Estimating sub-national wealth inequality in the US using ensemble learning

Keywords: Ensemble methods, machine learning, geography, economic inequality, wealth

Extended Abstract

In the United States, 10 percent of families possess 70 percent of national wealth (Piketty and Zucman 2015). Wealth inequality, defined as disparities in household assets net of debts, has grown continuously since 1980 (Kuhn, Schularick, and Steins 2020). While these disparities have contributed to the rise in income inequality (Piketty, Saez, and Zucman 2018), wealth and income represent distinctive facets of economic inequality, with potentially different roots and implications (Killewald, Pfeffer, and Schachner 2017; Cowell et al. 2017). Researchers and policymakers increasingly recognize the links between wealth inequality and a wide range of social, economic, political, and even epidemiological consequences (Neckerman and Torche 2007).

Despite growing awareness and expanding research into economic disparities across local labor markets (Moretti 2012; Manduca 2019; Kemeny and Storper 2020), information about the geography of wealth and how it has changed remains limited. At a national scale, wealth inequality in the US has been investigated using administrative data linked to taxes (Piketty, Saez, and Zucman 2018), or a range of smaller household surveys (Killewald, Pfeffer, and Schachner 2017). While each has advantages and disadvantages (Kuhn, Schularick, and Steins 2020), concerns around confidentiality mean that none of these data can be directly used to describe meaningful spatial disparities in wealth.

To overcome this challenge, we generate predictive models of household wealth using ensemble methods, built using rich survey information from the Federal Reserve's Survey of Consumer Finances (SCF), as a means to predict wealth among households in Census population surveys that include geographical identifiers. The end result is a dataset that permits description of inter-place variation in wealth, as well as the distribution of wealth within individual local economies, allowing researchers to track the evolving geography of wealth and wealth inequality between 1940 and 2020 across 722 local labor markets that span the entirety of the contiguous US.

Our construction of the data comprises three steps: (1) build a model (stacked ensemble) of wealth using the SCF; (2) predict wealth using Census population survey data; and (3) estimate wealth and wealth inequality at various spatial scales. Figure 1 shows the performance of the ensembles separately for households with positive and negative wealth in the held out SCF data (i.e. test sample, not used for fitting the ensemble). There is an evident strong fit for households with positive wealth – the models are highly accurate at the household-level, with a root mean squared error (RMSE) of 1.21. The fit is less good for those households with negative wealth (10% of sample; RMSE = 1.47) due to a relative lack of information which can be used to quantify negative wealth (i.e. housing value and income items provide little information as to the quantum of negative wealth). This performance is a substantial improvement over the ensemble base learners (7 models total, including random forest and linear model), and we validate the performance on an out-of-sample dataset derived from the Panel Study of Income Dynamics for 2019.

Armed with stack ensembles, we impute wealth for each observation in the census for years from 1960 to 2020. The top panel of Figure 2 compares our Gini coefficient from our country-level estimate with that provided by Saez and Zucman (2020), which is estimated using distributional macroeconomic accounts. The bottom panel compares our estimates of the top 1% share of total wealth, the top 10% share, and the bottom 50% share with those provided by Saez and Zucman (2016). The estimates, while differing slightly in levels, broadly agree in terms of trends. This provides us with comfort that our estimates derived from imputing household wealth in the census is picking up important aggregate level dynamics.

We then compute inequality at varying levels of geography, including the Commuting Zone (CZ) level, and examine how inequality has evolved over time. Figure 3 maps inequality by CZ for different years. In 1960, wealth inequality was relatively high across the country, although CZs in the Great Lakes region tended to be relatively more equal. As we see with the country-level trend, inequality declined significantly across the map by 1980. By 2020, wealth inequality had increased and was spread more evenly across the country. Rather than high-levels being concentrated in the coasts and South, the central and southern CZs of the Midwest began to see relatively higher levels. This is made more clear in Figure 4, which computes the change in the Gini coefficient over the 1960-2020 span. This reveals that wealth inequality has declined in most CZs, however there is a clear spatial clustering of increases in inequality in the West North Central region of the Midwest, as well as in small pockets in North Texas and parts of the South.

We hope that the resulting dataset, which will be made available publicly, will enable research into the trends and consequences of wealth inequality in the United States.

References

Cowell, Frank, Brian Nolan, Javier Olivera, and Philippe Van Kerm. 2017. “Wealth, Top Incomes and Inequality.” In *National Wealth: What Is Missing, Why It Matters*, 175–206. New York: Oxford University Press.

Kemeny, Tom, and Michael Storper. 2020. “The Fall and Rise of Interregional Inequality: Explaining Shifts from Convergence to Divergence.” *Scienze Regionali* 19 (2): 175–98.

Killewald, Alexandra, Fabian T Pfeffer, and Jared N Schachner. 2017. “Wealth Inequality and Accumulation.” *Annual Review of Sociology* 43: 379–404.

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins. 2020. “Income and Wealth Inequality in America, 1949–2016.” *Journal of Political Economy* 128 (9): 3469–519.

Moretti, Enrico. 2010. “Local Multipliers.” *American Economic Review* 100 (2): 373–77.

———. 2012. *The New Geography of Jobs*. Houghton Mifflin Harcourt.

Neckerman, Kathryn M, and Florencia Torche. 2007. “Inequality: Causes and Consequences.” *Annu. Rev. Sociol.* 33: 335–57.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. “Distributional National Accounts: Methods and Estimates for the United States.” *The Quarterly Journal of Economics* 133 (2): 553–609.

Piketty, Thomas, and Gabriel Zucman. 2015. “Wealth and Inheritance in the Long Run.” In *Handbook of Income Distribution*, 2:1303–68. Elsevier.

Saez, Emmanuel, and Gabriel Zucman. 2016. “Wealth inequality in the United States since 1913: Evidence from capitalized income tax data.” *The Quarterly Journal of Economics* 131 (2): 519–78.

———. 2020. “The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts.” *Journal of Economic Perspectives* 34 (4): 3–26.

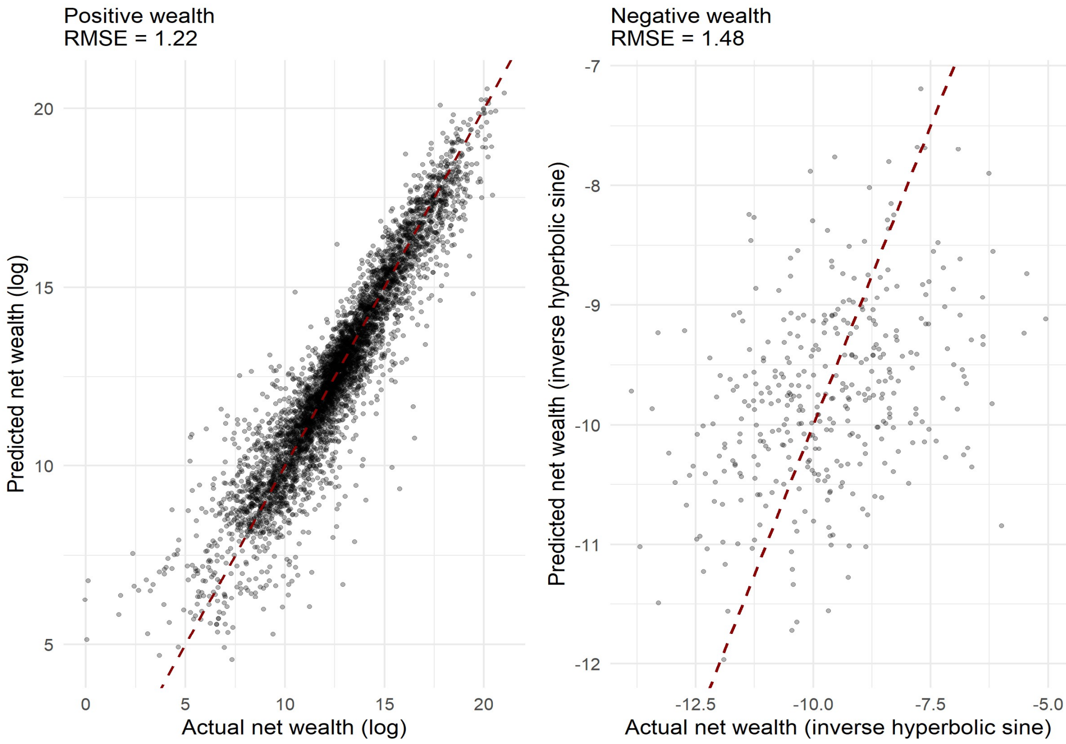

Figure 1. Performance of stacked ensemble on test sample

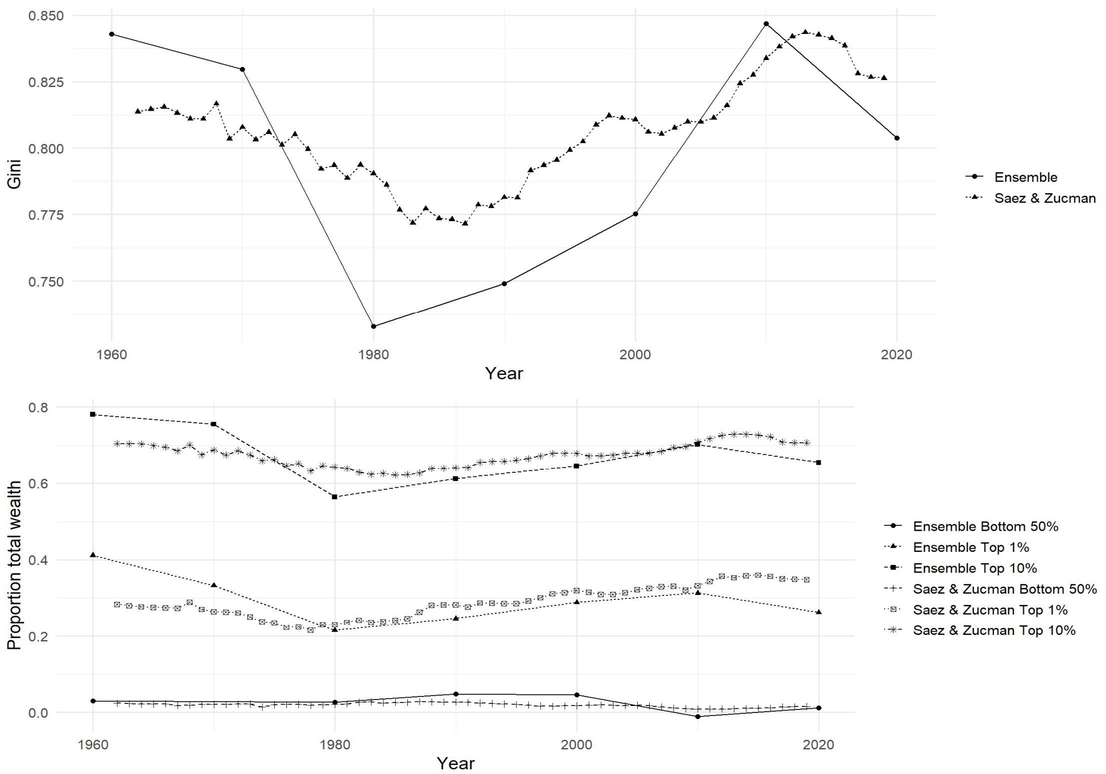

Figure 2. US wealth inequality, comparison of estimates

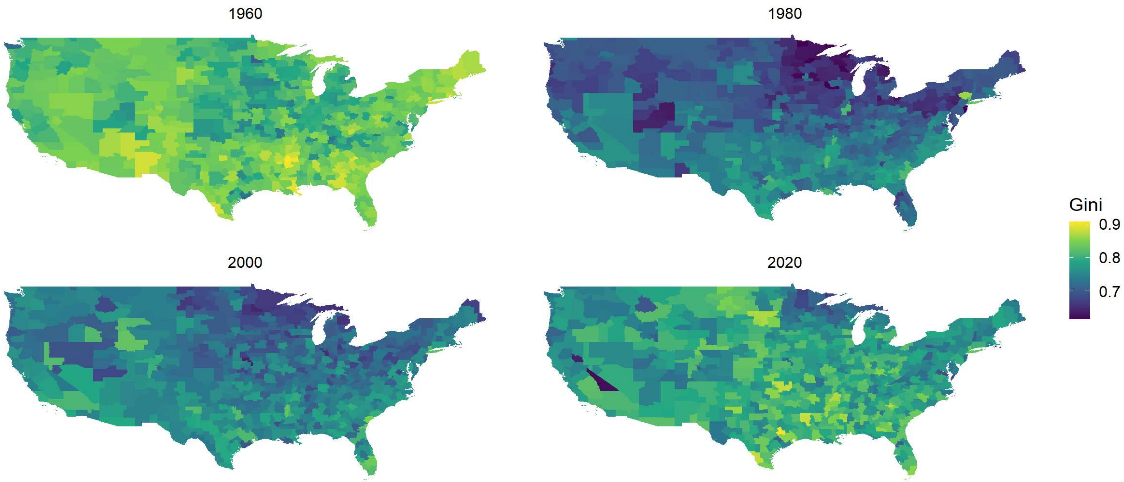

Figure 3. Wealth inequality by Commuting Zone, 1960-2020

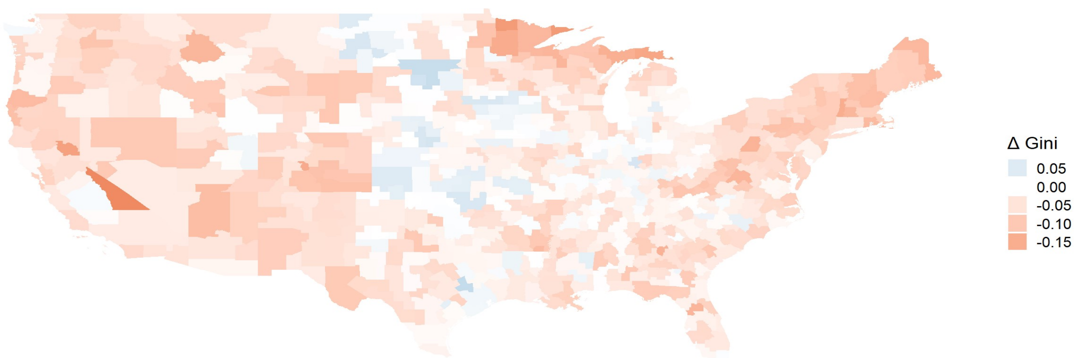

Figure 4. Change in wealth inequality by Commuting Zone, 1960-2020