

5. We do not want to rule out the possibility that some of the China specialists are government contractors, and their government ties may have contributed to the similarities in policy preferences. However, we still believe that the policy congruence is largely because of shared perceptions about cultural as well as strategic and political factors in China studies than links to the government. We did not ask whether people have government contracts in our survey because we think that it may complicate their answers to the other questions.

6. We suspect similar situations may also exist among other area specialists in political science. We do not believe that the case of China scholars is an anomaly. Our study may encourage other area specialists to do similar studies in their area to examine the gap between political science and public policy.

References

- Fairbank, John K. 1992. *China: A New History*. Cambridge: Harvard University Press.
- Herspring, Dale R. 1992. "Practitioners and Political Scientists." *PS: Political Science & Politics* 15(3): 554-58.
- Huntington, Samuel. 1988. "One Soul at a Time: Political Science and Reform." *American Political Science Review* 82: 3-10.
- Kammarck, E. C. 1990. "Political Scientists in Presidential Campaigns." *PS: Political Science & Politics* 23(3): 428-29.
- Leigh, Lawrence J. 1991. "Political Scientists as Expert Witnesses." *PS: Political Science & Politics* 24(3): 521-23.
- Maxwell, Mary. 1992. "The Gulf War and Political Science." *PS: Political Science & Politics* 25(4): 693-95.
- Moe, Ronald C. 1991. "Political Science and the Savings and Loan Crisis." *PS: Political Science & Politics* 24(3): 451-55.
- Mosher, Steven W. 1990. *China Misperceived: American Illusions and Chinese Reality*. New York: Basic Books.
- Myers, Ramon H., and Thomas A. Metzger. 1980. "Sinological Shadows: The State of Modern China Studies in the United States." *The Washington Quarterly* 3: 144.
- Oksenberg, Michel. 1970. *A Bibliography of Secondary English Language Literature on Contemporary Chinese Politics*. New York: East Asian Institute, Columbia University.
- Pye, Lucian. 1990. "China: Erratic State, Frustrated Society." *Foreign Affairs* 69(4): 58.
- Sabatier, Paul A. 1991. "Political Science and Public Policy." *PS: Political Science & Politics* 24(2): 144-46.
- Sundquist, James L. 1991. "Political Scientists and Public Policy Research." *PS: Political Science & Politics* 24(3): 531-34.

Procrustus and the Regression Model: On the Misuse of the Regression Model

James P. McGregor, *U.S. Information Agency*

In *The Logic of Scientific Revolutions*, Thomas Kuhn (1970) described the power of prevailing paradigms to assist—or impede—progress in the sciences. A paradigm channels thinking and observation; under conditions of what Kuhn calls "normal science," it dictates what is a fact and the methods by which that fact is to be studied.

While the social sciences are not yet subject to a single paradigm, scholars tend to behave as if a paradigm prevailed, especially with respect to method. The introduction of the scientific method to mainstream social science represented a powerful advance in the study of mankind and his institutions. Unfortunately, it also brought with it potent tendencies toward scientism and considerable resistance to

challenges to the methodological canon.

My argument is that one of the principal icons of social scientists—the regression model—has been misused and, in my view, has been a major impediment to progress in political science as well as in other social sciences. I submit that the method has not only been abused by general failure to apply it properly but also that the model is not a reasonable way of looking at the world of politics.

The Regression Model

The regression model dominates empirical work in political science. Rough evidence of this can be found in a review of articles in the *Ameri-*

About the Authors

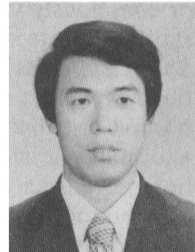
Tao Wu

Tao Wu is a Ph.D. candidate in American politics at Arizona State University. He is completing his dissertation on "Party Reforms and the Transitions of Political Elites: 1972-1988" in 1994.



Yang Zhong

Yang Zhong is an assistant professor in the department of political science at The University of Tennessee, where he teaches Chinese and comparative politics. His publications have appeared in *Studies in Comparative Communism, Armed Forces and Society, Journal of Asian and African Affairs*, and *Pacific Focus*.



can Political Science Review over the past year: almost every article that displayed findings in tabular form used some form of regression analysis—ordinary least squares, generalized least squares, probit or logit analysis. To be sure, many of these analyses were preceded by or supplemented with other methods such as simple marginals, crosstabulation or correlations—but the key method was regression.

The regression model has powerful attractions. Under the proper conditions, it can provide precisely the type of information political scientists need in their search for explanations of political phenomena. It can summarize the relationship among variables in a parsimonious and precise manner and has the additional virtues of being relatively easy to

understand and intuitively attractive. There is little wonder that the regression model has achieved its preferred status in the social sciences.

Yet the theoretical and intuitive allure of the regression model appears to have blinded many social scientists to its gross abuse in application. Despite its relative theoretical simplicity, the regression model is loaded with caveats and enabling assumptions that are almost always ignored, dismissed, left unexamined, or consciously violated. Standard statistical texts such as Blalock's *Social Statistics* set out these assumptions explicitly: interval level data, normal distributions (including bi- and multi-variate normality), and homoscedasticity (variances of the Y distributions are the same for each value of X) (Blalock 1960, 276).

Almost none of the *APSR* articles noted earlier paid any attention to these assumptions. One could (and many must) argue that the regression model is "robust" and can endure some level of violation of assumptions. This may be true. Yet there is no discussion in the *APSR* articles of the possible effects of assumption violations. This is not a new problem: over two decades ago Holt and Richardson identified the same situation in an essay on paradigms in comparative politics (1970, 60 ff).

A further sign that not all is well in the use of the regression model is the fact that there is no corpus of reinforcing findings on political phenomena based on regression. Studies of what purport to be the same set of factors (the covariates of democracy come readily to mind—see, e.g., Banks [1972], Bollen and Grandjean [1981], Jackman [1973, 1974], Bollen and Jackman [1985], Arat [1988], Muller [1988]) result in sometimes wildly varying regression coefficients and multiple R's. If the regression model is strong and robust, one could reasonably expect that similar studies should lead to similar findings—but they do not. Part of the problem could be bad data or violations of assumptions. In my view, however, the difficulty is that the regression model is simply inappropriate for most political phenomena.

Shortcomings of the Regression Model

The three assumptions noted earlier do not include the two most important: that the relationship under study must be linear and additive. These assumptions have both mathematical and theoretical significance, the latter of which is most important for the present discussion.

What kind of universe does the regression model envision? An exceedingly simple one, it would seem: phenomena are related to one another in a way that describes a straight line on a two-dimensional graph or an n-dimensional plane in the case of multicausal phenomena

Yet the theoretical and intuitive allure of the regression model appears to have blinded many social scientists to its gross abuse in application.

(assuming appropriate data transformations). Moreover, a phenomenon is (usually) the simple sum of its weighted components.

Are these reasonable theoretical assumptions? I think not. Very little in the physical sciences indicates that nature fits together according to the assumptions of the linear, additive model. There is absolutely no reason to suppose that this would be the case in the social sciences. Of considerable significance is the fact that physical scientists do not regularly use the regression model in their work.

This is more than a quibble about different methodological approaches in the social and physical sciences. I argue that the domination of the regression model in political science has led us down a blind alley, causing us to accept certain conclusions as true when they are not and—perhaps more important—to overlook true relationships because, not being linear and additive, they cannot be revealed by the regression model.

Three simple examples drawn from

the physical sciences will illustrate this point: Galileo's law of falling objects, Boyle's law of gases, and Newton's law of gravitational attraction. Suppose a social scientist inclined toward the regression model were to undertake an attempt to discover the relationship among the factors involved in these laws. The social scientist would, of course, have some theoretical reason to suppose that the factors were in some way interrelated. Suppose further that the social scientist followed the standard rules of gathering observations and had near-perfect measuring instruments. Table 1 shows the results of applying ordinary least squares using a stepwise procedure to the resultant data.

If we compare the actual forms of the relationships to those estimated by regression, two points are immediately apparent. First, none of the regression equations comes even close to capturing the real form of the underlying relationship. Despite the fact that the estimate for Galileo's law of falling bodies resulted in a very respectable multiple R square of .94, the linear, additive regression model was simply wrong. Second, the laws are all very elementary relationships, algebraically speaking. Yet even the simplest non-linear bivariate law (Galileo's) was incorrectly estimated and, as the laws became moderately more complex, the regression model failed completely even though the relationships were perfect. A social scientist examining the attempt to estimate Newton's law would reasonably conclude that there was no relationship among the variables that could not be attributed to random error.

Table 1 demonstrates the potentially disastrous inferences that could be drawn applying the regression model. Both Type I error (rejecting a true hypothesis) and Type II error (accepting a false hypothesis) are possible. Because our real data are so imprecise, almost any social scientist would be delighted to achieve a multiple R square at the level of the Galileo example—and he or she would be wrong if it were concluded that the regression model had captured the relationship. By contrast, there is a perfect relationship among the variables in both the Boyle and

Newton examples—yet the social scientist would probably reject both as unpromising based on the regression analysis results.

Uses of the Regression Model

Is the regression model then worthless to social scientists? Absolutely not. While routine application of regression methods in the context of discovery may be dangerous, use of regression in the context of confirmation is quite useful. If one's theory explicitly permits the use of the linear, additive model, then it is entirely appropriate to apply it, provided that the data assumptions of the model are met. More importantly, bivariate regression can be used to discover constants and confirm laws. In each of the examples in Table 1, had the scientist speculated that the nature of the relationships was what we know the relationships to be (i.e., that $d = t^2$ in the case of the law of falling bodies), regressing d on t^2 using the data would have resulted in a perfect relationship, plus the constant 16 would have been reported as the unstandardized B coefficient. A similar procedure found the constants in both Boyle's and Newton's laws. In all cases, the regression analysis revealed a perfect fit between the predicted and the empirical. This, I submit, is extremely useful.

Heal Thyself

Part of the problem with the uncritical application of the regression method lies in the striving of a young science to emulate its seniors. We have witnessed the role of mathematics in the physical sciences without being aware that science and mathematics were once separate disciplines (Kline 1953, 241). True, the major advances in the hard sciences in the eighteenth century were largely the result of the absorption of science by mathematics, but the two are distinct enterprises. But science can proceed at least to a point without advanced mathematics; the three laws in Table 1 far exceed anything uncovered to date in the social sciences, yet they are extremely simple from a mathematical point of view.

Is this the counsel of despair? I think not. We have the example of Galileo to show us the scientific enterprise can be rescued from a mistaken path. Galileo was convinced that science as practiced in his time and by the Greeks and Romans before him had not advanced knowledge very far. His answer was to go back to basics and attempt to focus on quantitative descriptions of phenomena (Kline 1953, 185). Perhaps political science need not take such a

drastic step—although, in my view, it would be healthy if more effort were put into description and measurement of political phenomena. But resort to models such as regression clearly impedes progress by seducing us to believe that by adopting mathematics we are doing science. Clearly, we are not.

The implications of the preceding discussion are straightforward. First, political scientists need to follow their own teachings and critically

TABLE 1
A Regression Approach to Certain Laws of Nature

Galileo's Law of Falling Objects	
Variables:	d = distance an object falls t = time in fall
Unstandardized Regression Results:	
$d = (1593.91882)t + (-26642.65)$	
Standardized Regression Results:	
$d = (.96819)t$	
Multiple R	= .96819
R Square	= .93733
Standard Error	= 11776.56953
Actual Form of the Relationship: $d = 16t^2$	
Boyle's Law of Gasses	
Variables:	p = pressure t = temperature v = volume
Unstandardized Regression Results:	
$v = (-.00861553)p + (.004329250)t + .42779$	
Standardized Regression Results:	
$v = (-.41021)p + (.20738)t$	
Multiple R	= .46083
R Square	= .21237
Standard Error	= .52446
Actual Form of the Relationship: $v = tc/p$, where c is a constant specific to the gas under study (in this case $c = .082$)	
Newton's Law of Gravitational Attraction	
Variables:	m_1 = mass of object 1 m_2 = mass of object 2 r = distance between the two objects x = gravitational attraction
Unstandardized Regression Results:	
$x = (-17966.23)r + (.000535)m_2 + (.000482)m_1 + 658165.97$	
Standardized Regression Results:	
$x = (-.22982)r + (.07054)m_2 + (.06348)m_1$	
Multiple R	= .24953
R Square	= .06226
Standard Error	= 2102591.7458
Actual Form of the Relationship: $x = m_1m_2g/r^2$, where g is the universal gravitational constant ($g = 6.67 \times 10^{-11}$)	

Note: Approximately 1,000 observations for each variable were generated using BASIC's random number function. Dependent variable was computed using the actual form of the relationship. Stepwise regression was done in SPSS PC taking all defaults.

The Profession

examine the model of political phenomena implicit in their use of regression methods. It may very well be the case that the assumptions of linearity and additivity are appropriate—but this is a critical theoretical as well as mathematical conclusion that merits extensive discussion and justification.

Second, if we decide that the regression model is appropriate, we need to apply it properly. We are not scientists if we consistently violate the statistical assumptions of regression without overriding and justified cause.

Third, in the great majority of cases where the linear, additive regression model is not suitable, we need to strive harder to develop other forms of relationships among political phenomena. Almost no work is done in this area.

Fourth, we need to do better groundwork before launching ourselves into advanced methods. Kuhn notes that renowned scientists such as Copernicus and Brahe made their contributions principally with their meticulous observations of the heavens, not their discovery of scientific laws (Kuhn 1970, 26). Yet, the mere gatherer of facts has little respect in the social sciences.

Finally, we need to study methods more appropriate to the underdeveloped state of data and theory in the social sciences. Despite the appeal

and apparent power of regression and other parametric methods, our data are simply not up to their requirements in most cases. Non-parametric statistics—even simple crosstabulation—may offer firmer footing than potentially misleading advanced methods.

Might the rejection of the regression model solve all or even any of the problems faced by political scientists in their pursuit of underlying order in political phenomena? While such a rejection may make us feel that we are regressing, I would argue that we would avoid grievous and highly misleading error. A blind alley such as that represented by the regression model may ultimately cost us years of work that could be better put to use finding the laws that are the building blocks of any science.

References

- Arat, Zehra F. 1988. "Democracy and Economic Development: Modernization Theory Revisited." *Comparative Politics* 21: 21-36.
- Banks, Arthur A. 1972. "Correlates of Democratic Performance." *Comparative Politics* 4: 217-30.
- Blalock, Hubert M., Jr. 1960. *Social Statistics*. New York: McGraw-Hill.
- Bollen, Kenneth A., and Robert W. Jackman. 1985. "Political Democracy and the Size Distribution of Income." *American Sociological Review* 50: 438-57.
- Bollen, Kenneth A., and Burke D. Grandjean. 1981. "The Dimension(s) of Democracy: Further Issues in the Measurement and Effects of Political Democracy." *American Sociological Review* 46: 651-59.
- Holt, Robert T., and John M. Richardson, Jr. 1970. "Competing Paradigms in Comparative Politics." In *The Methodology of Comparative Research*, ed. Robert T. Holt and John E. Turner. New York: The Free Press.
- Jackman, Robert W. 1973. "On the Relation of Economic Development to Democratic Performance." *American Journal of Political Science* 17: 611-21.
- Jackman, Robert W. 1974. "Political Democracy and Social Inequality: A Comparative Analysis." *American Sociological Review* 39: 29-45.
- Kline, Morris. 1953. *Mathematics in Western Culture*. Oxford: Oxford University Press.
- Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*, 2nd ed. Chicago: University of Chicago Press.
- Muller, Edward N. 1988. "Democracy, Economic Development, and Income Inequality." *American Sociological Review* 53: 50-68.

About the Author

James P. McGregor

James P. McGregor (Ph.D., Ohio State University, 1975) is a policy officer at the U.S. Information Agency. He is a specialist in Central and East Europe. He has served as an adjunct professor at George Washington University and American University where he conducted courses on East European foreign policy and research methodology.

