Feature Essay

Persistence and Place: History and Geography in Political Economy

Jennifer Alix-Garcia, Oregon State University and Emily A. Sellars, Yale University

What accounts for differences in long-run development across countries and regions? Geography plays a central role in many theories of political and economic development, from early works emphasizing land productivity and proximity to markets (e.g., Baldwin 1956, von Thünen 1966) to more recent arguments about resource endowments (e.g., Engerman and Sokoloff 1997), disease burdens (e.g., Sachs and Malaney 2002), and terrain ruggedness (e.g., Nunn and Puga 2012). Even works that do not directly engage with geography frequently rely on proxies of geographic advantage - soil quality, slope, elevation, distance to important places, and climate conditions - as control variables to rule out alternative explanations for both historical and contemporary patterns of development.

In this essay, we highlight some of the empirical and theoretical challenges associated with studying geography and political economy over the long term, drawing on some of our recent work (Sellars and Alix-Garcia 2018, Alix-Garcia and Sellars 2018) and that of others.

Physical characteristics may not change, but their effects do

Scholars often think of geographic fundamentals as remaining constant over time. Though humans may transform their natural environment in various ways, the baseline characteristics of a location - for example, whether the terrain is uneven or the soil fertile - are usually thought of as "exogenous" and causally prior to any human-created changes. After all, mountains do not usually move, soil quality evolves in geologic time, and climate conditions, until recently, tended to shift slowly relative to most political and economic outcomes of interest. However, technology and policy can alter the consequences of time-invariant geographic factors, which poses some challenges for empirical work.

Consider the two sets of maps below, taken from our study on Mexican urbanization (Alix-Garcia and Sellars 2018). A motivating question of our work is whether and how the influence of geography on urban development has changed over time. That is, are the places that were "geographically advantaged" during the colonial era still benefiting from superior geography, or have policies and technology succeeded in restructuring the nature of geographic advantage? In Map (a), we plot the predicted probability that each 225-km^2 grid cell contains a city in 1900 given its geographic characteristics. In these models, we allow the geographic fundamentals of each grid-cell - elevation, potential maize productivity, distance to coast, and distance to fresh water source - to shift, holding all other variables constant at the mean. In Map (b), we show the *change* in the predicted probability that a grid-cell contains a city given

Figure 1: The Changing Role of Physical Geography Over Time

Figure a: Probability of urbanization in 1900, based on geographic covariates

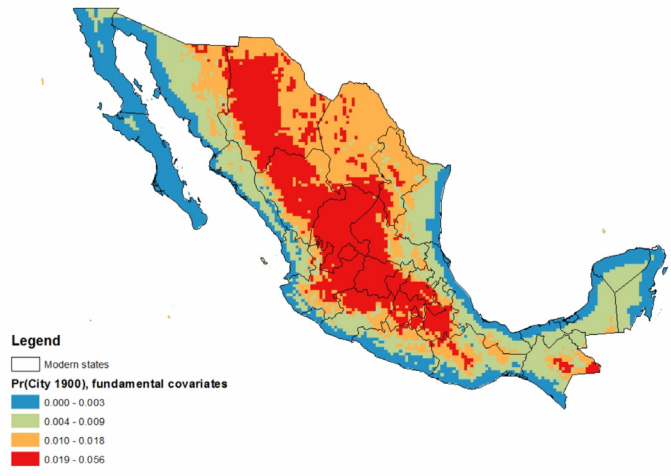
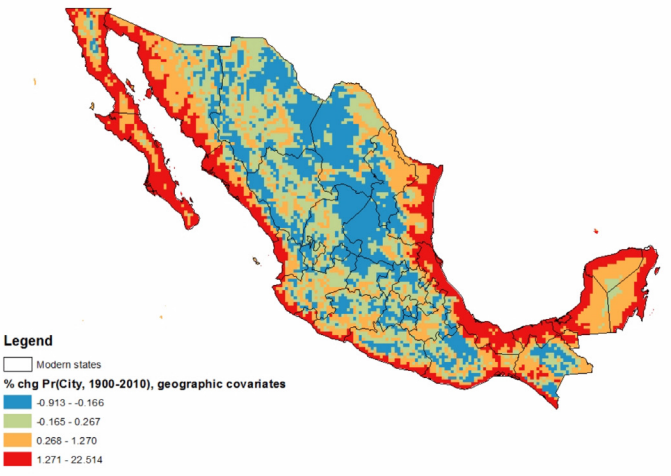


Figure b: Probability of urbanization in 2010, relative to 1900, as predicted by geographic covariates



its geographic characteristics in 2010 relative to 1900.

For simplicity, focus on one time-invariant geographic characteristic: whether the grid cell is located along the coastline. For much of Mexico's history, areas along the disease-prone coasts were underdeveloped relative to the central altiplano. By 2010, the geographic disadvantage of being in low-lying coastal areas had diminished substantially due to improvements in sanitation and the disease environment, and the government had invested large sums of money in the development of tourist areas in coastal regions.

continued on page 5

Alix-Garcia & Sellars Feature Essay...continued from page 4

Conversely, by 2010, and unlike earlier eras, terrain ruggedness had begun to act as a drag on urban development. Geographic conditions that were an asset in some eras thus became a liability in others and vice versa.

We are not the first to recognize that the effects of geography are not immutable and that technology conditions geographic advantage (e.g., Bosker et. al 2013; Henderson et. al 2018). However, the evolving relationship between geography and development has implications for empirical work that sometimes go unnoticed. For example, when examining the long-term effects of policy shocks or institutions, it may be not be enough to "control for" time-invariant factors through fixed effects or other strategies that do not flexibly change with time. What it means to be near a coast changes, and often so does the covariance between the distance to coastline and most outcomes of economic or political interest.

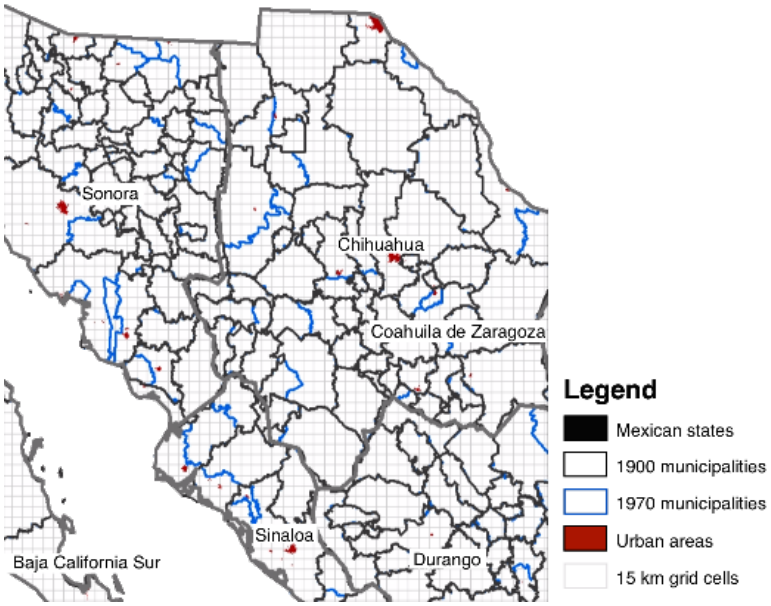
Selecting the appropriate unit of analysis can be challenging

Our research on long-term development and institutional change in Mexico has also changed the way we think about a seemingly simple topic: selecting the appropriate unit of analysis. Subnational territorial units are seldom stable over time; in fact, boundaries change endogenously based on political conflicts, population growth, and other factors that may be of direct empirical interest (e.g., Grossman and Lewis 2014). This can make it challenging to identify the correct unit of analysis in long-term studies, where important variables may be reported at different levels of analysis.

In our paper on the effects of Mexico's 16th-century population collapse on past and present land institutions (Sellars and Alix-Garcia 2018), we were faced with the problem of tracing continuities and changes in Mexico's land system over more than 400 years, through numerous changes in political structure and subnational territorial divisions. Mexico's 20th-century political boundaries, the level at which our outcomes and many covariates were measured, do not nest into the colonial administrative units at which colonial population data were reported. Even more complicated, a central contribution of this paper was in tracing how the historical legacy of the population shock changed over time, requiring us to examine intermediate outcomes measured at yet another level of aggregation. We ultimately decided to conduct the analysis at the level of disaggregation of the main outcome variable and key conditioning variables, addressing spatial dependence in various ways. Though our results ended up being insensitive to our choice of aggregation, the process of writing this paper highlighted the theoretical and empirical challenges associated with studying the long-term effects of historical events or institutions. Should we measure historical persistence at the level at which the initial shock was recorded, at which modern outcomes are measured, at which mediating factors either built upon or undermined the early legacies of the shock, or some combination of these?

For our paper on urbanization (Alix-Garcia and Sellars 2018), we had a dizzying array of choices for how to measure population concentration over time, from the over 100,000 localities of modern Mexico to the intendencies of the late colonial period. Even just since 1990, there have been considerable changes locality and metropolitan area designations, driven by population growth, urban consolidation, government policy, and other factors. Though somewhat more stable, municipalities (akin to counties in the U.S. context) and states vary greatly in size, in part because of past and present population concentration (Figure 2). The number of options forced us to think about which unit of analysis best reflected the data-generating process we wished to capture. In examining city emergence, we conceived country as an open canvas where population centers could theoretically emerge at any place or time. We thus opted to

Figure 2: Urbanization and Subnational Boundaries



continued on page 6

Alix-Garcia & Sellars Feature Essay...continued from page 5

examine 15 x 15 km grid cells (the size of the median municipality in contemporary Mexico) projected across the space of country for most of our analysis. However, this choice raised other questions about measurement and spatial dependence, as we discuss below.

The choice of aggregation poses measurement challenges

Even where the appropriate unit of analysis is clear, aggregating geographic data to this level can pose further challenges. Modern remote sensing data can now record geographic information at levels smaller than a square meter. Though these high-resolution data can be a goldmine for both natural and social scientists, researchers need to think carefully about the level at which decisions are made that generate these data (seldom as small as 1-meter), how to address spatial dependence, and how to merge data with information from other sources.

Figure 3 illustrates some of the difficulties with measuring geographic fundamentals at different levels of aggregation. Panel a. shows elevation measures in 30-m pixels for a portion of northwestern Mexico. Panel b. shows the average elevation within the 15 x 15 km grid cell units used in our urbanization paper, and panel c. shows elevation aggregated to the level of the 1900 municipality, or the smallest subnational unit that can be traced to 1900 (Sellars and Alix-Garcia 2018). Neither grid cells nor municipalities measure

Figure 3: Elevation at Various Levels of Aggregation

Figure a: Original digital elevation model

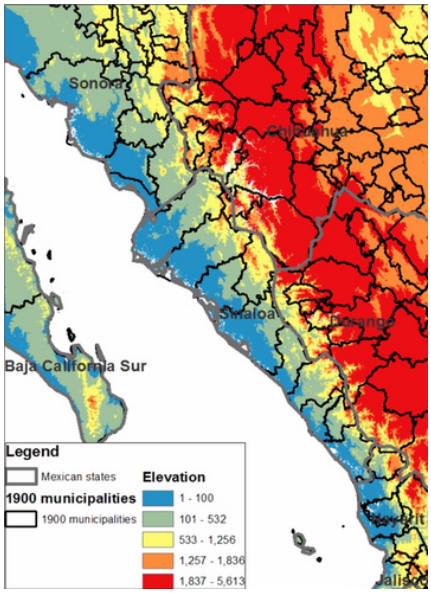


Figure b: 15 x 15 km grid cells

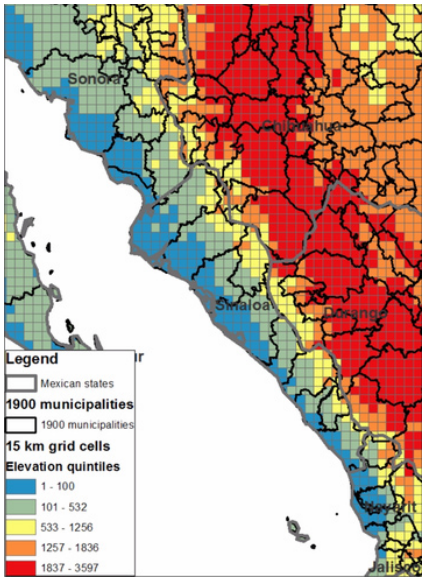
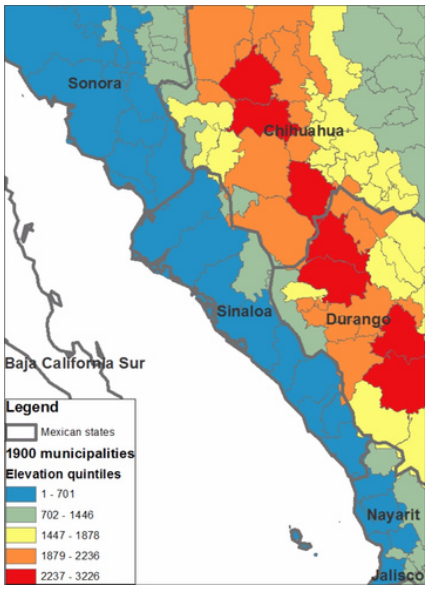


Figure c: Historical municipal boundaries



"elevation" accurately, and the discrepancy gets worse at higher levels of aggregation. Several municipalities in Figure 3 contain cells in the first two quintiles of elevation, while municipal averages put the entire area in the first quintile. Aggregating up thus introduces non-classical measurement error into the elevation measure, an example of the modifiable areal unit problem.

Is the answer to always use smaller units when possible? Though smaller grid cells more accurately reflect variation in elevation, selecting a unit that is too small can also be problematic. If other data are recorded with less precision, or if relevant decision-making really occurs at a higher level, using units that are too small can also introduce bias and inefficiency in statistical estimates. For example, consider using the elevation data to ask how terrain ruggedness (the standard deviation of elevation) has influenced patterns of urbanization over time. While elevation data are available for extremely small units, theories on how ruggedness influences urbanization (such as work on defensive position or urban infrastructure costs) typically operate at levels much higher than 1- or 5-meter pixels. Attempts to study this question at the level of units that are "too small" introduce noise by honing in on sources of variation that have little to do with the theory being evaluated, such as minute differences in ruggedness caused by fallen rocks or human-made ditches. Researchers exploiting highly disaggregated satellite data also need to remember not to treat these tiny units as independent observations. Ignoring broader spatial correlation can lead researchers to dramatically underestimate standard errors, leading to incorrect inference.

For an excellent overview of these technical issues, see Avelino et al. (2016). There is no simple solution, illustrating why it is important to think carefully about the units at which theories are thought to operate, spatial dependence, and measurement. In our

continued on page 7

Alix-Garcia & Sellars Feature Essay...continued from page 6

case, we thought about the scale or scales at which decision-making related to institutions or urbanization takes place and used this to guide our thinking. We also examined whether and how results changed at different levels of aggregation and using different methods to account for temporal and spatial dependence between units. In addition to illustrating the robustness of our empirical results, this exercise also required us to clarify our thinking about the theories we were hoping to evaluate and their empirical implications.

Works cited

Alix-Garcia, Jennifer and Emily A. Sellars. 2018. "Locational Fundamentals, Trade, and the Changing Urban Landscape of Mexico." *Unpublished manuscript*: https://drive.google.com/open?id=1OGJ6eOlZRZcTpgCBwh8zlrJ-RvA6qjaH.

Avelino, Andre F. T. et al. 2016. "Goldilocks and the Raster Grid: Selecting Scale when Evaluating Conservation Programs." *PLoS One* 11(12): e0167945.doi:10.1371/journal.pone.0167945.

Baldwin, Robert E. 1956. "Patterns of Development in Newly Settled Regions." *The Manchester School* 24(2), 161-179.

Bosker, Maarten, Eltjo Buringh, and Jan Luiten Van Zanden. 2013. "From Baghdad to London: Unraveling Urban Development in Europe, the Middle East, and North Africa, 800–1800." *Review of Economics and Statistics* 95(4), 1418-1437.

Engerman, Stanley L. and Kenneth L. Sokoloff. 1997. "Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies." In *How Latin America Fell Behind*. Ed. Stephen Haber. Stanford: Stanford University Press, 260-304.

Grossman, Guy and Janet I. Lewis 2014. "Administrative Unit Proliferation." *American Political Science Review* 108(1): 196-217.

Henderson, J. Vernon et al. 2018. "The Global Distribution of Economic Activity: Nature, History, and the Role of Trade." *The Quarterly Journal of Economics* 133(1), 357-406.

Nunn, Nathan and Diego Puga 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *Review of Economics and Statistics* 94(1), 20-36.

Sachs, Jeffrey and Pia Malaney. 2002. "The Economic and Social Burden of Malaria." *Nature* 415: 680-685.

Sellars, Emily A. and Jennifer Alix-Garcia. 2018. "Labor Scarcity, Land Tenure, and Historical Legacy: Evidence from Mexico." *Journal of Development Economics* 135, 504-516.

Thünen, Johann Heinrich von. 1966 (1850). *The Isolated State: English Edition*. Translated by Carla M Wartenberg. Ed. Peter Hall. Oxford/New York: Pergamon Press.

