A Methodological Education

Simon Jackman[1]

August 19, 2004

During my undergraduate studies at the University of Queensland, David Gow told me that I should go to America for graduate study. The Australian system, he argued, was a copy of the British system: after an undergraduate Honors degree one sat at the feet of one's advisor and one learned whatever it was he or she was yet to teach you, and hence was narrow, *ad hoc*, and could be a hit-or-miss affair. The American system, with its emphasis on graduate coursework and exams, generated excellence more reliably, and was the option I ought to pursue. I followed David's advice, and benefited tremendously. All the same, my training in political methodology seems to owe much to luck, to simply being around the right people at the right time.

Undergraduate Training: After three years as an undergraduate I had become deeply skeptical about the use of quantitative methods to study politics. I vividly remember two reactions while leafing through the *APSR*. One was purely ideological and literally sophomoric, a reflection of the content of my undergraduate studies up until then: models, tables, graphs and – of all things – survey data were staples of the hegemonic, neo-liberal program, and, as such, to be regarded with suspicion, if not derision and outright hostility. On the other hand (and this was not a popular view with most of my classmates), I was impressed by the seriousness with which American political science appeared to conduct itself. Staring at the pages of the *APSR* it dawned on me that vast numbers of (predominantly American-based) academics were building careers in pursuit of the notion that politics could be studied with categories and theories that transcended particulars of time and place, and that statistical modeling was a fruitful way to go about this.

Then I got to meet one of these people. In 1986 David Gow took up an appointment at the University of Queensland. Originally from Sydney, David had spent about 14 years in American political science, both as a graduate student and most recently as a professor.[2] Gow was and is a True Believer, one of the small group that helped found the Political Methodology section in the early 1980s and a student of the history of political methodology.[3] My friend Bruce Western (now Professor of Sociology at Princeton) and I were intrigued; here was Gow, an Australian, but conforming almost exactly to our imagined stereotype of

[1]Department of Political Science, Stanford University. e-mail: jackman@stanford.edu

[2]Bruce Stinebrickner, a Yale PhD (and co-incidentally, a classmate of Chris Achen and Neal Beck), was Head of Department at UQ at the time, and was instrumental in bringing Gow to Queensland.

[3]e.g., David John Gow, "Quantification and Statistics in the Early Years of American Political Science, 1880-1922", *Political Methodology*, V11 (1985): 1-18.

one of those *APSR* characters. Closer inspection revealed that in fact, Gow *is* one of those *APSR* characters![4] Gow was invited to give a seminar in the Honors sequence at UQ; later I would recognize his talk a masterful introduction to what is often called scope and method in political science PhD programs. Over the next year or so, in a series of impromptu late night tutorials, Gow led me on a tour spanning Karl Popper, Fisher, Neyman, Pearson, and Przeworski and Teune.

By now I was hooked. At Gow's urging, I took a class in econometrics and immersed myself in the American literature on voting behavior, eventually writing an undergraduate Honors thesis under David's supervision. Late night sessions on statistics and computation were also incredibly valuable: Gow was writing a companion computing volume to the 2nd edition of Judge et al.'s *Introduction to the Theory and Practice of Econometrics*, implementing almost everything in the book using SAS's `PROC MATRIX`, coding up Monte Carlo experiments to demonstrate properties of estimators in different scenarios, implementing some Newton-Raphson algorithms for simple MLE problems and the like. Getting this level of exposure to statistical programming and early on was incredibly valuable; my later transitions to `gauss`/`Splus`/`R`/`C` etc were relatively easy, and coming into graduate school with a reasonable degree of programming skill helped me secure research work in the summers (an especially important consideration for a foreign graduate student).

Finally, Gow and Bruce Western helped me understand that I could and should pursue a PhD in the United States. Gow had shown me Mo Fiorina's *Retrospective Voting in American National Elections* which struck me as an incredibly impressive piece of work, and as the kind of research I would like to do (cutting edge empirical work on political behavior informed by formal models), and Rochester seemed a great place to get that requisite training. Moreover, if Rochester was good enough for Fiorina in 1968 it would be good enough for Jackman in 1988, and besides, it was in New York State and hence close to New York City, right? Dick Niemi called to say I had been admitted, and I accepted over the phone. My professional socialization intensified: more statistics, the Australian version of the ICPSR summer school, an Honors workshop presentation on logit/probit, even a conference paper implementing robust regression co-authored with Western and Gow. Since the Australian academic year runs on the calendar year, I delayed my start at Rochester until January 1989.

Rochester/Princeton: The PhD methods sequence at Rochester in the late 1980s consisted of an introduction to probability and statistics taught by Linda Powell (which I missed, given my late arrival from Australia) and a Gujarati-level econometrics class taught by Dave Weimer (both required classes). Larry Bartels taught an advanced class in the 2nd year of the

[4] e.g., David John Gow, "Scale Fitting in the Psychometric Model of Judicial Decision Making," *American Political Science Review* V73 (1979): 430-441.

program, a quick review of regression models, before a set of topics covering discrete choice, systems of equations and measurement models (factor analysis and analysis of covariance structures). There was no required text but Larry's syllabus suggested that we own Hanushek and Jackson's *Statistical Methods for Social Scientists* or Johnston's *Econometrics Methods*. Glancing over my file of homeworks and lecture notes from Fall 1989 I see that for anything beyond regression, our homeworks were surprisingly light on exercises with real data. A typical Bartels homework from the late 1980s for an advanced class would be to show that a systems of equations or covariance structure setup was or wasn't identified. Larry also ran a fourth, workshop style class where students worked up projects into papers; the thrust of this course was to get us doing more than running econometric bells and whistles through political science data, but to work hard on identifying the substantive issues and precisely how a particular model or estimator would shed light on these issues.

In none of my Rochester methods classes were there lab sessions or classes devoted to implementation with available software; I think Larry's position was that we would figure that out for ourselves, and that teaching us how to figure out what is *estimable* is more important than the mechanics of *estimation* itself. Today it difficult to imagine a graduate methods class that would be so "hands-off" with respect to issues of implementation (my own teaching style is to switch over from theory to examples and real data and analysis in `R`); although frankly, with today's programs capable of optimizing user-defined likelihood functions, or the relative ease with which one can toss fanciful models into `WinBUGS`, it might not be such a bad thing if graduate methods classes spent more time on something as fundamental as identification.

While at Rochester I took a class on discrete choice econometrics in the Economics Department: a good deal of the class was spent on properties of maximum likelihood estimators (i.e., ch 4 of Amemiya's *Advanced Econometrics*), and definitions of various types of models (type 5 tobit, anyone?). This class was both eye-opening and incredibly frustrating. On the one hand, my econometric and statistical literacy sky-rocketed; on the other, in 14 weeks of lectures not once we did look at an actual application of any of the models with data. I've seen this pattern replicated repeatedly; good students go to Economics departments looking for advanced methodological training, and return having been exposed to a lot of asymptotics, but with their data analytic skills unimproved or degrading. As a result I try to steer students towards genuinely applied econometrics classes (which usually aren't in the econometrics PhD sequence at a place like Stanford, and are more likely to be in labor, public finance or trade), or to the Statistics Department.

The dearth of actual data analysis in the advanced classes I took had several causes, not least of which was the computing power available to political scientists at the time. Harold

Stanley's 286 was the fastest machine in the Rochester Political Science Department (in the summer of 1990 Harold let me use his office and machine to run the Monte Carlo experiments that appeared in Larry's quasi-IV paper[5]), and anything beyond linear regression was still costly. For instance, as a graduate student at Rochester in 1989, ordered probit usually meant a trip to the basement computing lab, cosying up to a monochrome green VT102 terminal and running LIMDEP on a remote IBM mainframe. Few graduate students owned their own desktops, and the PC lab at Rochester comprised only 3 or 4 painfully slow, loud, and hot XTs (largely faculty discards); this meant that the few PC-based programs for advanced modeling were beyond my reach.[6] The bottom line was that up until about 1991, at least at Rochester, anything beyond regression was a mainframe-only chore. And data analysis as we know it today (e.g., interacting with the data via spreadsheets and graphics) was just unimaginable on a mainframe, circa 1990.

Sometime around 1991-92 a lot changed, and quickly; as it turned out, my first years of graduate school (1989 and 1990) were the tail-end of the mainframe era. PCs got faster (and kept getting faster), and color monitors started showing up in computer labs. It became increasingly apparent that I needed to get my hands on `gauss`; Gary King was using the `maxlik` routine in `gauss` to implement MLEs, and Neal Beck had used it to implement a Kalman filter in a *Political Analysis* article. On a trip out to visit Bruce Western at UCLA I first saw `Splus` running on a PC in the impressive Social Science Statistics Lab there; the first time you click on a scatterplot to identify an outlier is a wondrous moment, and my approach to statistical computing and data analysis was forever changed.[7] I also discovered some new Sun workstations hidden away in a quite corner of the Rochester computing center; `Splus` running under the `X11` windowing system on the seemingly huge greyscale Sun monitors was tremendous fun, and if it meant going to another building to do my work, well so be it.[8] The Internet started to be more than an e-mail network, and things like `ftp` meant data and code was suddenly easy to share.

But overshadowing these technical advances was that in 1991 Princeton stole Larry away from Rochester. This was momentous; I teased Larry that I would kidnap his kids if I didn't get to go along with him (he assured me that that wouldn't be necessary), and in the end I wound up spending three years as a visiting student in the Woodrow Wilson School. Don

[5]Larry M. Bartels, "Instrumental Variables and 'Quasi-Instrumental' Variables", *American Journal of Political Science* V35 (1991): 777-800.

[6]For instance, Jeff Dubin and Doug Rivers' SST is/was a marvelous piece of software, managing to squeeze respectable performance out of the computing power available at the time.

[7]See my enthusiastic embrace of the brave new world in "`GAUSS` and `S-PLUS`: a comparison." *The Political Methodologist* V6, No.1 (1994): 8-13.

[8]From that point on my statistical computing slowly shifted from PC based to being based around various NIXs (UNIX, HP-UX, Linux and now Mac OS/X).

Stokes, the then Dean of the Wilson School, had made sure Larry was well catered for on the research side, with fellowships and office space for RAs. In turn, Larry entrusted me with ordering hardware and statistical software for his fledging group, and I obliged with getting the fastest PCs then available, and a nice suite of statistical software. Best of all, Larry was only two doors away, and he was incredibly generous with his time; in addition to serving as my dissertation advisor and mentoring me in the ways of the profession, Larry helped me explore the world of Bayesian statistics that I was increasingly interested in.

Bayes. In the winter of 1990 (my 2nd year at Rochester), I was the TA for the PhD regression class. Larry Bartels was invited to give a guest lecture on Bayesian approaches to econometrics. I knew nothing of the Bayesian approach at that stage, other than that it involved the use of prior information, which sounded uncontroversial enough. Bartels' lecture was the first and still the best Bayesian critique of frequentist statistics I've ever encountered. I'd always been troubled by aspects of the frequentist approach, at least as conventionally practiced in social-science settings: point null hypotheses, research findings that live-or-die at point-oh-five, inferential procedures that rest on properties of statistics over imaginary, repeated applications of sampling processes. Bartels' critique brought all of that together, showing that Bayes provided a coherent framework with which to make the probability statements about parameters that it seemed everyone wanted to make (and often do make), but are not what the frequentist approach supports.[9] My memory of this lecture was that it quite revolutionary in its implications; Larry reminds me that as a minimum, he wanted to show people that one could still run regressions as usual, but interpret the results in a more sensible, Bayesian way (i.e., with the right ignorance priors, least squares regression estimates are what a Bayesian would report as the posterior mean for β, etc). In light of the demolition job I thought Bartels had just delivered, I wondered how I could go back upstairs and run and report regressions with any degree of self-respect. Nonetheless, upstairs we went, back to whatever it was we were doing, some of us giving lip-service to the Bayesian approach in our interpretations of parameter estimates and their standard errors. But that really was that. Bartels' one hour lecture was as much Bayes as one would encounter in the Rochester quantitative methods program, although probably an hour more than one might see almost everywhere else in a political science PhD program in 1990. There was no Bayesian community at Rochester to go for additional guidance: again, this was 1990, before

[9] I shared Peter Lee's confusion: "When I first learned a little statistics, I felt confused... Not because the mathematics was difficult...but because I found it difficult to follow the logic by which inferences were arrived from data.... the statement that a 95% confidence interval for an unknown parameter ran from -2 to +2 sounded as if the parameter lay in that interval with 95% probability and yet I was warned that all I could say was that if I carried out similar procedures time after time then the unknown parameters would lie in the confidence intervals I constructed 95% of the time." *Bayesian Statistics*, Oxford University Press, 1989, p.vii.

the mid-1990s, MCMC-led Bayesian revival.[10]

On the other side of the country, in UCLA's Sociology program, Bruce Western was encountering a similar critique of conventional, frequentist practice, largely led by Dick Berk. On a summer trip out to LA, Bruce and I sat in on a biostats lecture by Rod Little: this was the first time I'd seen a practicing Bayesian in full-flight with applications, unapologetic about the use of priors, pointing out that one could "sod the data" if one's priors were sufficiently stringent. One of Little's texts was Peter Lee's 1989 book, the source of the quote in the footnote, above.

Back at Rochester, Dick Niemi passed along a paper by Andrew Gelman and Gary King using Bayesian methods to estimate seats-votes curves that appeared in *JASA*.[11] Niemi had shown me the seats-votes setup in my first year of graduate school (leading to my first two refereed publications, both co-authored with Niemi), but Gelman and King had clearly taken the literature to another level. My work with Niemi relied on the "multi-year" method a la Tufte to estimate seats-votes curves (and hence, estimates of electoral bias and responsiveness); each election contributes a x and a y (votes and seats shares, respectively) to a data set that one would then use to estimate seats-votes curves via a log-odds on log-odds regression. This method was useful in generating sweeping, historical characterizations of electoral systems, but not especially useful in assessing the bias and responsiveness of a just-implemented or yet-to-be-implemented districting plan in a specific jurisdiction. Gelman and King solved this problem with an elaborately parameterized simulation model (including a mixture model and multi-year analysis to bound the magnitudes of year-to-year, district-level shocks) all wrapped up in a fully Bayesian framework: estimation employed the *EM* algorithm and the Tanner and Wong data augmentation algorithm, the latter a special case and forebearer of the Gibbs sampler (the workhorse MCMC algorithm). I was enormously impressed by the statistical sophistication applied to a problem I had worked on myself, and, in particular, how helpful Bayesian simulation methods could be: coming up with a flexible parameterization to attack a problem is one thing (and impressive in itself), but it was clear to me that the computational tools Gelman and King used were incredibly powerful and of widespread applicability.

Shortly after the move to Princeton with Bartels, armed with a fast machine of my own and a copy of `gauss`, I set about replicating the Gelman and King setup. This work was far removed from my dissertation, and a less tolerant advisor would have warned me off

[10] Only after leaving Rochester did I realize that Martin Tanner was over in the biostatistics department. Tanner's *Tools for Statistical Inference* went through three editions with Springer, and was a very useful resource for me as I found my away around the fast-moving Bayesian literature in the mid-1990s.

[11] Andrew Gelman and Gary King "Estimating the Consequences of Electoral Redistricting", *JASA*, V85 (1990): 274-82.

it. Instead, Larry generously helped me decipher the Gelman and King article, helping me work through the Bayesian analysis of the mixture model ("...the Dirichlet is conjugate to the multinomial...") and exactly what was being computed at each step of the way. Under the guise of "I'm working on something for Larry," I took over a conference room in the Woodrow Wilson School for a week, whiteboarding up my assault on the Gelman and King piece; in addition to being my first foray into serious Bayesian computation, I was also learning `gauss` by throwing myself in the deep end. Gary King patiently answered many e-mails seeking help or clarification. I then disappeared into Princeton's fabulous Firestone Library, collecting Australian election returns, which I then fed to my implementation of the Gelman and King setup (with a slight generalization to deal with the rampant malapportionment in Australian jurisdictions). The result was a *BJPS* piece, far and away the longest and most technical thing I had done up until that time, convincing me that perhaps there was a future in this Bayesian business.

Around the same time Bruce Western and I worked up a draft of what would become an *APSR* piece.[12] We wrote this piece after realizing that via largely independent routes (Bartels for me; Dick Berk for Bruce), we had arrived at almost identical positions regarding Bayesian versus frequentist approaches in statistical work in the social sciences. We argued that repeated sampling was an inappropriate foundation for statistical inference in many settings (in particular, cross-sectional statistical analysis in comparative politics); Bayes, we argued, provided a solution to this problem, as well as a way to formally incorporate the often-abundant but non-quantitative prior information available to students of comparative politics (e.g., historical accounts and narratives).

I presented this work to the Political Methodology summer meetings at Florida State in 1993, and was shocked by the hostility of the reaction from the audience. I later realized I had walked into a mine field, that many in the Methodology group had encountered the Bayesian/frequentist debate (in the aftermath of Ed Leamer's wonderful 1978 book, *Specification Searches*), and were letting me hear their well-rehearsed positions (and all at once, or so it seemed). Punch-drunk, I flailed around trying to salvage my talk from the constant interjections flying across the room, even failing to recognize what should have been helpful comments from people like Chris Achen, friendly to the Bayesian position. Afterwards, John Jackson took me aside to cheer me up, consoling me with the view that the liveliness of the session was an indicator of the impact of the ideas in my talk. Throughout it all, Larry said nothing, not that I expected him to, and fair enough too; nothing that happened to me that morning in Tallahassee was unfair or unprofessional. In retrospect I was woefully

[12] Bruce Western and Simon Jackman, "Bayesian Inference for Comparative Research," *American Political Science Review*, V88 (1994): 412-23.

under-prepared to give *that* kind of talk to *that* kind of audience: the parts of the Bayesian program I was stressing in the talk — subjective probability and informative priors in a small n setting – were uncontroversial to Larry, Bruce, and myself, but clearly the rest of the profession wasn't so convinced.

The paper attracted a mixed set of reviews at the *APSR*: one reviewer was enthusiastic, one was doubtful, and another took the curious position that while Bayesian ideas were of course the right way to approach the problems we laid out, graduate students shouldn't be given journal space in the *Review* to make those arguments. Larry encouraged to us hold our ground in revise-and-resubmit, and in the end our piece was published. But the experience led me to be pessimistic about the prospects for the widespread adoption of Bayesian ideas in the profession. Besides, I'd invested considerably in the then-emerging Bayesian computational tools (data augmentation, the Gibbs sampler), and I could see tremendous payoffs down that less controversial avenue, where, for the most part, the priors are vague and the analysis via MCMC is formally Bayesian, but, for all practical purposes, is "maximum likelihood by any other means". We'll see what can be done about that in the years ahead.

Post-Graduate. When I joined the faculty of the University of Chicago in 1994, I lucked out again. Chicago's Business School was home to Arnold Zellner, the leading Bayesian econometrician of his generation, meaning that the Chicago GSB was a Bayesian beachhead of sorts. A group of younger scholars there were beginning to ride the Bayesian/MCMC wave, in particular, Peter Rossi, Rod McCulloch and Nick Polson. In addition, in my first year at Chicago the Statistics Department there ran a speaker series devoted entirely to MCMC. McCulloch and Polson met gave me helpful advice on how professional statisticians really *did* MCMC; in particular, McCulloch pointed out that implementing MCMC for non-toy problems usually meant writing your own code in a real programming language like `C`, and Polson gave me a gentle introduction to the literature on convergence results for MCMC algorithms.

Stanford has been a wonderful place for my continuing development as a methodologist, and principally because of my senior colleague Doug Rivers. Doug's methodological training is perhaps the one we should be reading about. Rivers excels at so many facets of all that we call methodology and has been nothing short of an inspiration for me. Doug was initially quite skeptical about MCMC and the Bayesian approach more generally (I'd say now that position has moderated to "a little skeptical"); Doug's blend of skepticism, enthusiasm, but above all, a demand for rigor, has been immensely valuable for Josh Clinton and myself in our joint work, using Bayesian models to analyze rollcall data. Via my colleague David Laitin, I was introduced to Persi Diaconis (quite simply, one of the world's most important living

Bayesians), who has also been a gracious and valuable resource for me and my students.

In short, I've been extremely lucky: Gow and Western turning me on to quantitative social science, and helping get me from Australia to Rochester; riding Larry's coattails from Rochester to Princeton and Larry presenting me with the Bayesian fork in the road; a job at Stanford, great students, and Rivers' interest in using Bayes to move the football on a range of problems. I don't regard my methodological education as complete: a good methodological training gives us what we need to keep learning over one's career (indeed, anything less has to be considered insufficient in our field). I hope my luck continues.