

Is Regression Analysis Really Leading Political Science Down a Blind Alley?

When I received my December 1993 issue of *PS: Political Science & Politics* I came across a scathing critique by James P. McGregor on the overutilization of regression models in political science research. Specifically, McGregor gave a detailed critique of the traditional ordinary least squares (OLS) linear regression models frequently used by political scientists. While some of his warnings need to be heeded by those of us conducting empirical political science research, his critique, for the most part, is misguided. Such commentary is a disservice to those who rely on articles such as these to serve as a guide for their research and teaching endeavors.

In his critique, McGregor ignores the many credible and common forms of regression analysis that have gained a steady following within the past ten years or so in the profession. These more recent regression-based techniques (e.g., probit, logit, event count models, nonlinear regression, nonlinear dynamics in time series data, etc.) are not bound to the traditional assumptions associated with the OLS linear regression model. I do not contend that regression-based methods are appropriate in every quantitative application; however, these methods are much more flexible than McGregor's article would lead one to believe. By ignoring these recent and important developments, McGregor presents a heavily skewed perspective on the use of regression analysis in the social sciences. My aim is to shed light on the various issues raised by Professor McGregor, and to suggest that rather than rely on elementary methods of data analysis as our discipline's foundation in

answering complex social science questions, we need to move forward by exploring more advanced quantitative methodologies that can better capture the complexity of social processes often encountered in political science.

Linear Regression Analysis: Its Uses and Limitations

The conventional OLS linear regression model is by far the most utilized tool of statistical inference by political scientists. It clearly has intuitive, theoretical, and descriptive appeal. The linear regression model has a host of assumptions that must be met in order for one to have confidence in the results generated (Lewis-Beck 1980, 26–30). McGregor is accurate in saying that many of these assumptions are often violated. As most of us who conduct quantitative political science research know, this point is extremely well understood. But McGregor fails to give adequate attention to the various means that we use to address these violations of linear regression assumptions, such as serial correlated errors, heteroskedastic errors, among others. McGregor's article suggests that researchers never conduct any diagnostic checks when employing their empirical models. To the contrary, political scientists routinely seek to account and/or correct for various possible violations of the statistical assumptions associated with the OLS linear regression model.²

McGregor's assumptions concerning the practice of regression analysis and its epistemological underpinnings are questionable and incomplete. I do, however, agree wholeheartedly with him on a few key points. First, most social science relationships are not truly linear nor additive;³ however, there are many instances (but clearly not all) where they are reasonably

close enough to lend support for these assumptions (Achen 1982, 36–37; Lewis-Beck 1980, 13).⁴ Second, in those many instances where OLS linear regression methods are inappropriate, we must utilize methods of data analysis that will provide robust results even when conventional OLS assumptions attached to the conventional linear regression model do not hold.⁵ But even with the shortcomings of traditional OLS linear regression, does this mean we need to forsake regression analysis in general? I believe that we do not have to abandon the regression model as an empirical analytical tool since recent developments in this area over the past dozen years makes it much easier to deal with the concerns raised by McGregor's largely uninformed criticism of regression analysis.

One such instance where McGregor unfairly criticizes regression analysis in a highly speculative manner occurs in his commentary on the covariates of democracy. He attributes the great variability associated with these findings to the use of regression analysis (p. 802). Based upon inspecting these articles, one notices that there are many research design and measurement differences between these studies (not attributable to regression analysis) that appear to be the main source for these diverse findings. These possible non-regression related problems (e.g., different time periods for each study, the countries in each sample are not identical, different research designs, and possibly biased sampling) seem to be the reason for the lack of robustness associated with these studies' findings. For instance, Banks (1972) and Jackman (1973) are cross-sectional studies with the former examining seven-year interval median data from 1816–1966, while the latter examined only the year 1960. On the other hand, Arat (1988) employs a

TABLE 1
A Very Brief Sampler of Popular Advanced Regression Methods
Employed by Political Scientists

(A) <u>Binary Choice Regression:</u>	
Logit:	$E(Y_i) = 1/(1 + e^{-\alpha - \beta X_i})$
Probit:	$E(Y_i) = (1/2\pi) \int e^{-z^2/2} dz_i$
(B) <u>Discrete Ordered Regression:</u>	
Ordered Probit:	$Y_i^* = \alpha + \beta X_i + \epsilon_i$
where:	$Y_i = 1$ if $Y_i^* < 0$, $Y_i = 2$ if $0 < Y_i^* < A_1, \dots, Y_i = M$ if $A_{m-2} \leq Y_i^*$
(C) <u>Discrete Unordered Regression:</u>	
Multinomial Logit:	$E(Y_i) = e^{\alpha_j + \beta_j X_i} / \sum_k e^{\alpha_k + \beta_k X_i}$
(D) <u>Discrete Event Count Regression:</u>	
Poisson:	$E(Y_i) = V(Y_i) = \lambda_i = \exp(X_i \beta)$
Negative Binomial:	$E(Y_i) = \lambda_i = \exp(X_i \beta)$
with	$V(Y_i) = \lambda_i \sigma^2 = \exp(X_i \beta) \sigma^2$
where:	$\sigma^2 > 1$ (overdispersion)
Generalized Event Count:	$E(Y_i) = \lambda_i = \exp(X_i \beta)$
with	$V(Y_i) = \lambda_i \sigma^2 = \exp(X_i \beta) \sigma^2$,
where:	$0 < \sigma^2 < 1$ (underdispersion)
	$\sigma^2 = 1$ (Poisson dispersion)
	$\sigma^2 > 1$ (overdispersion)
(F) <u>Time Series Regression</u>	
Polynomial Distributed Lag:	$Y_t = \alpha + \beta[\lambda_0 X_t + (\lambda_0 + m^n \lambda_n) X_{t-m}] + \epsilon_t$
where:	λ = the lag weights
	m = the number of lags
	n = the order of the polynomial

pooled cross-sectional time series research design from 1948–77 for 127 countries. In addition, many of these studies construct different measures and weighting schemes to operationalize the concept of democracy. Given the high degree of variability of regarding issues of measurement, variable construction, research design, and alike, it is not surprising that the findings vary to such a large extent. The results generated from regression analysis, like any other quantitative technique, can be influenced by a myriad of pre-data analytic factors.

Advanced Regression Analysis

Various forms of regression analysis have been introduced into political science in the past dozen years that are designed to address the various problems encountered with the traditional OLS linear regression model.⁶ Table 1 presents a very brief menu of advanced regression models designed to deal with many issues that cannot be adequately handled by elementary

OLS linear regression techniques.⁷ These advances deal with the limitations cited by McGregor (p. 802). For instance, recent forms of regression analysis are designed to deal with discrete dependent measures whether they are dichotomous (e.g., probit and logit regression—see Aldrich and Nelson 1984), ordinal (e.g., ordered probit regression—see McKelvey and Zavoina 1975), or event counts (e.g., poisson, negative binomial, and generalized event count regressions—see King 1988, 1989a, 1989b). In addition, there have been recent developments in the analysis of duration data which is interval data, yet has special peculiarities that often render traditional regression methods inappropriate (e.g., Allison 1984; King et al. 1990, Tuma and Hannan 1984). As one can see by the terms used for these advanced regression techniques, many of these maximum likelihood estimation approaches take on a statistical process that differs from a bi-variate or multi-variate normal distribution that is alluded to by McGregor (p. 802).

Also, many of these distributions possess inherently nonlinear functional relations.

McGregor also calls for less restrictive quantitative models to be used in political science research (p. 804). There have been recent advances in these type of approaches during the past five years. For example, nonparametric statistical regression methods such as *bootstrapping* have been recently advocated by political scientists (Mooney and Duval 1993). The advantage of bootstrapped regression analysis is that the researcher does not need to make distributional assumptions regarding the parameterization of the model under examination. In a similar vein, recent developments in the analysis of time series data have led to the proposal of “theory sparse” parametric regression models such as vector autoregression (Freeman, Williams, and Lin 1989). Vector autoregression methods are designed for instances when one is working with time series data and they are not sure about how to specify their empirical model since (1) multiple competing theories exist, or (2) there is an insufficient amount of existing theory as to serve as a guide, and (3) the autoregressive nature of the variables being examined is not known with certainty.

Conclusion

McGregor’s dual contention that we as social scientists do not do enough to check the robustness of our results in the face of potential violations of the statistical assumptions,⁸ and the need for data analytic techniques that circumvent the problems associated with the OLS linear regression model are both extremely accurate insights. However, by overlooking recent developments in political methodology in the past dozen years, McGregor does little to advance discussion on the subject of regression analysis in political science. As a methodologist, McGregor should well be aware of these fairly recent developments in political methodology. By not doing so, he commits nu-

merous glaring flaws of omission in his criticism of regression analysis. To intelligibly criticize a subfield's present condition and suggest future directions, one must possess some knowledge regarding the current state of the subfield.

McGregor's call to move away from more advanced methods of quantitative analysis in political science is both illogical and misguided (p. 804).⁹ I believe that theory (or the lack thereof) needs to be the single most important consideration when modelling social processes. For this reason, it is important for political science to pursue advanced methodologies (including but not limited to regression analysis), which do not lead us down a blind alley as McGregor suggests, but instead construct bridges between our theoretical insights and social reality. McGregor and like-minded individuals need to acknowledge that we as a discipline will not progress in our pursuit towards understanding political phenomena by reverting back to simplified methods of analysis as our main foundation. Advanced methodological approaches generally tend to be less restrictive and/or more theoretically plausible in their assumptions, impose weaker demands on our data, and give us a more accurate portrayal of the social processes under examination.¹⁰ Since we as social scientists attempt to explain and/or understand complex behavior emanating from complex individuals in complex institutions and settings, we must construct models that are a closer representation to reality. Many of the recent methodological advances over the past dozen years or so in political science (especially in the area of regression analysis) have supplied us with greater insight into the subtleties of behavior by taking into account the intricacies that are inherent in the real world (e.g., discrete dependent variables, nonlinearity and/or multiplicative relations, different levels of measurement, lagged distributed effects). Empirical political science research will not progress by forsaking the discovery and refinement of advanced quantitative methodologies. If any-

thing, political science will stagnate by clinging to limited tools, or even worse, abandon the pursuit of innovations in statistical theory and data analysis that are so richly needed for the discipline's development and contribution to knowledge.

George A. Krause
West Virginia University and
University of South Carolina

Notes

1. The first portion of this title is loosely borrowed from a statement made by James P. McGregor (1993, 802). Since this is a critique of McGregor's article, this title seemed most appropriate.

2. For instance, serial correlation is often addressed by employing either generalized least squares estimation, maximum likelihood estimation, or the inclusion of lagged dependent variable(s). Also, heteroskedasticity is addressed in various ways by political scientists conducting quantitative empirical research employing regression analysis (e.g., weighted least squares, heteroskedastic-consistent robust standard errors, or bootstrapped standard errors).

3. McGregor is correct in suggesting that many social science relationships are not linear and additive because the effects of variables are typically "contextual." Therefore, we need to specify how the effect of the independent variable of interest varies depending upon its own values and those from other variables under analysis in order to close the gap between political theory and social reality. I would like to acknowledge Bill Berry for bringing this point to my attention.

4. A parallel point can be made regarding a comparison of linear versus nonlinear time series regression models. Specifically, King (1989a, 167) notes that the Wold decomposition theorem does suggest that linear time series models may serve as a good approximation of more complex (i.e., nonlinear) relations in many cases.

5. In Gary King's significant contribution to the field of political methodology, *Unifying Political Methodology: The Likelihood Theory of Statistical Inference* (1989a), he acknowledges and addresses the need to construct alternative estimation approaches that go above and beyond traditional OLS regression-based methods, and are necessary for accurate statistical inference in many political science applications.

6. My discussion of advanced forms of regression analysis employed in political science is by no means exhaustive. Due to purposes of brevity, however, I only discuss a handful of the most popular forms of advanced regression analysis. Granato (1992) proposes a "reduction theory" approach for analyzing the appropriateness of model specifications by conducting a battery of diagnostic tests.

7. Many of these advanced techniques are discussed at greater length in Bartels and Brady (1993).

8. Achen (1982, 11) and Lewis-Beck (1980, 26-30) also mention the limiting assumptions associated with OLS linear regression analysis.

9. This is especially peculiar when one considers his contention that social processes reveal much more complexity (e.g., nonlinearity) compared to the processes generated in the physical sciences (p. 802-803). For if social processes are as complex as McGregor implies, it appears to make sense to employ advanced methods of data analysis (including regression) to obtain a more realistic portrait of social reality.

10. These points are at the heart of Gary King's (1989a) maximum likelihood estimation approach to political methodology.

References

- Achen, Christopher H. 1982. *Interpreting and Using Regression*. Beverly Hills, CA: Sage Publications.
- Aldrich, John H., and Forrest Nelson. 1984. *Linear Probability, Logit, and Probit Models*. Beverly Hills, CA: Sage Publications.
- Allison, Paul. 1984. *Event History Analysis*. Beverly Hills, CA: Sage Publications.
- Arat, Zehra F. 1988. "Democracy and Economic Development: Modernization Theory Revisited." *Comparative Politics* 21:21-36.
- Banks, Arthur S. 1972. "Correlates of Democratic Performance." *Comparative Politics* 4:217-30.
- Brady, Henry E., and Larry M. Bartels. 1993. "The State of Quantitative Political Methodology." In *Political Science: The State of the Discipline*, ed. Ada W. Finifter. Washington, D.C.: American Political Science Association.
- Freeman, John R., John T. Williams, and Tse-min Lin. 1989. "Vector Autoregression and the Study of Politics." *American Journal of Political Science* 33:842-77.
- Granato, Jim. 1992. "An Agenda for Econometric Model Building." *Political Analysis* 3:112-54.
- Jackman, Robert W. 1973. "On the Relation of Economic Development to Democratic Performance." *American Journal of Political Science* 17:611-21.
- King, Gary. 1988. "Statistical Models for Political Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model." *American Journal of Political Science* 32:838-63.
- King, Gary. 1989a. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.
- King, Gary. 1989b. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33:762-784.

- King, Gary, et al. 1990. "A Unified Model of Cabinet Duration in Parliamentary Democracies." *American Journal of Political Science* 34:846–71.
- Lewis-Beck, Michael S. 1980. *Applied Regression: An Introduction*. Beverly Hills, CA: Sage Publications.
- McGregor, James P. 1993. "Procrustus and the Regression Model: On the Misuse of the Regression Model." *PS: Political Science & Politics* 26(4):801–4.
- McKelvey, Richard, and William Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4:103–20.
- Mooney, Christopher Z., and Robert D. Duval. 1993. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Beverly Hills, CA: Sage Publications.
- Tuma, Nancy Brandon, and Michael Hanan. 1984. *Social Dynamics*. New York: Academic Press.

On Looking More Like America

Thomas R. Dye's analysis of the Clinton cabinet ("The Friends of Bill and Hillary," *PS*, December 1993) left out an important dimension. Gender, ethnicity, education, and occupation merits attention, but what about a more fundamental attribute, physical height? Consider a comparison to the Watergate era, with its suggestive correlation of above average height and criminality. (Let me confess that I speak from a protected position, standing

near the national average at 5 feet 6 ½ inches.)

Back in the Watergate days the president's men were significantly taller than the average American of their generation. A police lineup of the Watergate conspirators brought in Attorney General John Mitchell at 5'9" while Ehrlichman, Halde- man, and Colson all stood tall at 6'2". A photo reconnaissance as- sessment of the 1972 Nixon cabinet presents an estimated average height of 5 feet 10.8 inches com- pared to the relevant cohort in the American male population, which averaged 5 feet 8.9 inches (the all- American average including women was 5 feet 6.25 inches).

By contrast, President Clinton's government clearly looks more like America. To be sure, the president himself suffers a height handicap in this respect. Bill Clinton is essen- tially as tall as George Bush, de- pending on how the hair is handled, and he is clearly taller than any other President since LBJ, or, within this century, William Howard Taft. His cabinet, on the other hand, ranges from a Labor Secre- tary at 4'8" and a Secretary of Health and Human Services who stands five foot flat in her bare feet to an Attorney General who tops the 6 foot 1 ½ inch mark. This spread of 17 ½ inches demonstrates

greater diversity in height than the Nixon cabinet, which ranged from a 5 foot 6 inch Italian-American to a 6 foot 6½ inch Kentuckian.

The mean height of the Clinton cabinet probably also comes closer to the current American median of 5 feet 7 inches, though we can't calculate the precise figure for the whole cabinet since the photo made available by the Clinton White House is taken at an angle and shows some members in heels while others are partially obscured behind colleagues.

Some two hundred years ago our shortest (5 foot 4 inch) president argued that social diversity pro- vides a powerful restraint on the potential for tyranny. That proposi- tion remains true, but diversity in itself is not enough for good gov- ernment. During his first year in office President Clinton has in- vested so much of his attention into crafting how his government looks that he may have overlooked the more important issue of how it ac- tually governs. As in the new De- fense Department policy on sexual orientation the crucial question is not what the president's men and women *are* but what they *do*.

James F. Guyot
Baruch College

[*Hillary not included.]

SimHealth gives you everything you need to re-make the American health care system—except the First Lady.

A unique new simulation software created by Maxis and brought to you by the Markle Foundation, SimHealth elects you leader. Now you can build a health policy in tune with *your* values, and watch it in action on a virtual "Main Street." Will you be re-elected? You'll certainly be more engaged, entertained and educated about our national debate than you ever thought possible. So why not give us a call? Because at \$29.95—that's 25% off retail—and in a new, easy-to-install, 16-color version, SimHealth is just what the doctor ordered.



1-800-U-CHANGE
1-800-824-2643

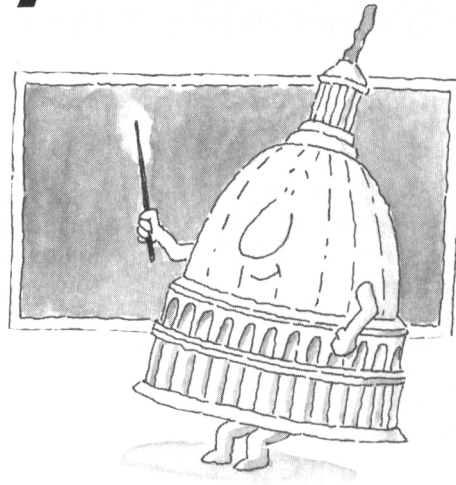


Mail orders:

SimHealth, PO Box 2096, Woodinville, WA 98072-2096

Add \$6 for shipping and handling. NY and WA residents, please add applicable sales tax. Requires an IBM or 100% compatible with DOS, a 386 microprocessor or greater, and a 16-color monitor.

Bring the Capitol to your classroom.



Put your students on Capitol Hill, with Congressional Quarterly publications and services.

The CQ Weekly Report keeps you and your students up-to-date with objective, thorough coverage of Congress every week.

And economical classroom subscriptions let you and your students follow congressional action as it unfolds!

The CQ Researcher offers in-depth background, analysis and bibliographies on one national or international topic each week.

With **CQ's Washington Alert**, you and your students have online access to more than 25 fully searchable databases of congressional information.

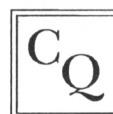
Call now for more information on CQ publications and services — and put your classroom in the thick of congressional action!

For FREE samples of print publications or information on CQ's Washington Alert, contact jsteiger@cqalert.com. and mention this ad!

Or call:

Vickie Binsted at (800) 432-2250 ext. 279, or (202) 887-6279, for print publications and classroom subscriptions.

Waldo Tibbetts at (800) 432-2250 ext. 350, or (202) 887-6350, for CQ's Washington Alert.



Congressional Quarterly
1414 22nd Street N.W.
Washington, D.C. 20037

4LPP93