeneity bias is also highly significant according to the Nakamura & Nakamura (1981) version of the Hausman (1978) test (more on this below). This is performed by introducing in the equation beside the endogenous regressor, here ODA, the residuals of the first-stage equation explaining the latter presented at column 3. The latter includes all the same controls as the two previous equations and public expenditures on order and security as the instrumental variable. Hence, right-wing governments, as revealed by their levels of expenditures on law and order, disburse more foreign aid to control immigration from poor countries than the others.

**Table 1: Flows of Legal Migrants from Low- and Lower-Middle Income Countries**

<b>Dependent Variable: Log of Number of Legal Migrants</b>	<b>Uncontrolled Endogeneity</b>	<b>Controlled Endogeneity</b>	<b>First-Stage Equation</b>
<b>Log of Official Development Assistance (ODA) Disbursement</b>	0.46 (0.32)	-3.68*** (1.15)	-
<b>Endogeneity Bias ODA</b>	-	4.47*** (1.26)	-
<b>Unemployment Rate</b>	-0.30*** (0.09)	-0.18*** (0.08)	0.03 (0.02)
<b>Social Expenditures as a Percentage of GDP</b>	0.32*** (0.09)	0.30*** (0.09)	-0.009 (0.013)
<b>Log of per capita GDP</b>	0.54 (1.42)	9.53*** (2.82)	1.87*** (0.36)
<b>Log of Stock of Foreign Population</b>	0.19 (0.57)	0.57 (0.50)	0.04 (0.08)
<b>Log of Public Expenditures on Order and Security</b>			0.21*** (0.07)
<b>Number of Observations</b>	118	117	159
<b>F-test</b>	9.50	9.84	48.59

**Source:** Azam and Berlinschi (2010)

It is thus clear that foreign aid is effective at something after all, but donors are not too proud to say what it is good for. It is effective for controlling immigration and it is actually used for that. I now have the strong suspicion that the (French) socialists prefer keeping fiscal resources for their own public expenditures and they raise the minimum wage and payroll and profit taxes, and hence unemployment, maybe to deter immigration. However, we did not think about performing such a test in 2009, when we researched this paper with Ruxanda Berlinschi. This is a pity as it would have been so easy to do then. Nevertheless this would explain the massive increases in payroll and corporate taxes implemented by Prime Minister Jean-Marc Ayraud from 2012 on under the aegis of President François Hollande, resulting in a predictable three percentage point increase in unemployment, at the time when the Libyan crisis was in full gear.

### 3 The Case for Political Cliometrics

Econometric investigation of historical data is basically a game between civic-minded econometricians and policy makers whose deep motivations are potentially hidden. It is a basic tenet of economic theory that preferences can never be observed directly but must be inferred instead by analyzing observable behaviors using revealed preference theory. Econometricians have devised various two-stage methods that help us to discover some

determinants of policy-makers' preferences, and thus can play a key part in informing citizens' oversight.

### 3.1 The Setting

The econometrician wants to test whether a policy tool  $p$  is effective for reducing the quantity  $q$  of an outcome that the policy-maker deems "bad". These two variables are linked by a linear relation:

$$q = \alpha + \beta x - \gamma p + \delta e + \varepsilon, \quad (1)$$

where  $x, e$  and  $\varepsilon$  are exogenous variables. The Greek parameters, including  $\{\alpha, \beta, \gamma, \delta\}$ , are positive unless specified otherwise, while  $\varepsilon$  is a random disturbance such that  $E(\varepsilon) = 0$ .

We make the following key assumptions:

**Assumption 1:** Asymmetric information:

- (i) The policy maker observes  $x, e$  and  $p$  *ex ante* and  $q$  *ex post*.
- (ii) The econometrician observes  $q, x$  and  $p$  *ex post*.

**Assumption 2:** Efficient Information Processing:

The policy maker uses her information efficiently so that  $E(e\varepsilon) = 0$ .

**Assumption 3:** Quadratic Loss Function:

The policy maker seeks to minimize the following loss function:

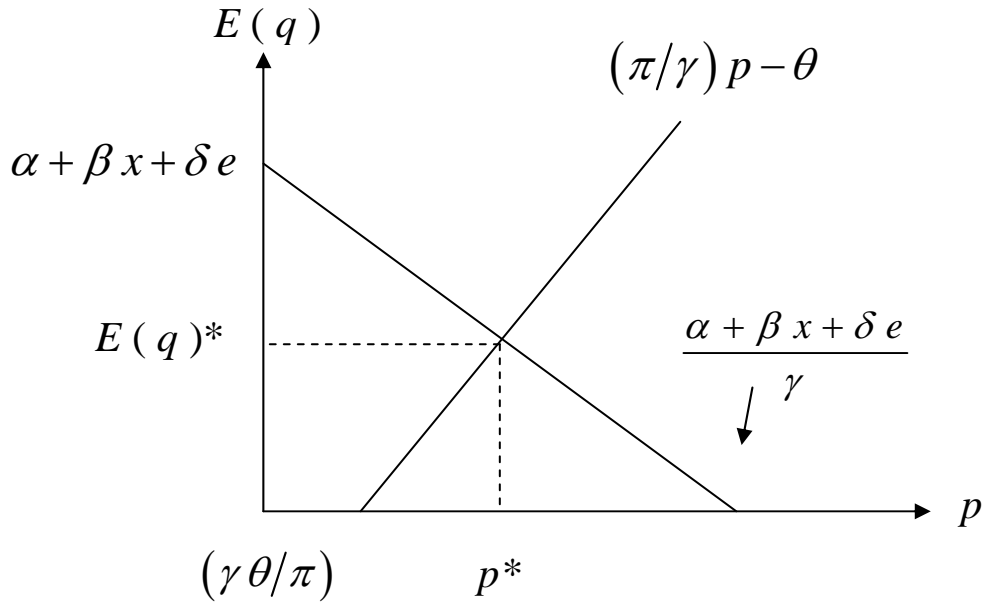
$$\min_p L = \left(\frac{\pi}{2}\right)p^2 + \frac{1}{2}(\theta + E(q))^2, \text{ s.t. } E(q) = \alpha + \beta x - \gamma p + \delta e. \quad (2)$$

The parameter  $\theta$  increases the policy maker's aversion to  $q$  and it is her private information, i.e., unobserved by the econometrician.

The policy decision is derived from this model by solving (2) to yield the first-order condition:

$$\pi p^* = \gamma(\theta + E(q)^*) \Leftrightarrow E(q)^* = \left(\frac{\pi}{\gamma}\right)p^* - \theta. \quad (3)$$

Figure 2 depicts the outcome of this optimization exercise performed by the policy maker. The downward-sloping line represents the causal equation taken as the constraint by the policy maker according to (2), while the upward-sloping line represents the first-order condition (3). Her choice is found at the intersection of the two lines. It is then obvious from the diagram that the chosen value of the policy variable  $p^*$  is a function first of the set of parameters  $\{\alpha, \beta, \gamma, \delta, \pi\}$  but also and more importantly of the pair of unobservable variables  $\{\theta, e\}$ .



**Figure 2: The Policy Maker's Optimal Choice**

The presence of this pair of unobservable variables lies at the heart of the identification problem faced by the econometrician. The key point is that they impact the econometrician's problem in two opposite directions. Changes in  $\theta$  across observations are playing on the econometrician's side, as they just shift the upward-sloping line, to the right in case of an increase in  $\theta$ , tracing out the causal relations under scrutiny. However, changes in  $e$  across observations are playing against the econometrician, as an increase in  $e$  would shift the downward-sloping line upwards, tracing out the upward-sloping first-order condition line. This would thus contaminate the estimation of the causal relation. Imagine now that the two unobservable variables  $\theta$  and  $e$  turned out to be positively correlated in the econometrician's sample, although they might in fact be independent in theory and would be

so in a very large sample. Then, the two lines would tend on average to move in the same direction and an upward-sloping regression line might result from the estimation process if the variance of  $e$  was larger than that of  $\theta$ . In such a case, regression analysis would tell you more about the fortuitous correlation that exists in your sample between the two unobserved variables than about the true slope of the causal relation that you are looking for.

### 3.2 Preference Proxies

Imagine now that, for whatever reason, the econometrician is tempted to guess that the preference proxy  $z$  such that:

$$\theta = \rho + \mu z + \varsigma \text{ and } E(\varsigma) = E(\varsigma e) = E(\varsigma \varepsilon) = E(\varsigma z) = 0 \quad (4)$$

might have some empirical relevance. This is a testable piece of guesswork using a two-stage approach to overcome the problem that  $\theta$  is not directly observable. The idea is to capture some of the beneficial identifying property of the changes in the policy maker's preferences over the sample thanks to the preference proxy  $z$  using a first-stage equation that explains the policy decisions actually made. Then, most of that equation's unobservable disturbance will reveal the changes in  $e$  rather than in  $\theta$ . Hence, that remaining unobservable disturbance will encapsulate some information about the policy-maker's information about  $e$  that can in fact be estimated via the first-stage equation. This is done by estimating a reduced-form equation for the chosen policy tool, which thus reveals the hidden information used by the policy maker as the residuals. This works as follows:

Substituting the preference proxy for  $\theta$ , the FOC becomes:

$$E(q) = \left( \frac{\pi}{\gamma} \right) p - \rho - \mu z - \varsigma. \quad (5)$$

We can substitute for  $E(q) = \alpha + \beta x - \gamma p + \delta e$  and rearrange the terms to get the reduced-form policy equation as:

$$p^* = \underbrace{\frac{\gamma(\alpha + \rho)}{\pi + \gamma^2}}_a + \underbrace{\frac{\gamma\beta}{\pi + \gamma^2}}_b x + \underbrace{\frac{\gamma\mu}{\pi + \gamma^2}}_c z + \underbrace{\frac{\gamma}{\pi + \gamma^2}(\delta e + \varsigma)}_f \quad (6)$$

where  $\hat{r}$  are the residuals of the regression of  $p$  on  $x$  and  $z$ . Notice that (i)  $\hat{r}$  is correlated with  $e$ , but (ii) it contains some noise produced by  $\varsigma$ , whose variance is lower the better  $z$  does proxy for  $\theta$ . Then,  $\hat{r}$  can be inserted in the second-stage equation as a control function, i.e., a noisy proxy for  $e$ .

Then, it is straightforward to show that the second stage regression equation yields under the control-function approach exactly the same estimates as 2SLS would in linear models:

$$q = \hat{\alpha} + \hat{\beta}x - \hat{\gamma}p + \hat{\phi}\hat{r} + \hat{\varepsilon} = \hat{\alpha} + \hat{\beta}x - \hat{\gamma}(p - \hat{r}) + (\hat{\phi} - \hat{\gamma})\hat{r} + \hat{\varepsilon}, \quad (7)$$

where  $p - \hat{r}$  is the fitted value of  $p$  from the first-stage equation, what 2SLS uses as the regressor, and  $\hat{r}$  can be dropped from this latter regression equation, as it is orthogonal to all the included regressors, by construction. This proves the exact equivalence between the two methods of controlling for endogeneity in linear models. However, the control function approach yields the natural test for endogeneity that  $\hat{\phi}$  is significantly different from zero (see Nakamura & Nakamura, 1981), and it can be used in many other contexts where 2SLS are inappropriate. In particular, it works when the assumed distribution is a negative binomial, the most popular assumption when analyzing count data. Notice that  $\hat{r}$  is a noisy proxy for  $e$ , as explained above. This entails that there is a measurement issue that will bias  $\hat{\phi}$  towards zero, i.e., against the endogeneity assumption. Hence, this attenuation bias will play against the econometrician and it will often reject the endogeneity diagnosis while it should be validated. In contrast, when the endogeneity test concludes significantly by rejecting exogeneity, then the econometrician's confidence in the endogeneity diagnosis is obviously strengthened by the bias.

#### 4 An Example of Political Interpretation

The residuals from the first-stage equation are a proxy for some information that (i) is observed by the policy maker when she makes her decision, (ii) is entering the causal relation, and (iii) is not directly observed by the econometrician. Hence, although it is biased towards zero, the control-function approach is providing the econometrician with a crucial piece of information about the information used by the policy maker to choose her policy intervention, if it is statistically significant. Azam & Bhatia (2017) have used this argument to



deny the claim that the rebels are the aggressors in India's Naxalite conflict. Table 2 reports their basic findings. The dependent variables are respectively the number of civilians killed by the police (1) and the number of people killed by the rebels (2).

**Table 2: Two-Stage Estimation of Numbers of Persons Killed by Each Side**

<b>Variables</b>	<b>(1) State Forces</b>	<b>(2) Rebellion</b>
<b>State Forces</b>		0.75** (0.0033)
<b>Residuals</b>		0.025* (0.015)
<b>Iron Resources</b>	3.239 (3.409)	1.285*** (0.404)
<b>Coal Resources</b>	9.453** (4.470)	0.877** (0.361)
<b>Tribal Population</b>	-0.492*** (0.142)	-
<b>Forest Cover Rate</b>	-0.137* (0.081)	-
<b>Tribal * Iron</b>	0.314** (0.129)	-
<b>Tribal * Coal</b>	-0.105 (0.192)	-
<b>Tribal * Forest</b>	0.020*** (0.003)	-
<b>Other Controls</b>	Yes	Yes
<b>Ln alpha</b>		1.087*** (0.135)
<b>Observations</b>	191	191
<b>Adjusted <math>R^2</math></b>	0.3848	
<b>F-Statistics (prob &gt; F)</b>	10.90 (0.000)	

**Source:** Azam & Bhatia (2017). Column (1) is estimated by OLS while (2) is a negative binomial regression.

#### 4.1 Basic Findings

We find here that the residuals of the first-stage equation are significant in the second-stage one. This reveals that there is a piece of information that is affecting simultaneously the killing performed by the two sides. This measure of the impact on the number of people killed by the local forces is just estimated as that part of the police violence that cannot be explained by the control variables and the instruments introduced in the first-stage equation, as explained above. On the one hand, this might simply capture some orders given by the hierarchy to the policemen or militiamen, when relevant, unobserved by the econometrician.

On the other hand, that information should have leaked to the rebels who got organized very quickly to retaliate. This tends to strain credulity given the rebels' known level of organization (Roy, 2011, Shah, 2010). More realistically, this might simply capture the fact that the rebels have observed the numbers of tribal people killed in the districts and have apportioned their retaliation to those numbers. It thus seems that the rebels know how much violence has been inflicted on the tribal people when they attack the policemen, what would not make sense if they really were the aggressors as claimed by the Federal and State governments. The significance of the positive impacts on the rebellion's activity of the presence of iron or coal resources hints at the part played by mining interests in triggering the uprising.

## 4.2 Political Diagnosis

Then, we find that the observed number of people killed by the local forces has a highly significant positive impact on the killing perpetrated by the rebels, an estimate that is "purged" of the endogenous reverse causality by the control for endogeneity. Hence, police killing of civilians is a major cause of the Naxalite uprising. This disproves the federal government's claim that the Adivasis are animated by the Maoist ideology with a view to topple the Indian democracy. The positive impact of the killing performed by the local forces deserves a bit more emphasis, by comparison with the theoretical framework discussed in the previous section. There, a negative sign was specified for the impact of the policy variable  $p$  to capture the idea that the outcome variable  $q$  was deemed bad by the policy maker. Azam & Bhatia (2017) introduce the concept of "provocation" to capture the idea that the rebels' violence does not seem to be regarded as a "bad" in that sense by the policy maker. Moreover, a closer look at the first-stage regression allows us to look deeper into the fundamental motivations for these attacks, via the preference proxies that are used as instrumental variables.

We observe first that the presence of a large Adivasis population and a large forest cover rate, where the former live most often, come up with a significant negative sign. Hence, the police seem to stay away from such poor areas. However, most of the killing of civilians performed by the state police occurs in areas where a large share of the people are tribal ones *and* there are significant deposits of iron ore and a significant forest cover. The latter mainly exists where bauxite deposits retain the mountains' moisture in these semi-arid areas (Padel & Das, 2010). Hence, it is the interaction between the presence of mining potential riches and

the largely tribal population of the area that entails significant violence against civilians by the police, although the presence of coal resources seems to attract police violence independently of the importance of the tribal population in the district.

#### 4 Conclusion

Carefully chosen preference proxies give the civic-minded econometrician a valuable tool to uncover the policy-makers' deep motivations that determine their decisions. Using two-stage econometric analysis of real-world data can thus play a key part in informing citizens' oversight in democracies and authoritarian regimes. The objective pursued is to estimate simultaneously whether the policy tool has a significant impact on the outcome of interest and is actually used for that by the policy maker.

#### References

- Arndt, C., Jones, S., Tarp, F. (2015): "Assessing Foreign Aid's Long Run Contribution to Growth and Development", *World Development* 69, 6-18.
- Azam, Jean-Paul, & Kartika Bhatia (2017): "Provoking insurgency in a federal state: theory and application to India", *Public Choice*, 170, 183-210
- Azam, Jean-Paul, & Ruxanda Berlinschi (2010): "The Aid-Migration Trade-Off", in Justin Yifu Lin and Boris Pleskovic (Eds.): *Annual World Bank Conference on Development Economics 2009, Global: People, Politics, and Globalization*, 147-171, World Bank: Washington, D.C..
- Boone, Peter (1996): "Politics and the Effectiveness of Foreign Aid", *European Economic Review*, 40, 289-329.
- Hausman, Jerry A. (1978): "Specification Tests in Econometrics", *Econometrica*, 46, 1251-1272.
- Nakamura, Alice, & Masao Nakamura (1981): "On the Relationships among Several Specification Error Tests Presented by Durbin, Wu and Hausman", *Econometrica*, 49, 1583-1588.
- Padel, Felix, and Samarendra Das (2010): *Out of this Earth. East India Adivasis and the Aluminium Cartel*, Orient Blackswan: New Delhi.
- Roy, Arundhati (2011): *Broken Republic*, Penguin Books: London.
- Shah, Alpa (2010): *In the Shadows of the State*, Duke University Press: Durham and London.