Soil consumption, organized crime, and WALS

Franco Peracchi
EIEF and UTV

UTV, January 19, 2024

Introduction

- In a sequence of papers with Giuseppe De Luca and Jan Magnus (De Luca, Peracchi & Magnus, 2018, 2019, 2021, 2022, 2023, 2024), I have addressed some general issues concerning covariate selection in linear models, inference after model selection, model averaging, and the combination of Bayesian and classical (frequentist) ideas.
- In more recent work with Cinzia Di Novi and Alessandro Flamini (Di Novi, Flamini & Peracchi 2023), I have also investigated the effect of organized crime on soil consumption, an important (especially for Italy) but understudied research area. This work has given me the opportunity to think about econometric practice as currently conducted.
- In this talk, I will try to bring together these two separate lines of research and to draw some conclusions about empirical work, which I hope you may find of interest.

Soil consumption and organized crime

- Soil is a non-renewable natural resource that provides a number of ecosystem services essential for life (carbon sequestration, water maintenance, food production, etc.).
- The "consumption" of soil, or "soil consumption", is the conversion of land with healthy soil and intact habitats into areas for industrial agriculture, traffic (road building) and especially urban human settlements (Wikipedia).
- Among the Italian regions, Apulia stands out because its soil consumption per capita between 2006 and 2021 is about 12.5%, the second largest in the country and more than three times the Italian average of 4% (SNPA, 2022).
- De Feo & De Luca (2017) provide indirect evidence of the effect of organized crime on soil consumption at the municipal level in Sicily, where "[p]ublic authorities may allow wilder urban expansion, overriding existing regulations, or obscurely award public contracts to mafia-related entrepreneurs to reward the mafia's electoral support".
- Using data from Apulia, I will provide direct evidence of this effect.

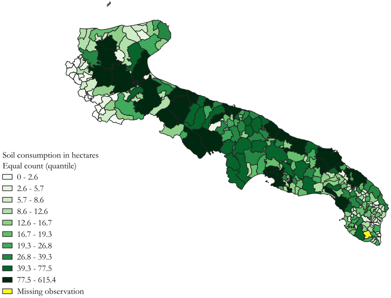
Mafia expansion in Apulia

- Apulia represents a peculiar success story of mafia transplantation.
- The presence of organized crime in Apulia is a phenomenon that only emerged in recent decades, as the Apulian *Sacra Corona Unita*–the fourth Italian mafia, after *Cosa Nostra* (Sicily), *Camorra* (Campania) and *'Ndrangheta* (Calabria)– only started operating between the late 1970s and the early 1980s.
- Three reasons help explain this phenomenon:
  - Close proximity of Apulia to Campania and Calabria, two regions characterized by the presence of long-established mafia-type organizations (Pinotti, 2015).
  - Shifting interest of organized crime away from the "Tyrrhenian route" for drug/tobacco smuggling (from Morocco to Marseille, Naples, and Sicily) towards the "Adriatic route" (from the Balkans to the Eastern-Adriatic ports to Apulia).
  - Process of mafia diffusion through the forced resettlement program (*soggiorno obbligato*) introduced in 1956, which required convicted/ suspected mafia bosses to take up residence in municipalities sufficiently far from the mafia's traditional areas of operation.

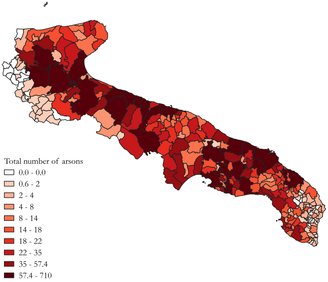
Data

- My units of observation are Apulia's 258 municipalities.
- The data have been constructed by merging several data sources containing the following variables, all measured at the municipality level:
  - soil consumption, 2006–2018 [Source: National System for Environmental Protection Database], the outcome of interest (Figure 1a);
  - number of arsons (*incendi dolosi*), 2004–2014 [Source: National Fire and Rescue Service Database], a proxy for the local strength of organized crime (Figure 1b);
  - physical, demographic, and socio-economic variables [Sources: Istat Population Census Database; Istat Municipal Database; Rete Urbana delle Rappresentanze, 2003], as exogenous or control variables.
- Excluding municipalities with missing or zero values for soil consumption or the number of arsons gives a working sample of $n = 231$ units.

Figure 1: Soil consumption and number of arsons in Apulia

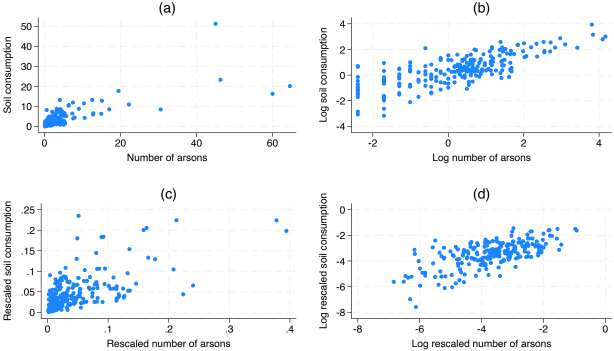


(a) Soil consumption, 2006–2018



(b) Number of arsons, 2004–2014

Figure 2: Soil consumption vs. number of arsons in Apulia



Notes: Municipal surface area as scaling variable.

The causal relation of interest

- Let C_i denote annual soil consumption in municipality $i = 1, \ldots, n$ during 2006–2018, let A_i denote the annual number of arsons during 2004–2014 (our proxy for the strength of organized crime), and let S_i be some scaling variable (e.g. municipal surface area, population size, or population density).

- I assume there exist smooth functions g and h such that

$$g(C_i/S_i) = \beta_0 + \beta_1 h(A_i/S_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where ϵ_i is an unobservable error term capturing the effect of all other variables that affect soil consumption.

- Letting $g(u) = h(u) = \ln u$ (as suggested by Figure 2) and $\epsilon_i = \beta_3' X_i + e_i$ for some vector X_i of observable controls gives the "structural equation" (the causal relation of interest)

$$c_i = \beta_0 + \beta_1 a_i + \beta_2 s_i + \beta_3' X_i + e_i, \quad i = 1, \ldots, n,$$

where $c_i = \ln C_i$, $a_i = \ln A_i$, $s_i = \ln S_i$, and the elasticity β_1 is the "focus parameter". All other parameters are treated as "nuisance parameters" as they do not generally have a causal interpretation.

Two approaches to estimating β_1

- The "conditional mean independence" approach:
  - Assume a_i is uncorrelated with e_i after conditioning on s_i and X_i, and estimate β_1 by OLS–a simple method with nice finite-sample properties if this assumption is correct.
  - The typical question with this approach is: What controls should be included in X_i to avoid omitted variables bias?
- The "instrumental variables" approach:
  - Allow a_i to be endogenous, i.e. correlated with e_i even after conditioning on s_i and X_i–either because a_i is an imperfect proxy for the presence of organized crime, or because of the omission of unobservable determinants of soil consumption that are correlated with a_i (e.g. the degree of moral integrity of municipal administrators/politicians).
  - Assume there exists some valid instrument z_i (i.e. uncorrelated with e_i but correlated with a_i after conditioning on s_i and X_i) and estimate β_1 by IV methods.
  - With this approach, an additional question arises: What instrument(s)?

The conditional mean independence approach

- The choice of controls should reflect expert knowledge about the problem.
- The set of controls is clearly not unique, and not all controls are equally useful.
- For a given sample size, the "best" set of controls should provide a "balance" between bias (if the controls are "not enough") and sampling variability (if they are "too many").
- This "bias-variance trade-off" is not a mechanical relationship: An additional control cannot decrease sampling variability, but does not guarantee that the bias goes down if the model remains "underspecified" (De Luca, Magnus & Peracchi, 2018), a point also made by Clarke (2005, 2009) and Pearl (2011).
- As shown by De Luca, Magnus & Peracchi (2019), unless strong assumptions are satisfied, this fact invalidates popular procedures, such as those proposed by Altonji, Elder & Taber (2005) and more recently by Oster (2019).
- In my case, the expert knowledge is quite limited. As a result:
  - there is little *a priori* knowledge of which controls to include in X_i;
  - even the few controls that one might include on *a priori* ground (e.g. "municipal income") are typically only defined in broad terms, so there is also uncertainty about their precise definition.

Free and doubtful controls

- Following Leamer (1985), I distinguish between controls that are “free” (always included in X_i) and “doubtful” (of which I am less certain).
- The free controls are binary indicators for provincial capital and coastal municipality.
- The doubtful controls include socio-economic variables:
  - indicators for educational attainments (fraction aged 15–19 with a secondary degree, fraction aged 30–34 with a university degree, fraction aged 6+ with a high school or bachelor degree, fraction aged 6+ with a primary degree or less);
  - indicators for local labor market conditions (unemployment rate, youth unemployment rate, youth employment rate);
  - composition of employment by industry or occupation (employment share in construction, in manufacturing, in the services, in non-retail services, fraction self-employed);
  - income (GDP per employee).
- They also include demographic variables: population growth and indicators for the age structure (aging index, fraction of children aged 0–5).
- I consider 16 doubtful controls in total.

Standard model selection

- Since my main source of model uncertainty is which of the $k = 16$ doubtful controls to include in X_i, my "model space" consists of the set $\mathcal{M}$ of all linear models for c_i containing the constant term, the variables a_i and s_i, binary indicators for provincial capital and coastal municipality as the "free regressors" (those always present), and any subset of the doubtful controls.
- Thus, $\mathcal{M}$ contains $J = \sum_{j=0}^{k} \binom{k}{j} = 2^k$ models, including the model with no doubtful controls (the "null model", indexed as $j = 1$) and the model with all of them (the "full model", indexed as $j = J$).
- I henceforth denote by $\hat{\beta}_{1j}$ the OLS estimate of β_1 from the jth model in $\mathcal{M}$.
- The typical approach to model uncertainty ("standard model selection") consists of two steps, both using the same data:
  - choose a subset $\mathcal{M}_0 \subseteq \mathcal{M}$ of models to explore, find the "best" model in $\mathcal{M}_0$, then estimate β_1 by the OLS estimate from the best model;
  - carry out inference about β_1 ignoring the data-driven first step.
- Standard model selection comes in two flavors, depending on whether the exploration of $\mathcal{M}$ is carried out informally or formally.

Informal explorations

- In this case, $\mathcal{M}_0$ typically contains a small number of models from $\mathcal{M}$, often cherry-picked after looking at the data.
- Table 1 illustrates this approach by presenting the OLS estimates of β_1 for the seven models (LS1–LS7) chosen by De Novi, Flamini & Peracchi (2023) under the constraint that $k \leq 6$.
- The table is an example of the tabular presentation of regression results that is now standard.
- Ignoring the exceedingly simple first model, the OLS estimates of β_1 vary little in the other six models (ranging between .530 and .545), which essentially makes it unnecessary to select a "best" model.
- But you should wonder why the table presents results only for this particular subset of models and not for others . . .

Table 1: OLS estimates of β_1 for selected models

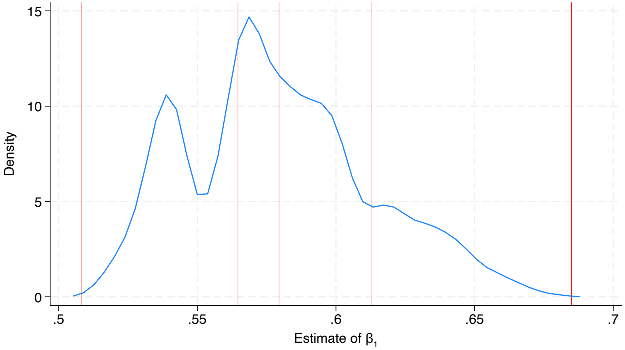
Covariate	LS1	LS2	LS3	LS4	LS5	LS6	LS7
Log arsons	.619	.534	.530	.531	.531	.533	.545
	(.041)	(.060)	(.047)	(.048)	(.049)	(.048)	(.049)
	[.540, .700]	[.416, .651]	[.437, .624]	[.437, .624]	[.435, .628]	[.438, .628]	[.449, .642]
Aged 15–19 with secondary degree		×					
Aging index			×	×	×	×	×
Children aged 0–5		×					×
Empl. in construction		×					×
GDP per employee							×
Log surface area		×					
Log population density			×	×	×	×	×
Pop. change 2001–2011			×	×	×	×	×
Self-employed				×	×	×	
Unemployment		×				×	
Youth empl. 15–29			×	×		×	
Youth unemployment					×		×
# observations	231	231	231	231	231	231	231
# parameters	4	9	8	9	9	10	11
# doubtful controls	0	4	3	4	4	5	6
Adjusted R^2	.595	.648	.618	.616	.616	.615	.628

Notes: All models also include the constant term and binary indicators for municipal capital and coastal municipality. Robust standard errors in parentheses, 95% confidence intervals in brackets.

Formal explorations

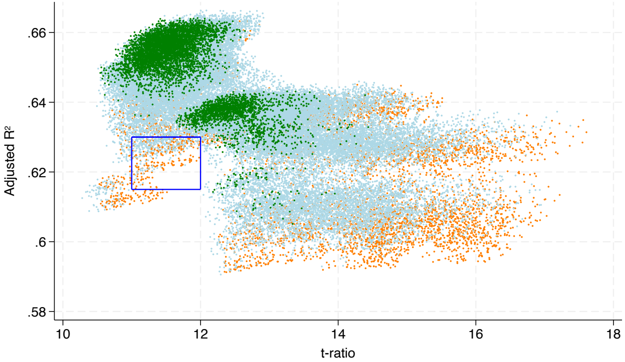
- Formal explorations of $\mathcal{M}_0$ or $\mathcal{M}$ follow a well-defined set of rules for finding the best model according to some selection criterion (e.g. adjusted R^2, Mallows' C_p, or AIC). The estimate of β_1 is again the OLS estimate from the best model.
- An example is best-subset selection, which looks for the best model in $\mathcal{M}$ that includes at most $h \le k$ doubtful controls after exploring $\sum_{j=0}^{h} \binom{k}{j}$ models.
- If $h = k$, one needs to explore the entire $\mathcal{M}$. Even for moderate values of k, say $k \geq 25$, this is computationally infeasible. For example, $2^{25} = 33,554,432$ and $2^{30} = 1,073,741,824$. Since $2^{16} = 65,536$, exploring the entire $\mathcal{M}$ when $k = 16$ is quite feasible (it takes about 2.5 hours on my MacBook).
- The average value of $\hat{\beta}_{1j}$ over $\mathcal{M}$ is $\text{avg}\{\hat{\beta}_{1j}\} = .579$. The variance of $\hat{\beta}_{1j}$ (a measure of model uncertainty) is .00111, while the average variance of $\hat{\beta}_{1j}$ (a measure of sampling uncertainty) is .00216. Hence, overall uncertainty about β_1 might be measured by $.00111 + .00216 = .00327$ or by $\sqrt{.00327} = .0572$.
- Figure 3: Distribution of the OLS estimates of β_1 over $\mathcal{M}$ (the "modeling distribution" in the terminology of Young and Holsteen, 2017).
- Figure 4: Scatterplot of t-ratios and adjusted R^2 over $\mathcal{M}$.
- Table 2: Effect of each doubtful control on the estimates (an example of "model influence analysis").

Figure 3: Distribution of the OLS estimates of β_1 over $\mathcal{M}$



Notes: Kernel density estimate with Epanechnikov kernel and Stata's default bandwidth. The vertical red lines correspond to $\min\{\hat{\beta}_{1j}\} = .508$, $\hat{\beta}_{1J} = .565$, $\text{avg}\{\hat{\beta}_{1j}\} = .579$, $\hat{\beta}_{11} = .613$, and $\max\{\hat{\beta}_{1j}\} = .685$.

Figure 4: Scatterplot of t-ratios and adjusted R^2 over $\mathcal{M}$



Notes: Coloring reflects the number h of included doubtful controls: Orange for $h \leq 5$, light blue for $5 < h \leq 10$, and green for $10 < h \leq 16$. The blue rectangle marks out the portion of $\mathcal{M}$ partially explored in Table 1.

Table 2: Effect of each doubtful control on the estimates

Doubtful control	Δ coeff.	Δ t-ratio	Δ adj. R^2
IndVec2001	-.05858	-4.2026	.05574
IncBamb2001	-.03592	-3.4901	.04600
Pop2011M2001	-.03562	-3.3378	.04558
IncGioC	-.03617	-3.3417	.04561
LivIstG	-.02060	-3.2220	.04585
AltIst	-.03694	-3.2761	.04985
BasIst	-.04354	-3.5784	.05002
TasDisGio	-.02980	-3.4077	.04469
TasOcc1529A	-.03310	-3.3122	.04485
CerSuFL	-.03354	-3.3493	.04433
OccTerzExtraCom	-.03364	-3.3595	.04462
OccManif	-.03453	-3.3859	.04555
OccCos	-.03182	-3.1283	.04569
OccSer	-.03254	-3.3638	.04449
OccAutonom	-.03423	-3.3249	.04433
PilxAddMigE	-.03091	-3.4876	.04464
Null model	.61298	15.8197	.59455

Notes: Each row shows the average difference between the estimate of β_1 from all models that include a particular control and that from the null model. Red and blue respectively denote the largest and smallest absolute differences.

The problem with standard model selection

- Irrespective of whether the exploration of the model space $\mathcal{M}$ is informal or formal, classical inferences after standard model selection need not have the properties established by classical statistical theory–which assumes that the model is fixed before seeing the data.
- The intuition is simple: "The selection of any aspect of a model or hypothesis using the data introduces sampling variability into the model or hypotheses, rendering random the specification process itself" (Kuchibhotla, Kolassa & Kuffner, 2022).
- Post-selection estimators are complex objects to analyze, but typically they are substantially biased, their distribution is quite far from the classical results derived under the assumption of a fixed model, and confidence sets based on classical distribution theory do not have the desired coverage level, not even asymptotically (see e.g. Leeb & Pötscher, 2005).
- Model averaging can be used to avoid some of these problems.

Enter WALS

- $\text{avg}\{\hat{\beta}_{1j}\}$ is an example of model-averaging estimator, one giving equal weight $1/J$ to the OLS estimators of β_1 from all models in $\mathcal{M}$.
- The general form of a model-averaging estimator of β_1 is:

$$\bar{\beta}_1 = \sum_{j=1}^{J} \lambda_j \hat{\beta}_{1j},$$

where the λ_j are non-negative weights that add-up to one and $\hat{\beta}_{1j}$ is the OLS (or some other) estimator of β_1 in the jth model in $\mathcal{M}$.

- There are many types of model-averaging estimators: Bayesian (with weights based on Bayesian priors), frequentist (with weights based on model fit or predictive accuracy), and hybrid. See Steel (2020) for a detailed review.
- An attractive one, for its computational simplicity and good MSE properties, is the Weighted Average Least Squares (WALS) estimator introduced by Magnus, Power & Prüfer (2010).
- WALS is simple to compute because it only requires three objects:
  - the OLS estimates of the free parameters in the null model;
  - the OLS estimates of the k doubtful parameters in the full model;
  - a set of k shrinkage factors $\omega_1, \ldots, \omega_k$ (the "WALS weights") obtained from a preliminary "Bayesian step" based on a "neutral" prior.

WALS estimates and WALS weights

- Using the updated Stata command `wals` (De Luca & Magnus, 2024) and its default Pareto prior, the WALS estimate of β_1 in my empirical problem is equal to .562–only slightly below the average value of .576 of the OLS estimates over $\mathcal{M}$–and its 95% (asymmetric) confidence interval is equal to [.469, .653].

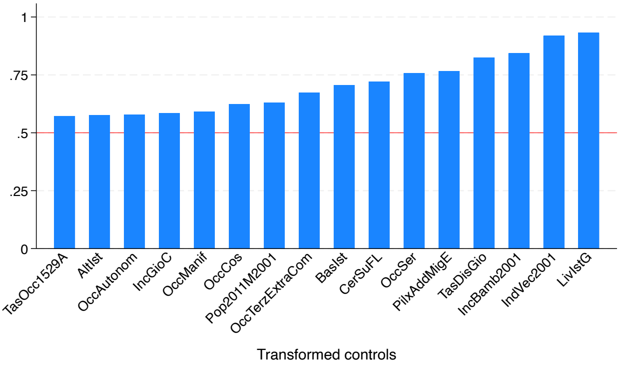
- The WALS weights ω_h are bounded between 0 and 1, but do not add up to one. How can we used them?

- $\omega_h = \sum_{j=1}^{2^{k-1}} \lambda_j$, $h = 1, \ldots, k$, so the WALS weights are akin to the posterior model inclusion probabilities in a Bayesian analysis and can similarly be used for model influence analysis.

- If the WALS weights are all equal to zero we obtain the OLS estimate $\hat{\beta}_{11}$ in the null model, and if they are all equal to one we obtain the OLS estimate $\hat{\beta}_{1J}$ in the full model.

- Figure 5: The WALS weights in my empirical problem.

Figure 5: The WALS weights in my empirical problem



The instrumental variables approach

- If a_i is endogenous, let z_i be a valid instrument for estimating β_1.
- In addition to the structural equation, now I can also consider the "first-stage equation"

  $$a_i = \gamma_0 + \gamma_1 z_i + \gamma_2 s_i + \gamma_3' X_i + u_i, \quad i = 1, \ldots, n,$$

  and the "reduced-form equation"

  $$c_i = \delta_0 + \delta_1 z_i + \delta_2 s_i + \delta_3' X_i + v_i, \quad i = 1, \ldots, n,$$

  where u_i and v_i are unobservable mean-zero prediction errors, by construction uncorrelated with z_i, s_i, and X_i.
- If the structural equation is correctly specified, then

  $$\delta_0 = \beta_0 + \beta_1 \gamma_0, \quad \delta_1 = \beta_1 \gamma_1, \quad \delta_j = \beta_1 \gamma_j + \beta_j \ (j = 1, 2), \quad v_i = e_i + \beta_1 u_i.$$

Just-identified IV estimators of β_1

- Since $\delta_1 = \beta_1\gamma_1$ if the structural equation is correctly specified, and $\gamma_1 \neq 0$ if the instrument is valid, solving out for β_1 gives

$$\beta_1 = \frac{\delta_1}{\gamma_1},$$

so the focus parameter is "just-identified".

- Clearly $\beta_1 = 0$ whenever $\delta_1 = 0$, a hypothesis that is easily tested using the reduced-form equation (Chernozhukov & Hansen, 2008).

- Although just-identified IV models are quite special, they are very common in empirical work.

- Because of the uncertainty about which of the k doubtful controls to include, the first-stage and the reduced-form equations may each be specified in $J = 2^k$ alternative ways, and so there are J alternative IV estimates of β_1.

- Given OLS estimates $\hat{\gamma}_{1j}$ of γ_1 and $\hat{\delta}_{1j}$ of δ_1 in the jth model, the resulting IV estimate of β_1 is

$$\check{\beta}_{1j} = \frac{\hat{\delta}_{1j}}{\hat{\gamma}_{1j}}, \quad j = 1, \ldots, J,$$

which coincides with the 2SLS estimate of β_1 in the jth model.

What instrument(s)?

- The forced resettlement program, 1956–1995.
- Figure 6: Number of forced resettlement in Apulia [Source: Questura di Bari].
- Validity of the proposed instrument when $\mathcal{M}$ consists of the $J = 2^{16}$ models with population density as the scaling variable:
  - Relevance: The value of the "first-stage F-statistic" (the square of the t-statistic on γ_1) is in most cases high enough to satisfy the conditions in Angrist & Kolesár (2023).
  - Assignment rule: The assignment of municipalities that are not provincial capitals to the forced resettlement program cannot be predicted based on available pre-1956 covariates and may be regarded as approximately random.
  - Balancing: Comparing means and standard deviations of the available pre-1956 covariates across "treated" and "untreated" municipalities shows no systematic differences between the two groups (provincial capitals are excluded from the test because "ineligible").
- Other possible instruments?

Figure 6: Forced resettlements in Apulia, 1956–1995

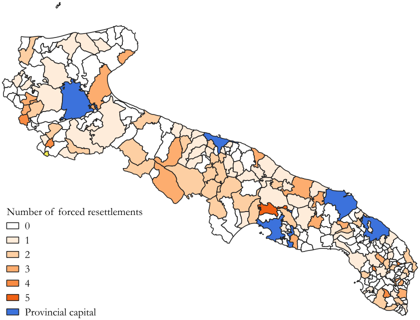
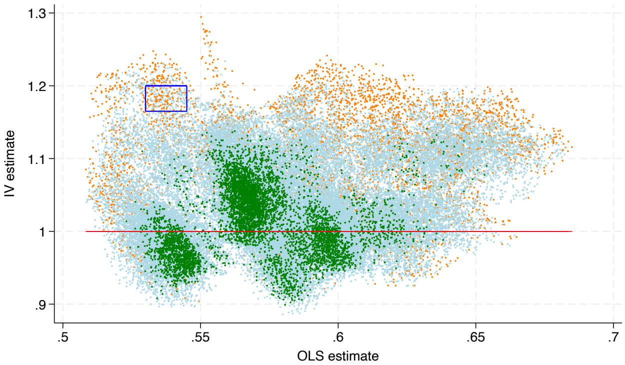


Table 3: OLS estimates of γ_1 and δ_1 and IV estimates of β_1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OLS estimate of γ_1	.457	.201	.506	.492	.486	.507	.494
	(.173)	(.140)	(.142)	(.141)	(.138)	(.139)	(.137)
	[.116, .799]	[-.076, .478]	[.226, .786]	[.215, .769]	[.215, .757]	[.233, .781]	[.224, .764]
OLS estimate of δ_1	.567	.230	.592	.586	.583	.591	.583
	(.144)	(.130)	(.126)	(.127)	(.124)	(.128)	(.125)
	[.284, .849]	[-.026, .487]	[.391, .888]	[.390, .889]	[.392, .881]	[.391, .895]	[.386, .878]
IV estimate of β_1	1.239	1.145	1.169	1.190	1.199	1.166	1.179
	(.326)	(.660)	(.267)	(.277)	(.280)	(.264)	(.267)
	[.597, 1.880]	[-.155, 2.445]	[.642, 1.696]	[.645, 1.735]	[.646, 1.751]	[.645, 1.686]	[.654, 1.705]
Aged 15–19 with secondary degree		×					
Aging index			×	×	×	×	×
Children aged 0–5		×					×
Empl. in construction		×					×
GDP per employee							×
Log surface area		×					
Log population density			×	×	×	×	×
Pop. change 2001–2011			×	×	×	×	×
Self-employed				×	×	×	
Unemployment		×				×	
Youth empl. 15–29			×	×		×	
Youth unemployment					×		×
# observations	231	231	231	231	231	231	231
# parameters	4	9	8	9	9	10	11
# doubtful controls	0	4	3	4	4	5	6
First-stage *F*-statistic	7.09	2.05	12.70	12.27	12.51	13.33	13.03

Notes: See Table 1.

Figure 7: Scatterplot of the OLS and IV estimates of β_1



Notes: Coloring reflects the number h of included doubtful controls: Orange for $h \leq 5$, light blue for $5 < h \leq 10$, and green for $10 < h \leq 16$. The blue rectangle marks out the portion of $\mathcal{M}$ partially explored in Tables 1 and 3.

Remarks

- The average value of the IV estimates of β_1 over $\mathcal{M}$ is avg$\{\check{\beta}_{1j}\} = 1.038$, much larger than avg$\{\hat{\beta}_{1j}\} = .579$ for OLS.

- In fact, the IV estimates are always larger than the OLS estimates. This may be explained by:
  - Downward bias of the OLS estimates because the number of arsons is a noisy measure of the local strength of organized crime.
  - Downward bias of the OLS estimates because X_i omits variables that are correlated with both soil consumption and the number of arsons, the two correlations having opposite sign. One example is the degree of moral integrity of municipal administrators/politicians.
  - Variation in β_1 across municipalities, in which case IV methods estimate a "local average treatment effect" (LATE) that can be larger than the average value of β_1 across municipalities (the "average treatment effect").

- The standard deviation of the IV estimates of β_1 over $\mathcal{M}$ (a measure of model uncertainty) is sd$\{\check{\beta}_{1j}\} = .062$, much larger than sd$\{\hat{\beta}_{1j}\} = .033$ for OLS, while the average standard error of $\check{\beta}_{1j}$ (a measure of sampling uncertainty) is .256.

- You may say: This is all very interesting, but what if one needs a single estimate of β_1 and is worried by the lack of transparency of dimensionality reduction or regularization methods?

Re-enter WALS

- Given WALS (or some other model-averaging) estimates $\bar{\gamma}_1$ of γ_1 and $\bar{\delta}_1$ of δ_1, I propose to estimate β_1 by
$$\tilde{\beta}_1 = \frac{\bar{\delta}_1}{\bar{\gamma}_1}.$$
- I instead ignore another estimate of β_1, namely the model-averaging estimate
$$\bar{\beta}_1^* = \sum_{j=1}^{J} w_j \check{\beta}_{1j},$$
since (at least for now) I am not sure how to weigh the IV estimates $\check{\beta}_{1j}$.
- In my empirical problem, using `wals` with its default Pareto prior gives:
  - $\bar{\gamma}_1 = .532$ against $\text{avg}\{\hat{\gamma}_{1j}\} = .498$,
  - $\bar{\delta}_1 = .495$ against $\text{avg}\{\hat{\delta}_{1j}\} = .480$,
  - $\tilde{\beta}_1 = .495/.532 = .930$ against $\text{avg}\{\check{\beta}_{1j}\} = 1.038$.
- Issues to be addressed:
  - Although $\sqrt{n}$-consistent, $\bar{\delta}_1$ and $\bar{\gamma}_1$ are biased in finite samples.
  - Consequently, $\tilde{\beta}_1$ is also biased.
  - Sampling distribution of $\tilde{\beta}_1$.

Conclusions about soil consumption and organized crime

- No matter what estimation method is employed, an increased strength of organized crime–proxied by the number of arsons–appears to increase soil consumption.
- The estimated elasticity of soil consumption to the number of arsons is less than one if OLS or WALS are used, but greater than one when the binary indicator for forced resettlement is used as instrument in a just-identified IV framework.

Some general conclusions about empirical work

- Practitioners recognize the importance of model uncertainty but address it in a haphazard and idiosyncratic way, and mostly ignore that standard model selection may create problems for classical inference.
- "Regression tables" are useful tools for communicating results and for local sensitivity analysis, but can only explore a small portion of the relevant model space.
- When the model space is huge, e.g. because of a large number of covariates or controls, you may question the choice of the necessarily small number of models included in these tables.
- This choice is usually not aimed at assessing sensitivity over the model space, or portions of it, but rather at convincing the audience (and, perhaps most importantly, the referees) that the "reference model" is sufficiently "robust".
- Whenever feasible, exploring the entire model space ensures transparency and can provide useful insights.
- If you need a single estimate of the effect of interest, model averaging is more intuitive than methods based on dimensionality reduction or regularization, and likely more efficient than those based on sample splitting.