# Introduction: Big Data in Political Economy

ATIF MIAN AND HOWARD ROSENTHAL

The massive growth in computing since the 1980s and 1990s has revolutionized data gathering and how people transact with one another. The result is that practically every economic and financial transaction is recorded somewhere by someone and can be linked to the individuals undertaking the transaction. Such proliferation of "big data" has made it possible for both economists and political scientists to empirically analyze questions that earlier could be addressed only theoretically. In particular, big data permits us to study behavior at both a high level of disaggregation and a high time frequency. For example, what is a household's spending behavior and how does it depend on changes in interest rates, asset prices, or political events? How do households form expectations of future events? How do ideology and electoral politics affect these expectations? What are the distributional consequences of macro shocks—such as the impact of monetary policy or housing collapse on the rich versus the poor? These are fundamental economic and political questions that can now be addressed using advancements in data collection and computing.

## BIG DATA: WHAT IS NEW AND DISTINCTIVE

There are numerous examples of research using new, disaggregated data sources, several appearing in this issue. These include data on mortgage originators (Igan, this issue); national data on individual voter registration and turnout (Catalist); data on the characteristics of individual professionals such as medical doctors (Bonica, Rothman, and Rosenthal 2014, 2015) or lawyers (Bonica, Chilton, and Sen 2015); government payments to contractors; Medicare payments to physicians; pharmaceutical company payments to physicians; campaign contributions (Bonica, this issue; Dimmery and Peterson, this issue); lobbying (Igan, this issue); tariffs (Kim 2014); traditional and social media content; government documents (O'Halloran et al., this issue); Google searches (Chae et al. 2015); and Twitter followers (Barberá 2015).

Of course, for big data to be seen as transforming research in political economy, it must be more than just the analysis of data sets with very large numbers of observations. Researchers have been exploiting the census for decades. Similarly, the pathbreaking research of Thomas Piketty and Emmanuel Saez (2003), using individual IRS records, dates from the turn of the century. In the 1980s, Keith Poole and Howard Rosenthal (1991) studied the entire congressional history of tens of millions of individual roll call voting decisions with a supercomputer. So what is distinctive about the current use of "big data" in political econ-

Atif Mian is Theodore A. Wells '29 Professor of Economics and Public Affairs at Princeton University. Howard Rosenthal is professor of politics at New York University and Roger Williams Straus Professor of Social Sciences, Emeritus, at Princeton University.

Direct correspondence to: Atif Mian at atif@princeton.edu, 26 Prospect Ave., Princeton, NJ 08540; and Howard Rosenthal at hr31@nyu.edu, NYU Department of Politics, 19 W. 4th St., New York, NY 10012.

omy? At least the following considerations appear relevant[1]:

1. *A new ability to link large data sets that are of far more limited use if unlinked* has emerged. For example, political activity in the form of lobbying can be linked to microlevel data on the firm's business activity, such as mortgage lending behavior in metropolitan statistical areas (see Igan, this issue). Another example is political activity in the form of roll call voting by a member of Congress, which can now be linked not just to aggregate economic characteristics such as median income but more finely to characteristics of small geographic units in congressional districts, such as mortgage foreclosure activity in portions of the district that have a high level of Republican voting (Mian, Sufi, and Trebbi 2010).

An important aspect of record linkage is the development of automated record linkage through the use of algorithms that assign a probability that a record from one data set can be matched to another. Record linkage is also facilitated by geocoding techniques. Researchers are recognizing that matches must carry an acceptable level of measurement error but need not be perfect. For example, political activity in the form of campaign contributions can be linked to the professional and demographic characteristics of individuals in most licensed professions (medicine, law, nursing) or state government employment and in some cases to income data (state government employees, including academics and physicians in university hospitals). More recently, researchers have been able to link public records such as bankruptcy filings (for example, Dobbie and Song 2015) with Social Security data to address questions like the impact of debtor relief on earnings and labor supply. Atif Mian, Amir Sufi, and Nasim Khoshkhou (2016) link constituent ideology and voting outcomes with consumer spending at the county level and with individual survey data on consumer sentiments to analyze the link between consumer spending and sentiments about government policy.

2. *A growing ability to extract data directly from web pages,* using Python and other tools, has become an important source of additional data. For example, Matthew Gentzkow and Jesse Shapiro (2010) use textual analysis of online newspaper data to construct measures of "slant" in various newspapers.

3. *The growth of computing capacity* remains important. For example, Chris Tausanovitch (this issue) carries out an ideological scaling using hundreds of thousands of public opinion surveys. The scaling takes advantage of special software that uses graphics chips to turn PCs into parallel processors. Changes in estimation strategy are also likely to accompany the use of big data. For example, Kosuke Imai, James Lo, and Jonathan Olmsted (2015) have recently proposed using efficient expectation-maximization (EM) algorithms for ideological scaling to replace the widely used Markov chain Monte Carlo (MCMC) methods. Computing capacity and estimation strategy are likely to be particularly important in the growing area of text analysis, as illustrated by O'Halloran and her colleagues in this issue.

4. *The private sector has become a large provider of big data* of potential usefulness to political economists. Big data about financial markets have been available for many years, accessible to academics through Wharton Research Data Services (WRDS) and other sources. More recently, data have begun to be accumulated about career paths (LinkedIn) and about housing and consumer markets. The private sector both complements and substitutes for the government sector. For example, LinkedIn can provide data about workers in unlicensed professions that can comple-

1. The software firm SAS characterizes big data as having volume, velocity, variety, variability, and complexity. See SAS, "Big Data: What It Is and Why It Matters," http://www.sas.com/en_us/insights/big-data/what-is-big -data.html?keyword=big%20data&matchtype=e&publisher=google&gclid=CjwKEAiAxfu1BRDF2cfnoPy B9jESJADF-MdJIJyvsnTWDXHchganXKpdoer1lb_DpSy6IW_pZUTE_hoCCwDw_wcB (accessed February 13, 2016). "Velocity" and "variability" refer to real-time applications, which are not yet present in political economy. The papers in this issue represent applications that have large volumes of data, data arising in different formats, and data with complex structures.

ment the data in government databases about those in licensed professions.

The growth of online payment and personal finance tools has given researchers access to people's spending and investing behavior. For example, Scott Baker (2014) uses data from individual accounts at a personal finance site to investigate how consumers respond to income shocks in the presence of debt. Mian, Rao, and Sufi (2013) use data on credit card spending to analyze the impact of the housing collapse on spending. Similar data have been used to analyze the impact of political shocks—as when the federal government approached the fiscal cliff or when it was threatened with shutdown—on consumer spending behavior as well.

A related feature of these data is that they are potentially available at high frequency, such as daily spending behavior. The high frequency enables researchers to exploit the sharp timing of certain events—such as the fall of Lehman Brothers in September 2008 or the attacks of September 11, 2001—to analyze the impact on consumer spending and investment behavior.

Credit bureaus, both in the United States and abroad, are another important private source of data. The credit bureaus contain data on all types of borrowing at the individual level at monthly frequency. These data also contain information on an individual's location and basic demographics and are thus potentially linkable to other data sets. Mian and Sufi (2014) describe a number of examples of research studies using credit bureau data.

A number of private firms specialize in collecting and consolidating data from a large number of public data sources. For example, the Securities and Exchange Commission (SEC) requires publicly traded corporations to file a variety of reports, including information on trading by insiders and on large block holdings. Since 1993, this information has been available in electronic form on the SEC's EDGAR platform. But the SEC has done little to summarize these reports in a way that would be useful to researchers. One cannot go to the SEC site and download a spreadsheet with the details of the largest owners of S&P 1500 companies. On the other hand, firms like Vickers

Stock Research have such data in more accessible forms.

5. *Government electronic record keeping has also expanded dramatically.* About the time the SEC created EDGAR, for example, government agencies in the fifty states, such as education departments, were shifting to electronic, web-accessible data. Records that were previously accessible only as copied or scanned documents became available in spreadsheet form. A transition to transparency has accompanied the technological transition to electronic record keeping.

Data on government payments to most contractors have long been a matter of public record, but the provision to the public of information on government payments to health care providers, long resisted by the providers, did not become federal government policy until 2014. Similarly, disclosure of payments to providers by pharmaceutical companies was required by the Physician Payments Sunshine Act, a part of the Affordable Care Act passed in 2010.

There have also been important shifts in the availability of large individual-level data sets at various governmental organizations. For example, academics have worked with the Internal Revenue Service (IRS) on tax return data and the Social Security Administration (SSA) on payroll data. These data have been extremely useful in illuminating trends in inequality and social mobility. At the same time, the granularity of the data sets is useful in helping us better understand the impact of changes in tax laws and other public policy interventions. The U.S. Census Bureau also maintains data on sales and employees by firm.

6. At the same time, *the development of optical character recognition (OCR) made it possible to process older data at relatively low cost*. Ten years ago, analysis of roll call voting data was largely limited to the U.S. Congress. Boris Shor and Nolan McCarty (2011) have extended this work to all fifty states.

## THE CHALLENGES OF BIG DATA

The use of big data does present some challenges for academic research. There are questions of data accuracy. There is a question of

equal access to data. There is a question of the ethics of the relationship of academic researchers to private-sector collectors of data. Although the challenges we identify apply more generally to the social sciences, political economy faces some particularly intensive challenges: because political economy addresses the interplay between political transactions and market transactions, the need for market transaction data makes political economists heavily dependent on the private sector.

### Data Accuracy

There are several potential problems with respect to data accuracy:

#### What Inferences Can Be Made from the Sampling Universe?

This question is particularly relevant for data from search engines and social media. Are individuals who search on Google representative of the larger population? Are heavy searchers representative of all searchers? Are Twitter users representative of broader population? Some of the data in the Tausanovitch paper in this issue come from surveys conducted through the Internet. In longitudinal studies, how will these data match up with data collected in the 1950s through door-to-door interviews or with telephone interviews in the 1990s?

Campaign contributions, explored in the Bonica paper in this issue, allow us to study groups that are not reported in sample surveys. For example, medical doctors would represent only on the order of 1 percent of the respondents in a survey of 2,000 adult Americans. But 145,000 physicians have made campaign contributions, with an indication of partisan preference, over the past twenty years (Bonica et al. 2014, 2015). Those 145,000 physicians, in turn, can be broken down into still large samples by specialty, gender, and employer type. But are these 145,000 representative of the nearly 900,000 physicians in the United States?

Another larger source of partisan preference could come from voter registration data put together by Catalist. The Catalist data have also been used to study physician preferences

about patient management (Hersh and Goldenberg 2016). Again, are physicians who are registered voters representative of physicians? Are former government employees with LinkedIn accounts representative of all former government employees?

A related big data development is represented by attempts to "bridge" different sampling universes by using common stimuli—for example, a legislative roll call vote on a bill and media editorials on the same bill. Jeffrey Lewis and Chris Tausanovitch (2015) survey this literature and discuss its promises and shortcomings.

Record linkage introduces inaccuracy. In the case of campaign contributions to candidates, the reports of individual candidates may be prepared by unpaid interns who lack strong incentives to be accurate. Even when the initial reports are filed accurately, reports across candidates can have a different name spelling and address for the same individual. Conversely, individuals with common last names can be confused. When the contributors are linked to another database, such as the National Provider Identifier (NPI) database that the government maintains for physicians, there is further opportunity for mismatch.

#### New Sources of Big Data May Contain Misrepresentation

Misrepresentation is hardly a new problem. For instance, the November Current Population Survey (CPS) has long been used to study voter turnout (Wolfinger and Rosenstone 1980), but turnout is substantially overreported in the CPS. Citizenship is also likely to be overreported (McCarty, Poole, and Rosenthal 2006). Income tax and estate tax returns are subject to fraud. Misrepresentation may be particularly important in loan markets (Griffin and Maturana 2013; Mian and Sufi 2015; Keys, Seru, and Vig 2012). Mian and Sufi (2015) show that income reporting on publicly available Home Mortgage Disclosure Act (HMDA) files was subject to large-scale overstatement by mortgage applicants during the mortgage credit boom of 2002 to 2006. The financial incentives of firms to misreport do, however, represent a new concern.

## Data Access

As we previously indicated, much of the new big data is being generated by organizations, both for profit (LinkedIn) and by nonprofits (ProPublica) that charge fees for data access. When an academic researcher uses proprietary data, what are the conditions for replication? Should journals allow publication if the entire data set cannot be made available for replication and further study? When the data are purchased, the purchase agreement may exclude the posting of replication materials.[2]

Government agency rules regarding data access have not been sufficiently streamlined yet. There is natural aversion by government agencies to "sharing" their internal data. The reluctance may be due to the fear of either lawsuits or scrutiny by outsiders of how the agency works. The latter excuse warrants greater transparency, as access might have the beneficial side effect of improving the functioning of some government agencies. Another source of reluctance is pressure from private interests. For example, until recently, such pressure kept Medicare payments to individual physicians from public scrutiny.

A related issue is the ability to link various government data sets, which raises a natural concern about privacy. Data are often anonymized before they are shared with researchers. Although this is a good practice to follow, anonymizing data makes it difficult to link them across different sources. It would be useful if the government came up with a mechanism to link the various data sets before anonymizing them so as to expand the scope of the research that could be conducted using governmental sources of data.

Along similar lines, there is also a need for the government to come up with uniform data access rules across its various agencies. Access to governmental data sometimes depends on who within the agency one knows and can collaborate with. As such, the playing field is not level when it comes to access to government-owned data.

Another question concerns funding: not all academic researchers have the resources to purchase the data in the first place. As government funding for research ebbs—the National Science Foundation (NSF) cut out political science for a period starting in 2013—researchers with large internal research funds, in either professional schools or elite universities, will have an advantage over others. It is also conceivable that private sources of data could grant differential access, essentially limiting access to those individuals whom an organization believes to be "safe." Many private data contracts already give the right of refusal to the data provider in case the provider objects to the research findings.

These important questions regarding access and scientific bias need to be addressed carefully as more and more private data sources are used by academics.

## The Ethics of Collaboration

An ethical issue arises when there is an academic collaboration with a for-profit generator of big data. The situation was highlighted by the Facebook deception study in 2014 (Albertson and Gadarian 2014). The study, which involved a researcher from Cornell University, had the "big data" advantage that it was possible to study the behavior of 700,000 individuals. The big data issue is that a private firm, such as Facebook, has proprietary interests and research objectives that can differ from those of a small, on-campus laboratory experiment monitored by a university's institutional review board (IRB). In the case of the Facebook deception experiment, the Cornell IRB approved the study with the argument that it was Facebook, not Cornell, that practiced the deception. The study was published in the prestigious *Proceedings of the National Academy of Sciences*. Certainly some researchers would argue that Cornell and *PNAS* made bad choices. Debate is needed about the wider issue of conflicts of interest generated by the interaction of non-academic data providers and academic research.

---

2. This issue arose with a recent publication (Lucca, Seru, and Trebbi 2014) that analyzed the revolving door. Although Francesco Trebbi, at a Harvard conference in 2013, orally stated that the data were from LinkedIn, the conference paper and the published version did not identify LinkedIn as the data source.

## A SUMMARY OF THE PAPERS IN THIS ISSUE

Research in political economy is increasingly focused on the role of money expenditures, as against votes, in shaping the outcomes of elections and policy. These expenditures can take the form of lobbying or campaign expenditures. Three of the papers in this issue—by Adam Bonica, by Drew Dimmery and Andrew Peterson, and by Deniz Igan—focus on political expenditures.

The consequences of political expenditures have been debated in the academic literature (compare Levitt 1994 and Erikson and Palfrey 2000). It is easy to identify cases where massive expenditures came up empty. One example is Michael Huffington's record-breaking personal expenditure of $28 million in his 1994 California Senate race against Diane Feinstein. Another is Sheldon Adelson's $140 million expenditure on the 2012 election, most of which went into Newt Gingrich's attempt to be the Republican presidential nominee. Comcast's attempt to acquire Time-Warner failed in 2015 despite massive lobbying and personal connections to the Obama administration. On the other hand, expenditure by Adelson and others is said to have forced a total alignment between the Republican congressional delegation in the United States and the Netanyahu government in Israel. Similarly, intense lobbying by hedge funds appears to have maintained the carried interest deduction in the 2012 tax bill.

We are, in terms of the research frontier, several steps away from tightly drawing the linkages between expenditures and the outcomes of elections or legislation. Research at this point, including the three papers on the subject in this issue, is more focused on the motivations and characteristics of the makers of political expenditures. At the individual level, what is the connection to income, wealth, and ideology? At the corporate level, what is the connection to firm characteristics, such as the propensity to take risks or to engage in fraud?

In his paper "Income, Ideology, and Representation," Chris Tausanovitch stresses the low level of voter information about the policy stances of their representatives. The electorate's awareness of where unelected candidates stand is quite arguably even lower. Tausanovitch also points to a very weak linkage between the policy preferences of voters in a constituency and the preferences of their representatives in Congress.

In "A Data-Driven Voter Guide for U.S. Elections," Adam Bonica develops a platform for better informing voters about candidates. So the ambition of the Bonica paper is potentially important. The paper exploits government electronic records, computing capacity, the linkage of a variety of different data sources, and text analysis, four of the important big data facets outlined earlier.

The central innovation of the Bonica paper is the use of the big data present in hundreds of millions of campaign contribution records. Informing voters in the United States is inherently a big data problem because of the decentralized aspect of both campaign finance and the political system, which only weakly controls candidate entry. In parliamentary systems, where online voter guides are important, the informational problem largely reduces to presenting the platforms of one or two handfuls of national parties. In the United States, politics can be described in one-dimensional liberal-conservative terms (Poole and Rosenthal 2007), but placing candidates on this continuum is challenging. Most candidates in an election have not previously been elected to a legislature, either because they are new entrants or because they have never won a past election. So their positions cannot be estimated from the well-established methods of roll call vote scaling developed for Congress (Poole and Rosenthal 2007) or state legislatures (Shor and McCarty 2011). But candidates—not only in federal elections but also for state legislatures and elected positions in state courts—can be placed on a common scale using the information provided by campaign contributors. If an individual, for example, contributes to a candidate for a U.S. Senate seat, a state lower house seat, and a judicial contest, the individual's contributions will provide information that glues together the continuum for two legislatures and a judicial body (Bonica 2013, 2014). More information is provided by candidates who, as is most often the case, are also contributors in other races.

To provide context to the contribution data, the platform also incorporates information from political text, election outcomes, and roll call scaling. Use of this additional information allows the platform to provide voters with information on candidate positions on specific issues. In practice, given the unidimensionality of American politics, information on specific issues is attractive in presentation but marginal in terms of information value. Bonica nicely refers to this problem as the "curse of unidimensionality."

Bonica's paper emphasizes the importance of disclosure of campaign contributions and roll call votes in providing information to the public. Disclosure, however, is not always implicit in democracy. Roll call votes in the Italian parliament were secret until 1988 (Giannetti 2010). In the United States, political expenditures by nonprofits—specifically, 501(c) organizations—are subject to minimal disclosure and have become increasingly important. Drew Dimmery and Andrew Peterson, in "Shining the Light on Dark Money," take a big data approach to identifying the political activity and expenditure of 340,000 nonprofits. The paper crosswalks government electronic websites and information from the websites of the nonprofits.

Dimmery and Peterson use automated techniques to identify the websites of nonprofits and then to scrape the websites of the organizations. They argue that the websites reveal more about these organizations than what the organizations report to the federal government or what has previously been gleaned by the Center for Responsive Politics. To ferret out political nonprofits, they match the larger set of nonprofits with a much smaller number of nonprofits whose names or IRS reports directly reveal them to be political organizations and with nearly 11,000 political action committees (PACs) that register with the Federal Election Commission (FEC). Nonprofits are deemed political when their websites use language similar to that used on the websites of known political organizations. The automated sources are validated by human evaluations that are crowdsourced.

The study is an important entry point to bring nondisclosing organizations into the disclosed world explored in the Bonica paper. For example, the websites are likely to identify the officers of the association (see the websites of Planned Parenthood and Crossroads GPS, two organizations mentioned in the paper), who in turn are very likely to have made individual political contributions. Record linkage of this type can "out" the expenditures and ideology of undisclosed nonprofits.

As we discussed earlier, a key challenge in the political economy literature is to draw a tighter connection between political expenditures and legislation or policy. One way to address this challenge is to focus on expenditure in specific industries and investigate the relationship between political spending and legislative impact. Deniz Igan takes this approach, with specific focus on the household credit, or mortgage, industry.

Focusing on the mortgage industry has some natural advantages, from both a political economy and a big data perspective. First, the financial industry is regulated in a number of different ways. The largest players in the mortgage industry—Freddie Mac and Fannie Mae—have heavy mandates from the government. There is thus a natural incentive for the private sector to try to influence the ways in which the industry is regulated. Second, large data are available for analysis, both for campaign contributions and for disbursements of mortgage credit. Igan describes a comprehensive data set on political influence exerted by financial institutions on Congress and links it to the mortgage lending activity of these institutions. She then describes the role of political influence in dictating financial regulation and credit disbursement during the U.S. credit boom of 1999 to 2006.

Results suggest that lobbying by financial institutions helped sway legislative decisions. Legislators who changed their vote in favor of deregulation under various bills were more likely to have been lobbied by the financial industry. At the same time, financial institutions that engaged in greater lobbying of the legislature were more likely to engage in risky lending behavior. For example, financial institutions that spent more on lobbying activity gave out loans with higher loan-to-income ratios, were more likely to securitize the loans, and had higher delinquency rates ex-post.

By linking lobbying and campaign contribution data with actual voting and lending behavior, Igan presents evidence that suggests that lobbying by the financial sector influences legislators' voting behavior. Moreover, the financial institutions that benefit the most from deregulation—such as subprime lenders—are more likely to devote greater resources to lobbying activity.

Another prominent topic in political economy is income inequality (see Piketty 2014, as well as the papers in the summer 2013 issue of the *Journal of Economic Perspectives*). Politics, in turn, can exacerbate income inequality if the political process overweights those with high incomes. Larry Bartels (2009) filed the opening claim that members of Congress represented the views of their rich constituents and largely ignored the views of poor ones. Bartels's methodological and measurement groups have subsequently been challenged (Bhatti and Erikson 2011; Brunner, Ross, and Washington 2013).

Tausanovitch brings big data to this problem by making substantial increases in the number of respondents used in the analysis. He estimates an item response model for 362,000 respondents. The large sample size permits analysis of the U.S. House of Representatives, whereas the earlier studies were limited to the Senate. Doing so required developing special software that took advantage of graphical processing units in desktop computers. The paper innovates in a way that goes beyond increasing sample size. Whereas Bartels and Erikson and Bhatti used responses to a single survey item, five-point or seven-point ideological self-placements, Tausanovitch applies the item response model to policy questions. He can then measure ideology on a continuum and eliminate the granularity in the other measurements. For a similar policy question approach but with smaller samples, see Stephen Jessee (2012).

The bottom line in the results is that how the distribution of income in a district influences whether Democrats or Republicans represent the district is far more important than how differences in income affect within-party representation. Moreover, the mean overall preference of the district, which is likely to have less measurement error than either the mean preference of the poor or the mean preference of the rich, is a better predictor than the mean of either group.

One limitation of the Tausanovitch study is that income is top-coded so that the "rich" in the study are all respondents reporting an income over $100,000—hardly the infamous 1 percent (Edsall 2013). One could apply the big data capacity of the Bonica study to use contributor zip codes to compute a money-weighted average ideology of contributors in a district. This might be a better measurement of the opinion of the truly rich, and it could be run through the Tausanovitch analytics.

Rather than looking at contributions, roll call voting, public opinion, or cheap talk text, the paper by Sharyn O'Halloran, Sameer Maskey, Geraldine McAllister, David K. Park, and Kaiping Chen goes directly to a policy analysis of financial regulatory structure. A major objective, shared with the Bonica and Dimmery and Peterson papers, is to replace tedious hand-coding of volumes of text with automated procedures. And volumes there are— the paper ambitiously tackles all regulatory legislation since 1950. The analytical problem has worsened over time as legislation has become increasingly wordy. (Dodd-Frank alone has over 30,000 words.) The main topics of interest, classic in the political science literature, are regulatory delegation and procedural constraints. The work shows that traditional coding and automated coding are complementary.

The authors use their processing of text to test two hypotheses: (1) that there is more discretion when the president and Congress have similar preferences or there is more market uncertainty, and (2) that higher risk aversion leads to more regulation, but with more discretion.

To summarize the methodology, O'Halloran and her coauthors started by identifying the texts of all financial regulation laws to the exclusion of those dealing with mortgage lending. The laws were then coded for delegation and procedural constraint. Both delegation and constraint were reduced to one-dimensional indexes, and discretion was measured as the product of the delegation index and one minus the constraint index. The anal-

ysis shows that discretion is least with a Democratic president and a Republican Congress.

## REFERENCES

Albertson, Bethany, and Shana Gadarian. 2014. "Was the Facebook Emotion Experiment Unethical?" *Washington Post,* July 1.

Baker, Scott. 2014. "Debt and the Consumption Response to Household Income Shocks." April. http://web.stanford.edu/~srbaker/Papers/Baker_DebtConsumption.pdf (accessed May 31, 2016).

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1): 76–91.

Bartels, Larry M. 2009. *Unequal Democracy: The Political Economy of the New Gilded Age.* Princeton, N.J.: Princeton University Press.

Bhatti, Yosef, and Robert S. Erikson. 2011. "How Poorly Are the Poor Represented in the U.S. Senate?" In *Who Gets Represented?* edited by Peter K. Enns and Christopher Wlezien (New York: Russell Sage Foundation).

Bonica, Adam. 2013. "Ideology and Interests in the Political Marketplace." *American Journal of Political Science* 57(2): 294–311.

——. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58(2): 367–86.

——. 2016. "A Data-Driven Voter Guide for U.S. Elections: Adapting Quantitative Measures of the Preferences and Priorities of Political Elites to Help Voters Learn About Candidates." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(7). doi: 10.7758/RSF.2016.2.7.02.

Bonica, Adam, Adam S. Chilton, and Maya Sen. 2015. "The Political Ideologies of American Lawyers." *Journal of Legal Analysis.* First published online October 13. doi: 10.1093/jla/lav011.

Bonica, Adam, David J. Rothman, and Howard L. Rosenthal. 2014. "The Political Polarization of Physicians in the United States: An Analysis of Campaign Contributions to Federal Elections, 1991–2012." *JAMA Internal Medicine* 174(8): 1308–17.

——. 2015. "The Political Alignment of U.S. Physicians: An Update Including Campaign Contributions to the Congressional Midterm Elections in 2014." *JAMA Internal Medicine* 175(7): 1236–37.

Brunner, Eric, Stephen L. Ross, and Ebonya Washington. 2013. "Does Less Income Mean Less Representation?" *American Economic Journal: Economic Policy* 5(2): 53–76.

Chae, David H., Sean Clouston, Mark L. Hatzenbuehler, Michael R. Kramer, Hannah L. F. Cooper, Sacoby M. Wilson, Seth I. Stephens-Davidowitz, Robert S. Gold, and Bruce G. Link. 2015. "Association Between an Internet-Based Measure of Area Racism and Black Mortality." *PLOS ONE* (April 24).

Dimmery, Drew, and Andrew Peterson. 2016. "Shining the Light on Dark Money: Political Spending by Nonprofits." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(7). doi: 10.7758/RSF.2016.2.7.04.

Dobbie, Will, and Jae Song. 2015. "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection." *American Economic Review* 105(3): 1272–1311.

Edsall, Thomas B. 2013. "When Class Trumps Identity." *New York Times,* October 29. http://www.nytimes.com/2013/10/29/opinion/edsall-when-class-trumps-identity.html (accessed May 23, 2016).

Erikson, Robert S., and Thomas R. Palfrey. 2000. "Equilibria in Campaign Spending Games: Theory and Data." *American Political Science Review* 94(3): 595–609.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Newspapers." *Econometrica* 78(1): 35–71.

Giannetti, Daniela. 2010. "Secret Voting in the Italian Parliament." Paper presented at the annual meeting of Rationalité et Sciences Sociales. College de France, Paris (June 3–4).

Griffin, John M., and Gonzalo Maturana. 2013. "Who Facilitated Misreporting in Securitized Loans?" Working paper. University of Texas, Austin.

Hersh, E. D., and M. Goldenberg. 2016. "Political Spillover Effects on Physician Clinical Practice." Working paper. Yale University, New Haven, Conn.

Igan, Deniz. 2016. "Home Truths: Promises and Challenges in Linking Mortgages and Political Influence." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(7). doi: 10.7758/RSF.2016.2.7.05.

Imai, Kosuke, James Lo, and Jonathan Olmsted. 2015. "Fast Estimation of Ideal Points with Massive Data." Working paper. Princeton University.

Jessee, Stephen A. 2012. *Ideology and Spatial Voting*

*in American Elections.* New York: Cambridge University Press.

Keys, Benjamin J., Amit Seru, and Vikrant Vig. 2012. "Lender Screening and the Role of Securitization: Evidence from Prime and Subprime Mortgage Markets." *Review of Financial Studies* 25(7): 2071–108.

Kim, In Song. 2014. "Intra-industry Trade and Trade Liberalization: Evidence from Dyad-Level Tariff Data." Working paper. Massachusetts Institute of Technology, Cambridge.

Levitt, Steven D. 1994. "Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House." *Journal of Political Economy* 102(4): 777–98.

Lewis, Jeffrey B., and Chris Tausanovitch. 2015. "When Does Joint Scaling Allow for Direct Comparisons of Preferences?" Working paper. University of California, Los Angeles.

Lucca, David, Amit Seru, and Francesco Trebbi. 2014. "The Revolving Door and Worker Flows in Banking Regulation." *Journal of Monetary Economics* 65: 17–32.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge, Mass.: MIT Press.

Mian, Atif, and Amir Sufi. 2014. *House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again.* Chicago: University of Chicago Press.

———. 2015. "Fraudulent Income Overstatement on Mortgage Applications During the Credit Expansion of 2002 to 2005." Working paper. Princeton University and University of Chicago.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics* (July 25). doi: 10.1093/qje/qjt020.

Mian, Atif, Amir Sufi, and Nasim Khoshkhou. 2016. "Government Economic Policy, Sentiments and Consumption." *NBER* working paper.

Mian, Atif R., Amir Sufi, and Francesco Trebbi. 2010. "The Political Economy of the U.S. Mortgage Default Crisis," *American Economic Review* 100(5): 1967–98.

O'Halloran, Sharyn, Sameer Maskey, Geraldine McAllister, David K. Park, and Kaiping Chen. 2016. "Data Science and Political Economy: Application to Financial Regulatory Structure." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(7). doi: 10.7758/RSF.2017.2.6.06.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century.* Cambridge, Mass.: Belknap Press of Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998. *Quarterly Journal of Economics* 118(1): 1–39.

Poole, Keith T., and Howard L. Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35(1): 228–78.

———. 2007. *Ideology and Congress.* New Brunswick, N.J.: Transaction Publishers.

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3): 530–51.

Tausanovitch, Chris. 2016. "Income, Ideology, and Representation." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(7). doi: 10.7758/RSF.2016.2.7.03.

Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* New Haven, Conn.: Yale University Press.