

Empirical Research in Political Economy in the Years to Come

James M. Snyder, Jr.

What findings should we believe when the problems associated with Researcher Degrees of Freedom (and Waste Baskets and Publication Bias) meet the opportunities of Big Data?

One area in political economy where this might be especially important is the research on the “persistence of institutions” and the impact of “culture” on economic and political outcomes.

This is becoming a large literature

Acemoglu and Robinson (2001, *AER*), "The Colonial Origins of Comparative Development."

Nunn (2009), "The Importance of History for Economic Development."

Nunn and Wantchekon (2011, *AER*), "The Slave Trade and the Origins of Mistrust in Africa."

Dell (2010, *Econometrica*) "The Persistent Effects of Peru's Mining Mita," Districts subject to forced labor in colonial Peru are less developed today, especially through land tenure and infrastructure provision.

Sacchi de Carvalho (2015), "Diamonds Are Forever: Long-Run Effects of Mining Institutions in Brazil." Areas inside the Diamond District have greater luminosity, higher literacy and greater road density.

Vidal-Robert (2012), "The Persistence of the Inquisitorial Mind: Long-Run Effects of the Spanish Inquisition." Regions more affected by the Inquisition have lower economic development and generate fewer patents.

Nunn, Alesina and Giuliano (2013, *QJE*), "On the Origins of Gender Roles: Women and the Plough."

Jha (2013, *APSR*), "Trade, Institutions and Ethnic Tolerance: Evidence from South Asia."

Becker et al. (2015), "The Empire Is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy." Effects of Hapsburg empire.

An example from a recent NBER conference (chosen due to recency bias, nothing especially wrong with this particular paper):

Schultz (2016), "The Church's Ban on Consanguineous Marriages, Extended Kin-groups and Democracy"

- ▶ Church ban on cousin marriage from 800-1500 causes democracy today
- ▶ It also causes higher population density in 1500
- ▶ It also reduces penetration of mafia in southern Italy

[See also: Woodley and Bell (2012), "Consanguinity as a Major Predictor of Levels of Democracy: A Study of 70 Nations"]

- ▶ Why these outcome variables?
- ▶ "Exogenous" (it happened in the distant past) vs. "Exogenous" (it is uncorrelated with the error term)
- ▶ *In fact, precisely because it happened in the distant past we have no idea what it might be correlated with and why*
- ▶ What is the theory?

One example to see the size of the problem:

In the 1870 U.S. Census, including the census of agriculture and the census of religion, there are over 300 variables at the county level. Moreover researchers can easily justify using the difference or ratio of two or more variables. So just by chance, for any **one** dependent variable from the 2000 census there will be about

$$.05 \times 300 \times 299 \approx 4,500$$

combinations of variables from the 1870 census that are statistically significant at the .05 level and

$$.01 \times 300 \times 299 \approx 900$$

that are significant at the .01 level. And there are hundreds of variables in the 2000 census – e.g. more than 250 in the county datebook alone – so there will be more than

$$250 \times 900 = 225,000$$

coefficients that are significant at the .01 level. Lots to play with.

And there are about 3,000 counties and 50 states, adding lots of ways to “cut the data” (what about the south? urban counties? states major ports on the coastal or on rivers?)

For example, the fraction of students going to public schools rather than private academies in 1850 is a highly significant predictor of both the percent voting democratic in recent presidential elections as well as per-capita income in 2000 (both relationships are negative).

The theory, of course, is that public schooling is designed to conservative, solid citizens who are conservative, and who are good workers but not creative and therefore do not produce high economic growth.

In the introduction to the paper I include quotes such as this: Public schools “maintain the existing social and economic order of society... rewarding students who have the norms and values of the upper middle class, and bringing about the reproduction of the existing class structure – a monument to the preservation of power relations.” (Fusarelli and Young, *Journal of Thought*, 2011)

Etc. etc.

Solutions?

Pre-registration? But how do you prevent researchers from data-mining then pre-registering?

Cross-validation? This really only helps when model over-fitting is the main concern.

Replication: Exact replication is impossible but there are often “similar situations” in other countries than the one studied or in other time periods. Maybe we should reward this more. Currently we do not: For example: “The Persistent Effects of Peru’s Mining Mita” vs. “Diamonds are Forever.” Despite its clever title, the likely response from editors and reviewers is, “Another geographic matching paper on mining in the 17th century and GDP today? How boring. Not for [*top journal name here*].”

Theory: Why we need Ken Shepsle (and others)

On the other hand there are some benefits