Journal of Information Technology & Politics, 6:166–175, 2009

ISSN: 1933-1681 print/1933-169X online
DOI: 10.1080/19331680902821569

Political Science Computing: A Review of Trends in Computer Evolution and Political Science Research

Euel W. Elliott
Karl Ho
Jennifer S. Holmes

ABSTRACT. The increase in computing power in the last four decades has had an enormous impact upon the way political scientists conduct research. In this article, we review the development of computing resources and the lifting of computational limitations of the use of political methodologies. Using descriptive analysis, we demonstrate that Moore's Law, or the doubling of power in computing hardware every 18 months, is closely associated with the adoption of advanced statistical methods and database availability in political science research. Our findings are surely not startling to researchers, but our analyses do highlight and provide important details as to the evolution of the profession over the last few decades. We aim to provide hints as to the future direction of political science research as computational capabilities continue to grow.

KEYWORDS. Electronic library database, Moore's law, multivariate models, political methodology

This article explores the synergistic relationship between computing capacity and the way it has shaped political science research over the past four decades. The article examines the

Euel Elliott is Professor of Public Policy and Political Economy in the School of Economic, Political and Policy Sciences at the University of Texas at Dallas, where he also serves as Senior Associate Dean for Academic Programs. Dr. Elliott's teaching and research interests are in the areas of U.S. public policy (general), institutions and public policy, and the application of the theory of complex adaptive systems to understanding political and social phenomena. He has published in a wide array of journals and is the author, with Mark Dobeck, of *Money* (Greenwood Press, 2007).

Karl Ho (PhD, University of North Texas) is Director of Academic Computing at the University of Texas at Dallas School of Economic, Political and Policy Sciences. He also serves as a Senior Lecturer teaching political science and research methods. His research covers comparative politics, democratization and peace process, and elections and politics in East Asia. He has also taught research methods and survey research across multiple social science and medical disciplines. He also gives courses on statistical programming in multiple programs and languages. His works have appeared in *Human Rights Quarterly*, *Electoral Studies*, and *Journal of African and Asian Studies*.

Jennifer S. Holmes (PhD, University of Minnesota) is an Associate Professor of Political Economy and Political Science at the University of Texas at Dallas. Her major areas of research are political violence, terrorism, and political development with an emphasis on Latin America and Southern Europe. In addition to publishing in numerous journals, she is the author of *Terrorism and Democratic Stability* (Manchester University Press, 2001; Transaction, 2006), *Terrorism and Democratic Stability Revisited* (Manchester University Press, 2008), and *Guns, Drugs & Development in Colombia* (University of Texas Press, 2008), and the editor of *New Approaches to Comparative Politics: Insights from Political Theory* (Lexington Books, 2003, 2008).

Address correspondence to: Euel W. Elliott, School of Economic, Political and Policy Sciences, University of Texas at Dallas, 800 W. Campbell Rd, MS GR31, Richardson, TX 75080-3021 (E-mail: eelliott@utdallas.edu).

evolution of quantitative methodologies as well as the impact of the dramatic increase in access to, and proliferation of, electronic journals and databases. We analyze the articles published in a top political science journal, particularly focusing on the statistical methods in these studies, and examine the impact of Moore's Law on the trends of political research.[1]

The discipline of political science has changed dramatically over the past half-century. Empirical research methodologies, along with other research approaches involving highly formal mathematical models of political behavior, have become more common as limits on computing capacity have diminished. Leading journals, including the flagship journal of the American Political Science Association, *The American Political Science Review* (*APSR*), and other top journals, such as the *American Journal of Political Science* (*AJPS*), reflect these trends.

The advent of the computer in the late 1940s was an epoch-making event, on a par with the discovery of DNA by Watson & Crick in 1953. The ability to manipulate and analyze massive amounts of data opened up incredible new horizons for researchers in all of the social sciences. Problems of data comparability and quality remain issues to be dealt with, but computational limitations have decreased. Of course, those of us of a certain age are cognizant of the rapid evolution of computational capabilities over the last few decades.

The traditional image of the computer as a massive "mainframe," housed in the generic and ubiquitous "Computing Services" department and serviced by squadrons of technicians, comes readily to mind. Over the past decades, by virtue of the steady increase in computational power resulting from the ability to steadily pack more and more transfers on a microchip, roughly a doubling of computation capacity has occurred every 18 to 24 months. Such expansions in power have resulted in an increase of more than one billion (yes, billion)-fold over the course of the last 35 years.

Expansions in computational capabilities have allowed for a range of powerful statistical techniques on large data sets that would have otherwise proven impossible. Simple bivariate or multivariate regression models that would use enormous computing resources a few decades ago are done today in the twinkling of an eye. Today highly complex analyses are performed in seconds on laptop computers weighing less than two pounds, while desktop computers' power has grown in commensurate fashion. Below we discuss the impact of advances in computing on empirical research in political science, and conclude with a brief assessment of future prospects for political science, given the likelihood of continued advances.

COMPUTING POWER AND ITS IMPACT

Moore's Law, which says that computing power doubles every 18 months, influences political science research in two directions. Faster computers with cheaper and more abundant memory and disk space not only propel and extend the economic boom the United States enjoys, but also the way scientists conduct research. Computationally intensive models that most researchers used to avoid have now become the convention. The fast growth of microchip central processing units (CPUs), along with parallel development in memory and storage, weans individual users from centralized computing that feature mainframe computers and round-reel magnetic tapes. For instance, as of 2005, a standard medium-range desktop computer can outrun by more than 10 times 1976's fastest supercomputer. The original supercomputer Cray-1 at the Los Alamos National Laboratory, fastest in its time, was capable of 100–250 MFLOPS (a million floating point unit operation per second), compared to most current (2005) desktop computers measuring in GFLOPS or thousands of MFLOPS. Decentralization of research computing took place in the early 1990s. Users who used to rely on terminals to submit Job Control Language (JCL) batch statistical programs can now run most of these programs on desktop or even laptop personal computers. Computationally more demanding jobs can still be run in batches submitted to UNIX boxes (formerly called minicomputers), but merely for background processing in most cases, in order to save local computer resources. Most political scientists, like their counterparts in other social sciences, can nowadays run research programs entirely independent of centralized computing centers.

Consistent with Moore's Law, along with increased computational capacity there has, not surprisingly, been more interest in sophisticated and computationally demanding statistical methodologies. Ordinary least squares (OLS) regression was considered "cutting edge" in the 1960s and into the 1970s, and is still a very common technique for well behaved interval and even ordinal-level data. It has since been supplemented by an array of methodologies relying upon far more intensive approaches using maximum likelihood estimation (MLE) and simulation-based estimations. Standard probit, logit, and later, multinomial logit, and nested or mixed logit models have become relatively common in the empirical research work in political science.

Moreover, whereas nearly all of the empirically grounded political science publications in the 1960s and early 1970s utilized relatively simple cross-sectional data, new computational capabilities allowed increasingly for sophisticated examination of time series data using a wide range of models such as generalized least squares (GLS) and auto-regressive moving average models (ARMA and ARIMA) that had been popularized by Box and Jenkins (1970). The particular properties of time series data and the unique problems often associated with their external structures would have remained opaque without the continued expansion in computation capacity.

Over the last few years, Bayesian econometrics has increasingly caught the attention of political scientists. Challenging the standard "frequentist" assumptions of normally interpreted probabilities, Bayesian statistical theory has been increasingly successful in forcing a rethinking of many assumptions in political science. Works by scholars such as Brandt and Freeman (2006), Gelman and King (1994), Western and Jackman (1994), Martin (2001), and others are many times more demanding in computational requirements than the statistical models of a few decades ago.

The development of increasingly demanding computational capabilities has also allowed for the emergence of increasingly sophisticated simulation methodologies. Simulation methodologies have been growing in influence in the social sciences, when other more traditional computational approaches have been found to be of limited usefulness.

Even with a standard desktop computer, political scientists can run replications and simulations to validate models in a split second. Simulation-based statistical techniques such as bootstrapping and the Monte Carlo and Markov chain models are not new and unique to modern day researchers. When W. S. Gosset popularized his famous Student's t distribution in 1908, he bolstered his faith with a validation method involving random number experimentation. Such a method was promising at the time, but lacked vigor due to slow and formidable manual computations. This simulation method, which was later formally named Monte Carlo in the 1950s, remained too costly to be practical and applicable in most of the 20th century until the 1970s (Ulam, 1976). The advent of a modern, fast speed microcomputer chip ushered in a new phase of computational statistics. Today, most researchers employing statistical models can solve and validate stochastic problems with the Monte Carlo method and other simulation-based techniques. In fact, any undergraduate student can easily perform thousands of simulations in seconds using built-in features of most statistical packages and a moderate rank personal computer. Only two decades or so ago, such endeavors would have taken months of manual calculations, usually done by brave and relentless graduate students. Another critical event in the boom of simulation-based modeling is the introduction of estimation tools based on Markov chains: the Metropolis–Hastings algorithm and the Gibbs sampler developed in disciplines of chemistry and engineering (see Gill, 2004).

One simulation approach involves what is commonly referred to as agent-based modeling. Agent-based models (ABM) allow for individual "agents" or single software programs to interact with each other based on simple rules governing their behavior. Such computational innovations have proven useful in attempting to model highly complex, "uncertain" systems. Typically known and referred to as complex adoptive systems (CAS), they are characterized by multiple, interacting elements that are capable of interacting in unpredictable and nonlinear ways.

IMPACT ON RESEARCH METHODS

A simple tally of the articles published in *American Political Science Review* and *American Journal of Political Science* witnesses the "power" of Moore's Law in political science research. In the 1970s, only about 10 articles applied or projected the use of simulations, Monte Carlo experiments, Markov Chain Monte Carlo methods, or Bayesian models (see Table 1). The number jumps to 15 and 25 in the 1980s and 1990s, respectively, while the number of published articles remains generally stable.[2] Noteworthy is the tremendous growth in the first decade of the new millennium. Twenty-six studies have already been published in the decade's first half!

To further assess the development of political research propelled by the Moore's Law, we went through each issue of the *APSR* from 1970 to 2006, and coded the articles as to whether a quantitative approach or approaches were employed and the type of methodology used.

TABLE 1. Computationally Intensive Models Published in *American Political Science Review* and *American Journal of Political Science* Since 1970s

Decade	Number of studies
1970	10
1980	14
1990	25
2000–2006	26

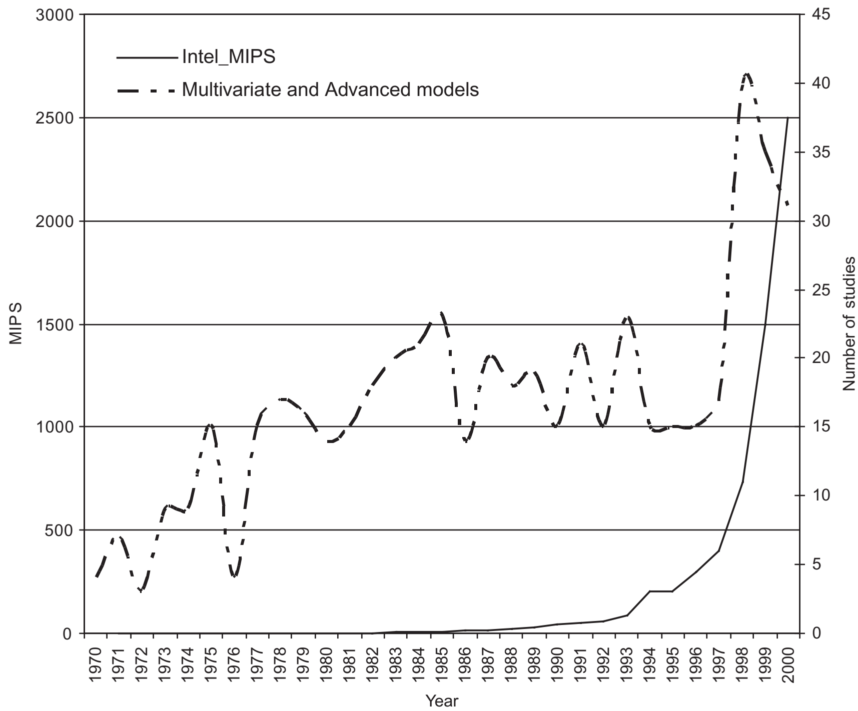
Note: Tally based on search of key words including "simulation," "MCMC," "Monte Carlo," and "Bayesian" in abstracts of *American Political Science Review* and *American Journal of Political Science*.

We were primarily concerned with distinguishing articles that would require the exploitation of computing resources, at whatever level, versus those that did not. A total of 1,330 articles were coded from the *APSR*. We initially coded for the categories of (1) Theory (Normative and Positive/Formal) that would not require computing power, as well as historical descriptive prices versus those analytical approaches that would require computational ability.[3] For the Positive/Formal category, we sought to distinguish between (2) purely descriptive approaches; (3) bivariate measures of association and correlation, such as crosstabular analysis; (4) basic multivariate models, including multiple regression using OLS, logistic regressions, and other models that are applicable to studies where a dichotomous or other limited choice-dependent variables may suggest analytic approaches other than standard OLS regression; and (5) advanced multivariate models that are computationally intensive. The methodologies in the fourth category included probit, tobit (where censoring occurs), multinomial logit, and factor analysis. This final category included, in addition to multinomial probit models, structural equation modeling and various time series approaches including but not limited to Box-Jenkins, London School of Economics (LSE), and pooled cross-sectional time series analyses.[4]

To operationalize the independent variable, we employ the variable of millions of instructions per second (MIPS) of Intel's CPUs to track computing power change.[5] The variable provides a measure of computer processor speed, which captures the broad contours of the gains in computing power over the past three decades.

Figure 1 illustrates the evolution of methodological approaches in political science over the past three decades. First, observe the two trend lines in Figure 1. The solid line shows the increase in computing power, measured by MIPS, from 1970 to 1999. It exhibits a near exponential increase in computing power, with the inflection point beginning around 1990. Incredibly, computing power as measured by MIPS soared from fewer than 50,000 operations per second in 1970 to more than 30 million operations per second by 1999, with an enormous

FIGURE 1. Computer power and methods used in *American Political Science Review* studies, 1980 to 2000.



spurt in computing power in the mid and late 1990s (from about 3 MIPS in 1995 to the 30 million just noted above).

The dashed line depicts the trends in multivariate statistical techniques reported in the *APSR* from 1970 to 1999. Although a dramatic increase in number of multivariate and advanced multivariate studies occurs well before the explosion in computing ability (as witnessed by the solid line) that takes place around 1990, there is nonetheless a fairly clear relationship. There is a general upward drift in these studies through the decade of the 1970s and 1980s, ending with an upward spurt in the mid to late 1990s before tapering off slightly in the year 2000.

Table 2 provides a breakdown of the number of articles, including non-quantitative theory pieces (including positive theory), for the period 1970–1999. Interestingly, theory articles have more than held their own across the period of

TABLE 2. Methods Used in *American Political Science Review* by Decade, 1970s to 1990s

			Method					Total
			Theory	Descriptive	Bivariate	Multivariate	Advanced multivariate	
Decade	1970	Count	183	65	124	87	6	**465**
		Row %	39.4%	14.0%	26.7%	18.7%	1.3%	**100.0%**
	1980	Count	179	39	49	160	11	**438**
		Row %	40.9%	8.9%	11.2%	36.5%	2.5%	**100.0%**
	1990	Count	199	21	19	165	23	**427**
		Row %	46.6%	4.9%	4.4%	38.6%	5.4%	**100.0%**
Total		Count	**561**	**125**	**192**	**412**	**40**	**1330**
		Row %	**42.2%**	**9.4%**	**14.4%**	**31.0%**	**3.0%**	**100.0%**

the study. More than 39 percent of all articles were designated theory in the 1970s, increasing to 41 percent in the 1980s, and more than 46 percent in the 1990s.

At the same time, and not surprisingly, articles using purely descriptive techniques dropped precipitately from 14 percent of the total in the 1970s to 9 percent during the 1980s and less than 5 percent by the 1990s. Clearly, the expectations of editors (and readers) of what an acceptable quantitatively oriented article would require have changed. Similarly, a dramatic decline from 27 percent in the 1970s to less than 4 percent in the 1990s is observed for simple bivariate techniques.

During the same period, multivariate studies soar in the percentage of their use, going from nearly 19 percent in the 1970s to more than 36 percent in the 1980s. Interestingly, according to our measure of multivariate approaches, only modest increases in multivariate approaches were observed during the 1990s, with nearly 39 percent of articles so identified.[6] Advanced multivariate techniques, while showing an increase of 400 percent from the 1970s to the 1990s, nonetheless accounted for only 5.4 percent of all articles in the 1990s.

While recognizing the growth of sophisticated models, we have to give a caveat on measurement of our four-point scale. It is not perfect, and in at least two aspects, it is far from being an ideal gauge of measuring advances of CPU use. One aspect we have yet to factor in is the programming costs that have been alleviated by widely available statistical software packages, such as Stata and R. This development—regardless of whether it is driven directly or indirectly by Moore's Law—allows researchers much greater opportunities to use sophisticated methods that were previously expensive to program and execute, and in turn render the distinctions across the scale more difficult from time to time. For instance, writing a program for multinomial logit models in the late 1980s could be costly on both CPU and programming prices, but nowadays these models can be easily executed using a point-and-click approach. Aside from the "opportunities" factor, we also acknowledge that even within a certain category, there are cases where we are unable to measure the CPU power use, particularly in the top category. Two models, for example, can both be coded in the top category, but one can be accomplished within seconds while another may take days to complete.

This review of the trends of political methods focuses on the growth of computing power and information technology. With the introduction of the new CPUs such as the Pentium chip (circa mid-1990s) and other innovations, more and more researchers have adopted advanced multivariate methods rather than simple bivariate tests such as t-test and one-way ANOVA. After 2000, few studies published in *APSR* used methods coded in our lower categories (see Appendix A). Evidence from the early 2000s suggests a complete regime change in terms of the usage of a variety of mixed-method approaches. Expansion beyond 2000 would require a new scale that further subdivides category 4 and develops a more sophisticated classification by computing power. However, this new classification would not help illuminate the changes that began in the 1970s.

ANALYSIS ON DATABASE USAGE AND AVAILABILITY

Concurrently with how technology has changed the statistical tools available to political scientists, numerous innovations have changed the way in which political scientists investigate. For instance, online forms of research tools open up possibilities for more efficient searching of the secondary literature, both in the discipline and across the social sciences. Additionally, new databases are expanding what is available to be empirically investigated.

Data and research tools have significantly changed over the last fifty years. Since 1890 and until 1951, punch cards were the standard for data analysis computing. After that, the U.S. Census Bureau began using electronic files on mainframe computers (Thomas, 2002, p. 192). In 1962, the Inter-university Consortium for Political and Social Research (ICPSR) was established, with the aim of creating an archive of electronic data. In the 1980s, the Internet emerged, and paper indexes began a transition to CD-ROMs. In the mid-1980s, it was estimated that more than 2,800 electronic databases were available (Williams, 1985, p. 445). Within a decade, more than 50,000 electronic databases were available online (Pool, 1993, p. 841). By the 1990s, many of these electronic databases were increasingly linked to full-text sources (Herman, 2001, p. 433).

Extensive research involves "comprehensively assembling relevant data and information, in-depth analysis of the information, and identifying relations among the concepts in the related and even apparently unrelated subjects" (Neelameghan & Iyer, 2004, p. 115). The shift to electronic journals and indices has reduced the difficulty of doing so. Online research increases access to sources and information and promotes sharing, cooperation, and training (Trenton, 1999), and the growth of such databases is virtually certain to continue into the 21st century.

So far, definitive studies of the effects of electronic journals and databases on objective measures of research productivity are unclear, since, among other reasons, younger scholars tend to use them more (Tenner & Yang, 1999, p. 5). However, survey research demonstrates that most users agree that electronic searches are in many ways superior. Two important studies (Stanford University Libraries eJournal User Study, 2002a, 2002b; Wenger, 2002) suggested that there are very substantial advantages to the use of electronic databases as opposed to traditional search techniques. Users overwhelmingly report a shorter retrieval time, with 98 percent of respondents giving an affirmative response and with a similar 99 percent indicating an overall savings in time (Wenger, 2002). Other important indicators of the superiority of electronic databases are that 94 percent of respondents claim an enhanced ability to conduct research, 81 percent state that e-databases help prevent the duplication of previously published research, and 52 percent assert an enhanced ability to organize information. While admittedly not comprehensive in scope, the two surveys also indicated two additional potentially crucial features of electronic databases. These features are the likelihood of reading papers outside the discipline, which enhances the possibilities for inter- and multidisciplinary approaches to research, and the likelihood of enhancing the possibilities for exchanges with colleagues. On both counts, 71 percent of respondents indicated an affirmative response. Numerous other studies have documented the utility of interdisciplinary databases and multiple searches (Schaffer, 2001, p. 53; Walters & Wilder, 2003). Clearly, these studies are not conclusive, since there is an obvious bias toward younger researchers, as they are more likely to use electronic databases (Tenner & Yang 1999, p. 5). However, these studies do suggest some intriguing possibilities for new modalities of electronic research in the coming years, as software and hardware systems continue to advance.

Some of the databases may be new and innovative. For example, Yu, Kaufmann, and Diermeier (2008) use automatic speech classification systems to examine Congressional speeches. They note that such analysis was difficult if not impossible until recently, due to the previous absence of rigorous methods for extracting and processing relevant information (also see Guerini, Strapparava, & Stock, 2008). Similarly, van Atteveldt, Kleinnijenhuis, Ruigrok, and Schlobach (2008) use a "machine learning approach" to classify relations, performance descriptions, and evaluative descriptions in their semantic analysis, which would be unmanageable without both the increased computing power and the availability of relevant databases. Dyson (2008, p. 8) highlights how advances in text analysis, using automation, coupled with the "ready availability of the large volumes of text produced by leaders and the increasing ease with which text can be accessed online, substantially mitigates the traditional problems of access." In this case, he is referring to the study of elites and the use of political psychology, but other uses are becoming more common as well. Demiros, Papageorgiou, Antonopoulos, Pipis, and Skoulariki (2008) examine the usefulness of multimedia databases and media monitoring in concert with new technologies of access and analysis.

CONCLUSION

The availability of computing resources and computing power have increased enormously over the last three or four decades. Statistical techniques and methodologies that just 20 years ago would have required supercomputing facilities for their solution can today be solved on small laptop computers. It is truly a different world from the days when many of us took stacks of punch cards to the computing center, then waited, in some cases for a day or more, on the results.

The increase in computing power has had a massive impact on the kinds of methodological approaches employed by political scientists. In this study, while we are not seeking to establish a causal relationship, we demonstrate the strong association between computing power and use of computationally intensive methods and computerized library databases. Although there are other variables that may affect the adoption of innovative methodologies, in this article we focus on Moore's Law, which provides opportunities for researchers to move forward quantitative research with advanced methods that are not easily implemented without fast, powerful computers. The transformation from simple, descriptive statistics, primarily observed in the 1970s, to multivariate regression techniques that were deployed in the 1980s, to yet more sophisticated maximum likelihood and other estimation techniques that gained popularity in the 1990s tracks very well with the gains in computational capabilities observed over this period of time.

Our findings are surely not startling to anyone having any familiarity with the direction that political science has taken in the last several decades. But our analyses do highlight and provide important details as to the evolution of the profession. At the same time, we certainly are aware that our analysis is limited to the extent that we have relied exclusively upon an analysis of the *APSR*. But, while employing other journals such as *AJPS* might yield slightly different results, we are nonetheless confident that our analyses are robust enough to capture the broad trends extant in the political science profession.

While it is always risky to attempt to predict the future, we have little reason to doubt that the relationship between computing technology and computing capabilities on the one hand, and political science research on the other will become bound even more closely together. Many of the changes will result from the enormous expansion in the ability to store data in a convenient and easily accessible manner. In just a decade, we have progressed from card punch systems for database storage to mainframe computers and CD-ROMs. As just a single case in point, we indicate that by the 1990s, electronic databases were available and typically linked to full-text sources (Herman, 2001, p. 433). Clearly, the proliferation of electronic databases with extraordinary large data storage capacity will dramatically promote and encourage the use of sophisticated and computationally demanding methodologies with a wide array of research applications at the same time that more and more data are available for empirical analysis (see King, 2009).

What might these methodologies look like? What quantitative approaches in political science will dominate the journals over the next decade or so? Clearly, prediction is chancy at best. With that disclaimer, we think the future of quantitative research will move in several directions. First and foremost, we would expect, in part, a continuation of current trends. In other words, the use of sophisticated maximum likelihood techniques applied to larger and larger data sets will be the norm. Logit, multinomial logit, censored probit, and related approaches will be standard, but rather than analyses utilizing a few thousand cases, as is typically the case today, data sets with hundreds of thousands or even millions of cases may not be unusual. Survey researchers, for instance, with access to large public opinion in voting studies may find these methodologies very useful. Others may work with time series data that take advantage of a similar abundance of data points. As a hypothetical example, those working in the area of public opinion and attitudes might find useful physiological data in heart rate, blood pressure, and the like when literally millions of data points are available over some extended period of time. This could open up a new era for time series, ARIMA, and other kinds of time series approaches to analysis.

Other newer methodologies will also become far more prominent. Economists have for the last decade or so been taking advantage of Monte Carlo techniques in various types of econometric modeling. This will also become more prevalent in political science, as Monte Carlo simulations software and access to large databases become increasingly common. Such simulations may provide much more sophisticated approaches to understanding an array of political science questions. Another methodological innovation, Bayesian analysis, has become much more accepted in the social sciences in recent years. Recent advances in computational capacity should open new doors to the widespread adaption of Bayesian techniques in political science, as such approaches have been heretofore constrained by the relative lack of computing power.

Finally, we expect that more radical quantitative methodologies will play a far more dominant role in the political science research of the future. It is well-known that the theory of complex adaptive systems, which makes certain assumptions about the role of nonlinear feedback and other processes has made considerable progress in explaining physical phenomena, and increasingly a variety of social phenomena. Miller and Page (2007), for example, demonstrate that concepts such as self-organized criticality, positive feedback, sensitivity to initial conditions, and other important concepts can be effectively modeled. These modeling exercises require in many cases very substantial applications of computing power, since the aim of the modeling exercise is to create "self-contained worlds" in which artificial agents, acting on the basis of simple rules governing their behavior and interaction with each other, produce complex and variegated social phenomena. Since thousands or even hundreds of thousands of parameters must be estimated in such simulations, computing power is at a premium. Such computationally demanding massive computer simulations are not unlike the efforts by climatologists or cosmologists to model the dynamics of weather and climate, or the evolution of the universe. Such approaches in political sciences and the social sciences more generally may yield important new insights into the underlying dynamics of behavior. Other approaches, including the application of classifier systems utilizing neural networks and other computing-intensive approaches, are likely to make their way into the political science toolkit over the next several years.

NOTES

1. Replication data are available at the JITP Dataverse referenced by Euel W. Elliott, Karl Ho, Jennifer S. Holmes, February 12, 2009, "Replication data for: Political Science Computing: A Review of Trends in Computer Evolution and Political Science Research," hdl:1902.1/12360. Available at: http://hdl.handle.net/1902.1/12360.

2. For instance, number of articles published in APSR in 1970s, 1980s, and 1990s are 465, 438, and 427, respectively.

3. We include formal model articles in the theory category since these studies do not usually involve statistical methods. Given the recent boom of formal studies, this category actually sees an increase in proportion compared to the other categories.

4. When considering the use of such a scale, we weight between the ordinality (real order of scale) and generality (practical use), acknowledging that classification of methods can be fuzzy. In fact, certain studies may use a third category method (e.g., probit) with combined use of other more computationally intensive method which should fall into the highest category. The general rule is if more than one method is involved, it is coded using the more computationally intensive method. Otherwise, in a case not clearly stated, we adopt the more conservative approach in giving a lower score.

5. We have considered using other measures such as number of transistors and clock frequency, but MIPS proves to be a more valid measure that provides the speed and capacity to accommodate computational procedures actually involved in statistical methods.

6. This could be attributed to a surge of formal theory articles in the 1990s that attenuated the growth of the multivariate studies.

REFERENCES

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Brandt, P. T., & Freeman, J. R. (2006). Advances in Bayesian time series modeling and the study of politics: Theory testing, forecasting, and policy analysis. *Political Analysis*, *14*(1), 1–36.

Demiros, I., Papageorgiou, H., Antonopoulos, V., Pipis, A., & Skoulariki, A. (2008). Media monitoring by means of speech and language indexing for political analysis.

Journal of Information Technology & Politics: JIPT, *5*(1), 133–146.
Dyson, S. B. (2008). Text annotation and the cognitive architecture of political leaders. *Journal of Information Technology & Politics*, *5*(1), 7–18.
Gelman, A., & King G. (1994). Enhancing democracy through legislative redistricting. *American Political Science Review*, *88*(3), 541–559.
Gill, J. (2004). Introduction to the special issue. *Political Analysis*, *12*(4), 323–337.
Guerini, M., Strapparava, C., & Stock, O. (2008). CORPS: A corpus of tagged political speeches for persuasive communication processing. *Journal of Information Technology & Politics*, *5*(1), 9–32.
Herman, E. (2001). End users in academia. *Aslib Proceedings*, *53*(10), 431–457.
King, G. (2009). The changing evidence base of social science research. In G. King, K. Scholzman, & N. Nie (Eds.), *The future of political science: 100 perspectives* (pp. 91–93). New York: Routledge Press.
Martin, A. D. (2001). Congressional decision making and the separation of powers. *American Political Science Review*, *95*(2), 361–378.
Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computation models of social life*. Princeton, NJ: Princeton University Press.
Neelameghan, A., & Iyer, H. (2003). Information organization to assist knowledge discovery. *Cataloging & Classification Quarterly*, *37*(1/2), 115–130.
Pool, R. (1993). Beyond databases and e-mail. *Science*, *261*(August), 841–843.
Schaffer, T. (2001). Databases and political science research. *Online Information Review*, *25*(1), 47–53.
Stanford University Libraries e-Journal User Study. (2002a). Retrieved November 11, 2007, from http://ejust.stanford.edu/findings/report_survey1.html#part1.
Stanford University Libraries e-Journal User Study. (2002b). Retrieved November 11, 2007, from http://ejust.stanford.edu/findings/table7.pdf.
Tenner, E., & Yang, Z. Y. (1999). End user acceptance of electronic journals. *Technical Services Quarterly*, *17*(2), 1–14.
Thomas, W. L. (2002). Social science data and the digital library. In C. F. Thomas (Ed.), *Libraries, the Internet, and scholarship tools and trends converging* (pp. 191–203). New York: Dekker.
Trenton, G. (1999). What does using the Internet for education mean? *Educational Technology*, *39*(4), 15–23.
Ulam, S. M. (1976). *Adventure of a mathematician*. New York: Scribner.
van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., & Schlobach, S. (2008). Good news or bad news? Conducting sentiment analysis on Dutch text to distinguish between positive and negative relations. *Journal of Information Technology & Politics*, *5*(1), 73–94.
Walters, W., & Wilder, E. (2003). Bibliographic index coverage of a multidisciplinary field. *Journal of the American Society for Information Science and Technology*, *54*(14), 1305–1312.
Wenger, C. (2002). The value of multiple database searching to scholarly research productivity. *Science and Technology Libraries*, *23*(1), 17–29.
Western, B., & Jackman, S. (1994). Bayesian inference for comparative research. *American Political Science Review*, *42*(4), 412–423.
Williams, M. (1985). Electronic databases. *Science*, *228*(4698), 445–450.
Yu, B., Kaufmann, S. & Diermeier, D. (2008). Classifying party affiliation from political speech. *Journal of Information Technology & Politics*, *5*(1), 33–48.

APPENDIX A

Classification Scheme of Statistical Methods

Category	Methods
1	Descriptive statistics
	Probability
2	t-test/mean difference test
	Correlation analysis
	Cross Tabulations
	ANOVA
	Chi-square test
	z-test
3	Factor analysis
	Regression
	Logistic regression
	Probit
	Tobit
	Log linear models
	Multinomial logit
	Principal components
4	Multinomial probit
	LISREL
	Structural equation models/LISREL
	Time series analysis
	Pool cross-sectional time series analysis
	Bayesian models