Comparative Democracy: the Economic Development Thesis

Ross E. Burkhart

Michael S. Lewis-Beck

University of Iowa

DOI: https://doi.org/10.2307/2082715
COMPARATIVE DEMOCRACY: THE ECONOMIC DEVELOPMENT THESIS

ROSS E. BURKHART and MICHAEL S. LEWIS-BECK University of Iowa

In comparative politics, an established finding—that economic development fosters democratic performance—has recently come under challenge. We counter this challenge with a dynamic pooled time series analysis of a major, but neglected data set from 131 nations. The final generalized least squares-autoregressive moving averages estimates (N = 2,096) appear robust and indicate strong economic development effects, dependent in part on the nation’s position in the world system. For the first time, rather hard evidence is offered on the causal relationship between economics and democracy. According to Granger tests, economic development “causes” democracy, but democracy does not “cause” economic development. Overall, the various tests would seem to advance sharply the modeling of democratic performance.

Political sociology has few, if any, “iron laws.” However, certain central hypotheses seem so established as to be almost beyond challenge. According to Lipset “Perhaps the most widespread generalization linking political systems to other aspects of society has been that democracy is related to the state of economic development” (1959, 75). Among the many reasons offered for this empirical connection, a common idea is that increasing economic benefits for the masses intensify demands for political benefits of democracy. Economic development can spread authority and democratic aspirations among a variety of people, thus fostering democracy (Dahl 1989). The notion of economic development as a “requisite” to democracy (Lipset 1959) has survived increasingly sophisticated statistical tests.

Leading relevant quantitative, multivariate studies in the political science and sociology literature include Jackman (1973), Bollen (1979, 1983), Bollen and Jackman (1985) and Brunk, Caldeira, and Lewis-Beck (1987). In each of these regression-based analyses, data are gathered on a sample of nations, and a nondichotomous measure of democratic performance is predicted from economic development and, usually, other independent variables. Economic development consistently emerges as a statistically and substantively significant influence on democracy. For example, Brunk, Caldeira, and Lewis-Beck recently found that economic development alone accounts for more variance in democracy than the other independent variables taken together (1987, 46).

Given the strong results of these varied regression analyses, one might imagine that the “economic development thesis” merited little further testing. However, these studies, careful as they are, have noteworthy limitations. The samples are rather small (average = 78 nations) and sometimes explicitly non-representative. (Bollen and Jackman [1985], for example, exclude communist nations.) Further, they are all based on democracy measures that are by now quite old, from either 1960 or 1965. Finally, and most importantly, they are cross-sectional, not time-series, in design (though Bollen’s [1979] study contains a measure from an earlier cross section). Of course, the absence of a temporal dimension makes it difficult to overcome the criticism that the positive economics-democracy connection represents a spurious association, occurring within a particular time-slice.

These data difficulties receive some remedy in Arat (1988) and Gonick and Rosh (1988), two more current quantitative democracy studies. Both use larger samples of nations (95 and 116, respectively), including communist ones. Moreover, they extend the measurement of democracy into the 1970s to 1977 and 1979, respectively. Of special note, in their model estimations, they explore a series of cross sections over time. Given the previous quantitative work, one might expect economic development to continue to exhibit strong, unrivaled, effects. But it does not. To quote Arat: “Only a few countries fit the models suggested by modernization theory. . . . It can be concluded that increasing levels of economic development do not necessarily lead to higher levels of democracy, even for the less developed countries” (1988, 30). The conclusions of Gonick and Rosh have a similar ring: “Economic development . . . is not the most important factor affecting the degree to which a political system can be characterized as a ‘liberal democracy’. . . . Our application has allowed us to reject the findings of Lipset” (1988, 189, 196).

These surprisingly weak findings, based on apparently better data, force a reevaluation. Perhaps economic development actually has little influence on political democracy. Or, possibly, these analyses possess shortcomings that still leave the question open. Let us consider three potential problem areas: design, specification, and measurement. Although both studies examine cross sections over time, neither takes full advantage of the statistical leverage a pooled time-series design allows. Arat (1988) simply applies ordinary least squares (OLS) to the pool before discarding it for a country-by-country time-series examination. Gonick and Rosh apply a “variance component” model to the pool and neglect to discuss the central problem of autocorrelation from the time series (1988, 195). Nor does either explore lagged effects or Granger tests for causality. In other words,
they ignore the dynamic, analytic opportunities that the time component of the pool affords.

With regard to specification, Arat does not include any independent variables other than economic development, a modeling strategy not followed after Jackman's (1973) initial baseline work (1988, 27). While Gonick and Rosh have an equation that is more fully specified, they include independent variables (e.g., urbanization, literacy) that the other quantitative studies do not include in their "direct effects" equations (1988, 194-95). This specification difference could generate an unusual amount of collinearity, perhaps accounting for their weaker findings on economic development effects.

With respect to measurement, they both rely heavily on the cross-national time-series data of Banks (1979). Arat (1988) uses the Banks data for five of the six scales in her democracy measure. Gonick and Rosh (1988), for their democracy measure, use Banks data exclusively. Of course, the Banks data have been widely used in quantitative democracy studies. The difficulty is that key Banks democracy indicators, as they are commonly used, contain considerable measurement error, as Bollen (1993) has recently shown.

Bollen (1993) employs structural equation modeling with latent variables to arrive at the estimated "validity component" of democracy variables that have been constructed from different data sources. The "validity component" is what remains of the measure after the random measurement error and systematic measurement error components have been extracted. A variable with no measurement error could be assigned a validity component rating of 100%. According to Bollen (1993), the validity component ratings on the Banks democracy variables he studied are as follows: chief executive elected (14%), competitiveness of the nomination process (62%), effectiveness of legislative body (79%), freedom of group opposition (92%). These four variables, with the possible exception of the last, contain serious measurement errors. Gonick and Rosh (1988, 184) use all four variables in building their democracy index, while Arat (1988) uses the first three. Given the high degree of measurement error that is thus contained in their democracy indices, it becomes difficult to have confidence in their conclusions.

We shall try to overcome the foregoing research problems. We examine data on more nations, over a longer and more recent span of time. The data set is not Banks but, rather, Gastil. The research design takes full advantage of the pooled time series structure, ending with a generalized least squares-autoregressive moving averages (GLS-ARMA) estimation, with appropriate diagnostics along the way. Model specification builds on previous work, but introduces promising theoretical innovations that at the same time guarantee a rigorous test of the economic development thesis. Additionally, the issue of measurement error (which affects Gastil data as well) is dealt with. Finally, firm conclusions are drawn about the economic development thesis.

A DEMOCRACY INDEX: THE GASTIL DATA

Democracy can have many meanings. See the current provocative philosophical exchange between Mueller (1992) and Lienesch [1992]. Here we follow the tradition of the above quantitative studies, conceiving of democracy as a continuous variable, scored numerically from low values to high. The particular democracy measure we select aims to parallel the Bollen (1979) measure. Working for Freedom House, Gastil has provided annual (1972-89) ratings of nations on two dimensions (see Appendix A). The first, here labeled "political rights," assesses the right to vote, election meaningfulness, multiple political parties, opposition power, and government independence from foreign or military control. The second, here labeled "civil liberties," covers the freedoms of speech, assembly, and religion from terrorism or blatant inequality. (Details on the items appear in Appendix A.) By adding together the ratings on these two 7-point scales (after reversing their coded order for readability), each nation acquires an overall democracy index, ordered from a low of 2 to a high of 14.

The Gastil data have received considerable use in social science research (see Bollen 1993) but not for constructing an explanatory model of democracy either in cross-sectional or time-series studies. In certain respects, these data represent a splendid opportunity. Using them, we assembled annual observations on 131 nations 1972-89, yielding a total sample size of 2,358. (The nations and their democracy scores are listed in Appendix B.) Hence, it becomes possible to incorporate simultaneously analysis of over-time, as well as cross-nation, variation in democracy. Sources of variation that would have been absent in a single cross section are now captured. In sum, a pooled methodology offers the prospect of serious gains in statistical efficiency.

MODEL SPECIFICATION

In the seven studies at hand, there is little agreement on the proper independent variables, beyond the inclusion of economic development (E) itself, which is invariably postulated to have a curvilinear effect on democracy (D). A few variables do appear in more than one model: public control of the economy, Protestant population, historical timing of development, and world-system position (W). Each of these attains statistical significance in at least one model, which suggests that other social forces (S), as well, influence democratic performance. However, the only variable to achieve statistical significance at .05 in more than one study is world-system position (Bollen 1983; Bollen and Jackman 1985; Gonick and Rosh 1988). This suggests the conceptual specification:

\[ D = f(E, S, W) \]
where \( D = \) democracy, \( E = \) economic development, \( S = \) other social forces; \( W = \) position in the world-system.

We now proceed to the operationalization of these independent variables. For \( E \), we employ energy consumption per capita (logged), the economic development measure most often used, and one correlated .9 with gross national product per capita (Bollen 1979, 578). For \( S \), we utilize a proxy, the lagged value of democracy, \( D_t - 1 \). The rationale is straightforward. The other social forces acting on democracy are uncertain (and perhaps unmeasurable across the pool). However, to the extent they are structural processes (e.g., public control of the economy, Protestant population, historical events), they will be essentially summarized in the democratic performance of the nation during its previous year. In other words, \( D_t - 1 \) acts to control for omitted independent variables. With such a pervasive control in place, it is more difficult for spurious economic effects to be reported. (Still, Bollen (1979) found economic development significantly related to democracy in a 1965 cross section, even after including the 1960 democracy index as a predictor.)

With regard to a measure for \( W \), a few preliminaries are in order. The notion of a world system represents a major theoretical step forward in the literature. Accordingly, a nation’s chances for democracy may depend on where it is placed within the world economy. (For the theoretical inspiration, see Wallerstein 1974.) The core economies (high wages, high profits, capital-intensive), are bound in a lopsided exchange with the periphery economies (low wages, low profits, and labor-intensive). The constraints imposed by this international economic system (so the argument goes) seriously diminish the likelihood of successful democracy in the periphery.

Operationalization of the world-system position differs somewhat from study to study, but always involves dummy variables for different sets of countries (e.g., semiperiphery \( M \) and periphery \( P \)) entered additively into the estimation equation. Sometimes, expected effects have fallen short of statistical significance. Further, even when significance is attained, the form of the effect is seriously misspecified.

The last point deserves amplification. Gonick and Rosh who are leading world-system effects advocates, argue that “the effects of economic development are fundamentally shaped by the country’s position vis-à-vis the World-Economy” (1988, 195). This is an unambiguous argument for the world-system position’s operating nonadditively. In particular, world-system position conditions the impact of economics on democracy. To incorporate this dependency, then, the preferred specification is an interaction of the world-system position dummies with the economic development variable. This specification should actually increase the likelihood of uncovering world-system effects. (Details on construction of the world-system categories appear in Appendix B.)

These considerations lead to the model

\[
D_t = a + bD_{t-1} + cE_t + d(M \times E_t) + e(P \times E_t) + u_t, \quad (1)
\]

where \( D_t \) is the democracy index at time \( t \); \( D_{t-1} \) is the democracy index from the year before; \( E_t \) is energy consumption per capita (logged to the base 10) at time \( t \); \((M \times E_t)\) is the dummy variable for semiperiphery status multiplied by \( E_t \); \((P \times E_t)\) is the dummy variable for periphery status multiplied by \( E_t \); \( a, b, c, d, \) and \( e \) are parameters to be estimated; and \( u_t \) is an error term.

**MODEL ESTIMATION**

When OLS is applied to a pool of cross-sectional and time-series data, the violation of classical regression assumptions is highly likely. In fact, Stimson claims that autocorrelation and heteroskedasticity are virtually “inherent” in pooled data. (1985, 919). (Unfortunately, Arat [1988] relied exclusively on OLS in her pooled analysis; Gonick and Rosh [1988] ignore the autocorrelation problem.) From the time-series component, we must anticipate the former problem, from the cross-section component, the latter. Therefore, our estimation utilizes what Kmenta (assuming the same autocorrelation per unit) calls a “cross-sectionally heteroskedastic and timewise autoregressive model” (1986, 618), appropriately modified to cope with the special problem of having a lagged dependent variable, \( D_t - 1 \), as a predictor. (This condition also renders the estimates of another common technique, least squares dummy variables (LSDV), inconsistent [Hsiao 1986, 72]). We shall treat the question of the lagged dependent variable, then go on to assess heteroskedasticity. Finally, we shall develop the case for the preferred estimator, GLS-ARMA.

Because \( D_t - 1 \) is on the right-hand side of the equation, usual methods (e.g., the Durbin-Watson test) for diagnosing autocorrelation are misleading. However, when the appropriate \( h \)- and \( m \)-tests were applied, we found the OLS estimates exhibited significant first-order autocorrelation (Ostrom 1990, 66). Therefore, we turned first to an instrumental variables (IV) approach. These results provided an estimate of this first-order autocorrelation (\( p = .90 \)).

Furthermore, these IV–OLS estimates revealed highly significant cross-sectional heteroskedasticity, \( \text{LM} = 1,776.86 \) (on this Breusch-Pagan LaGrange Multiplier statistic, see Greene 1990, 465-69).

Thus, the variables were adjusted for this first-order autocorrelation, and heteroskedasticity was corrected (with the “force homoscedastic model” option in Microcrunch). The estimation procedure was GLS-ARMA (to allow a final check on any residual autocorrelation). In addition, we evaluated the possibility of mutual unit correlation, concluding that it was not a serious threat.

These GLS-ARMA estimates appear below. The residuals from this estimation were scrutinized in an examination of the pooled diagnostics. Following
Stimson, we paid special attention to the summary value of unit residuals, the ratio of residual variance, and any leftover nonstationarity in the autocorrelation function (1985, 939). According to these three diagnostic criteria, GLS–ARMA does edge out rival estimators: IV–OLS, GLS–error components (GLSE), or LSDV.¹⁰

\[ D't = .35* + .09*D't - 1 + 2.49*E't \]

\[
\begin{array}{ccc}
(0.06) & (0.02) & (0.22) \\
5.89 & 3.97 & 11.43 \\
- 1.33*(M \times E't) - 1.54*(P \times E't)' \quad (2)
\end{array}
\]

with pseudo-R = .71 and N = 2096, where all variables are defined as in equation 1 but transformed (as indicated by the ') on the basis of the instrumental variable first-order autocorrelation estimate, then submitted to GLS–ARMA estimation, under the “force homoscedastic model” in Microcrunch. The figures in parentheses are standard errors, * means statistical significance at .01 (t > 2.58), the figures in brackets are absolute r-ratios, and pseudo-R is the correlation between the observed Dt and the Dt predicted from using original equation 1 variables with equation 2 coefficients. Sources for the democracy and world-system data are described in the Appendix B. The energy consumption data were gathered from United Nations publications.

On the basis of these results, economic development has a highly significant impact on democratic performance. Moreover, when that impact is conceived of as interactive, rather than additive, the interaction formulation has the edge.⁵ Economic development matters most for nations in the core, where c = 2.49; it still matters, but about half as much, in the semiperiphery (2.49 - 1.33 = 1.16 net economic effect); and for nations in the periphery, the economic effect is just a bit less (2.49 - 1.54 = .95). Taken together, economic factors, both international and domestic, appear decisive in shaping a nation’s democratic future. Still, there is not economic determinism. Other important social forces, captured at least partially by past values of the democracy index itself, exercise an effect. The model, by blending economic with social explanation, manages a good fit to the data, as indicated by the pseudo-R. But to expand the implications of the model, three methodological concerns need to be addressed: measurement error, ceiling effects, and causality.

METHODOLOGICAL CONCERNS

An assumption underlying the estimates of equation 2 (as well as the estimates from other quantitative democracy studies) is that the variables are measured without error. However, common measures of democracy have considerable error. As mentioned, Bollen (1993) demonstrated substantial error in the Banks measures. He also explored Gastil measures (essentially the two components of our study), assigning political rights a validity rating of 93% and civil liberties a validity rating of 78%. Encouragingly, these validity scores have a higher average (86%) than the Banks indicators (average validity = 62%). However, they are not error-free. According to Bollen, the two Gastil components have systematic error, respectively, of 7% and 16%, induced by the Gastil rating method itself.

The question is whether this “methods factor” (as Bollen calls it), even though relatively small, seriously biases our inference about economic effects. Some critics of Gastil have argued, on the basis of their impressions, that the measures he provided Freedom House have a conservative bias. For example, nations that are staunch American allies are alleged to get more democratic ratings than they deserve (Hartman and Hsiao 1988). Using 1980 data, Bollen (1993) assesses the methods bias in the Gastil measure nation by nation, where a high standardized score implies that the judge is too favorable, a low score, not favorable enough, and zero, a neutral judgment. If the Gastil measure has a conservative bias, then that could bias the economic development parameter estimate. To test this possibility, we sequentially reestimated equation 2, progressively eliminating nations with positive scores (i.e., first greater than 2 standard deviations above the neutral value of 0, then greater than 1.5, then greater than 1.0). The economic development coefficient remained statistically significant, its magnitude virtually unchanged: 2.51, 2.52, and 2.52, respectively. Thus, there is no evidence that conservative bias in the Gastil measure, to the extent it exists, has biased the economic development coefficient. (We repeated the same test with the negative scores, in case of a possible liberal bias, and again concluded that the economic development coefficient was unaffected.)

Of course, methods factor error is not exclusively a question of conservative or liberal bias. Other issues, such as the occasional adjustment of democracy scores to reflect poor economic conditions, have been raised Freedom in the World (1978:13). What would be useful is some general way of comparing our Gastil measure to a “true” measure of democracy. Fortunately, Bollen has constructed an index for 1980 that is highly correlated with the latent variable of democracy and virtually uncorrelated with the distorting methods factors (1993, app.). These uncontaminated 1980 index scores may be rather close to “true” democracy scores. While available only for 1980, they nevertheless provide invaluable baseline data. We find that this 1980 Bollen democracy index correlates .93 with our Gastil democracy measure for the same year. (All nations were measured on both, except for one, giving N = 130.) This suggests that very little error remains in the Gastil index as constructed, since it correlates almost perfectly with Bollen’s virtually error-free democracy measure. In other words, the
While the Gastil index has considerable validity, it is not flawless. Like other leading democracy scales, it has limited refinement. Nations can score only so high and no higher. For instance, in the widely used 1965 political democracy index, also created by Bollen, 27 out of the 123 nations sampled scored at the top, 90–100 (1980, 387–88). Of course, these scores reflect the scale “ceiling,” not the attainment of ideal democracy. Our Gastil measure allows a bit more variation here, with only 20 out of the 131 nations attaining a top score of 13 or 14 across the time period. Moreover, because of the time-series component of the design, nations obviously can attain more than one score. Overall, the added variance on the dependent variable afforded by this pooled design makes for more efficient statistical inference.

A last methodological concern is over causality, an issue that plagues this literature. Does economic development “cause” democracy? Or does democracy really “cause” economic development? The cross-sectional work has had to assume causality runs from economics to democracy, while the limited-time-series work has not fully utilized the testing opportunities that temporal observation allows. Fortunately, Granger (1969, 1988) offers some straightforward causality tests applicable to time series. Accordingly, if X causes Y, then past values of X should aid in the prediction of Y (even after controlling for past values of Y). Further, Y should not aid in the prediction of X.

To carry out the test (here extended to these pooled time-series data), we first established the optimal number of lags for the independent variables, using the final prediction error test (see Mahdavi and Sohrabian 1991, 44–45). With regard to predicting democracy (Y), optimal lagged variables turn out to be $Dt - 1$, and $Et - 1$. With regard to predicting economics (X), the optimal lagged variables are more elaborate, $Et - 1 ... Et - 6, Dt - 1$. The block $F$-tests on the unrestricted versus restricted $OLS$ equations on the pool are, for democracy, $F (1,2224) = 18.18$ (significant at .05), and for economics, $F (1,1572) = .00$ (not significant at .05).

According to the first test, the lagged economic development variable is a significant predictor of democracy even after controlling for past values on democracy. Therefore, we reject the hypothesis that economic development does not cause democracy. Following the second test, the lagged democracy variable is not a significant predictor of economic development after controlling on past values of economic development. Therefore, we accept the hypothesis that democracy does not cause economic development. These Granger results seem especially robust, even surviving controls on world-system position.6

CONCLUSION

The economic development thesis, long a staple of comparative democracy studies, has recently come under challenge. The pooled-time-series studies of Arat (1988) and Gonick and Rosh (1988), while in some ways the most advanced quantitative pieces in this critical literature, appear to be characterized by five flaws: (1) they measure democracy with error-laden data (2) the data do not extend past the 1970s, (3) the models are misspecified, in particular with regard to world-system-position interaction effects, (4) estimation procedures are not dynamic and fail to utilize more efficient estimation techniques such as GLS-ARMA, and (5) causality tests are not administered. In response, we have analyzed a very large sample of nations (131) over a long period of time (1972–89), employing a relatively neglected data set on democracy. After refining the model specification, a pooled design was applied, with final parameter estimates provided by the GLS-ARMA technique. The results were demonstrated to be methodologically robust. In particular, the Gastil democracy index has considerable measurement validity and promises to be increasingly used in future work.

We established that the causal arrow most probably runs from economic development to democracy, rather than vice versa. Further, that effect is highly significant statistically. Substantively, the economic development coefficient of equation 2 implies that for every tenfold increase in per capita energy consumption, the nation could expect about a two-and-a-half-point rise on the democracy scale. How important an effect is this? A standard deviation increase in per capita energy consumption (logged) is predicted to yield, on average, a .48 standard deviation change in the democracy score. By this measure, the economic impact is not small, and as effects cumulate over time, the impact can grow. A simple measure of importance comes from correlating average (over nations) world energy consumption with the average (over nations) world democracy score over time (1972–89), $r = .82$.

On balance, it is clear that economic development substantially improves a nation’s democratic prospects. However, the full magnitude of that effect depends on the location of the nation in the world system. As the nation moves from the core, to the semiperiphery, to the periphery, the effect diminishes. Even in the periphery, however, the effect remains statistically and substantively significant. Thus, around the world, economic development works to foster democracy. Indeed, our Granger results indicate that the relationship works in that direction but not the other. To the extent that this finding holds for nations currently in democratic transition, the implication is that democratic reform by itself cannot be counted on to bring about the needed economic development. However, this is no counsel to dictatorship. Just as clearly, we found that democracy, while not apparently a direct cause of economic development, certainly does it no harm.
Moreover, as the lag pattern of the structural model shows, past democratic performance breeds future democratic performance. Democracy, then, can be furthered for its own sake, without sacrificing economic development.

**APPENDIX A: THE GASTIL DEMOCRACY DATA**


**Checklist for Political Rights**

1. Chief authority recently elected by a meaningful process
2. Legislature recently elected by a meaningful process
3. Fair election laws, campaigning opportunity
4. Fair reflection of voter preference in distribution of power
5. Multiple political parties
6. Recent shifts in power through elections
7. Significant opposition vote
8. Freedom from military or foreign control
9. Major groups allowed reasonable self-determination
10. Decentralized political power
11. Informal consensus, de facto opposition power

**Checklist for Civil Liberties**

1. Media/literature free of political censorship
2. Open public discussion
3. Freedom of assembly and demonstrations
4. Freedom of political organization
5. Nondiscriminatory rule of law in politically relevant cases
6. Free from unjustified political terror or imprisonment
7. Free trade unions or peasant organizations
8. Free businesses or cooperatives
9. Free professional or other private organizations
10. Free religious institutions
11. Personal social rights

**APPENDIX B: THE SAMPLE OF NATIONS**

Nations are listed with their range of democracy scores constructed as described, from the material in Appendix A (2-14 being the widest possible range) and their world-system Position (c = core, m = semiperiphery, p = periphery). To construct the world-system-position dummies, we combined the ratings in nine studies, following a “panel of experts” strategy (Arrighi and Drangel 1986; Bollen 1983; Chase-Dunn 1983; Chirot 1977; Frank 1969; Gonick and Rosh 1988; Nemeth and Smith 1985; Smith and White 1992; Snyder and Kick 1979). There was high agreement among the judgments. However, when discrepancies occurred, more weight was give to the more recent evaluations. As can be seen, the classification has high “face validity”; and when we experimented with alternate classifications, the results did not change substantively.

Afghanistan 2-7 p, Albania 2-2 p, Algeria 3-6 p, Argentina 4-13 m, Australia 14-14 c, Austria 14-14 c, Bahamas 11-13 p, Bahrain 5-8 p, Bangladesh 4-10 p, Barbados 13-14 p, Belgium 14-14 c, Benin 2-4 p, Bolivia 4-11 p, Brazil 6-12 m, Bulgaria 2-4 m, Burksina Faso 3-11 p, Burma 2-4 p, Burundi 2-4 p, Cambodia 2-5 p, Cameroon 3-6 p, Canada 14-14 c, Central African Republic 2-4 p, Chad 2-4 p, Chile 4-13 m, China 2-4 p, Colombia 9-12 m, Congo 2-5 p, Costa Rica 14-14 p, Cote d’Ivoire 4-6 p, Cuba 2-4 p, Czechoslovakia 2-4 m, Denmark 14-14 c, Dominican Republic 9-13 p, Ecuador 4-12 p, Egypt 4-8 p, El Salvador 6-11 p, Equatorial Guinea 2-4 p, Ethiopia 2-5 p, Fiji 5-12 p, Finland 12-14 m, France 13-14 c, Gabon 4-5 p, Ghana 3-11 p, Greece 4-13 m, Guatemala 4-12 p, Guinea 2-4 p, Guyana 6-12 p, Haiti 3-7 p, Honduras 6-11 p, Hungary 4-9 m, Iceland 14-14 m, India 9-12 m, Indonesia 4-6 p, Iran 4-6 m, Iraq 2-3 m, Ireland 13-14 m, Israel 11-12 m, Italy 12-14 c, Jamaica 11-13 p, Japan 13-14 c, Jordan 4-6 p, Kenya 4-7 p, Kuwait 5-9 m, Laos 2-6 p, Lebanon 5-12 p, Liberia 4-7 p, Libya 2-4 s, Luxembourg 13-14 c, Madagascar 4-8 p, Malawi 3-3 p, Malaysia 5-11 m, Mali 2-4 p, Malta 10-14 p, Mauritania 3-5 p, Mauritius 10-12 p, Mexico 8-9 m, Mongolia 2-2 p, Morocco 6-9 p, Nauru 12-13 p, Nepal 5-9 p, Netherlands 14-14 c, New Zealand 14-14 c, Nicaragua 4-9 p, Niger 3-4 p, Nigeria 4-11 p, North Korea 2-2 p, North Yemen 4-8 p, Norway 14-14 c, Oman 3-4 p, Pakistan 4-10 p, Panama 3-9 p, Paraguay 4-9 p, Peru 4-11 p, Philippines 6-12 p, Poland 4-9 m, Portugal 5-13 m, Qatar 4-6 m, Romania 2-3 m, Rwanda 3-4 p, Saudi Arabia 3-4 m, Senegal 4-9 p, Sierra Leone 5-7 p, Singapore 6-8 m, Somalia 2-3 p, South Africa 5-7 m, South Korea 5-11 m, Spain 5-14 m, Sri Lanka 7-12 p, Sudan 2-7 p, Sweden 13-14 c, Switzerland 14-14 c, Syria 2-5 p, Tanzania 4-4 p, Thailand 4-11 p, Togo 2-4 p, Tonga 8-10 p, Trinidad 11-14 p, Tunisia 5-8 p, Turkey 6-11 m, Uganda 2-7 p, United Kingdom 14-14 c, United Arab Emirates 4-6 m, Uruguay 4-13 p, United States 14-14 c, USSR 2-5 m, Venezuela
12–13 m, Yugoslavia 4–7 m, Zaire 2–4 p, Zambia 5–7 p, Zimbabwe 5–8 p.

Notes

We express special appreciation to B. Dan Wood for his advice on estimation for this pooled time series design.

1. The LSDV technique is also less preferred here because it is a "fixed effects" model, which implies that the observations cannot be treated as a random sample from a larger population. In contrast, these data compose a large, apparently representative, sample from a population (of nations and of time). Thus, country dummies should not enter our analysis wholesale. However, we do include selective introduction of geographic dummies (i.e., the periphery, semiperiphery variables). These variables, in their interaction form, explicitly allow for slope heterogeneity of economic effects across nations but do so in a statistically efficient and theoretically defined way. (To confirm our economic effects results, another treatment for differing slopes or intercepts across nations—the random effects model of GLS-error components—was eventually applied. See n. 4.)

2. First, OLS was applied to the pooled of variables in equation 1. The Durbin's h = 7.87 indicated the need to reject at the .05 level the hypothesis of no first-order autocorrelation. The m test also showed significant first-order autocorrelation but no significant second-order autocorrelation. Finally, more general Breusch-Godfrey testing revealed this significant AR(1) process but uncovered no other processes of any order. To estimate this first-order autocorrelation, we used an instrumental variables approach (see Ostrom 1990, 65–71). An instrumental variable, \(Dt - 1\), was constructed (from the exogenous variables of equation 1, lagged to \(t - 1\)). The instrument appears a good proxy, \(Dt - 1\) correlating .64 with \(Dt - 1\), for which it is a stand-in.) This \(Dt - 1\) was substituted for \(Dt - 1\) in equation 1, and OLS was applied to achieve consistent estimates of the parameters and the residuals. Using these residuals, we arrived at an estimated pooled rho (the first-order autocorrelation) of .90.

Having thus obtained a desirable estimate of the troublesome first-order autocorrelation, we moved on to correct the original equation for it. Following a Cochrane-Orcutt-type procedure, each original variable of equation 1 was transformed (e.g., \([Dt - 1] - [1.90(Dt - 2) - 2] / 1.90\)). These transformed variables were then submitted to the comparative pooled estimations (see n. 4).

3. Besides heteroskedasticity and no autocorrelation, the assumption of the mutual independence of the units can also be a concern with analysis of pooled data. Here we are dealing with a sample of independent nations, rather than a population of more homogeneous geographic units like the American states (Kmenta 1986, 622–25). Therefore, we did not suspect a mutual unit correlation problem. Nevertheless, we attempted to correct for it, utilizing SHAZAM but were initially thwarted by the very high correlation between the error vectors of certain nations. One solution—exploration of the phi matrix to identify the offending high correlations—poses its own set of problems, involving a search through about 8,500 correlations (from our 131 \(\times\) 131 country matrix). Another solution—eliminating countries by trial and error and reestimating—not only raises practical difficulties but would hardly call into question the integrity of the remaining sample of nations. In the end, we chose to stick with our GLS-ARMA estimates, which have been corrected for autocorrelation and heteroskedasticity. We are quite comfortable with this because, in general, a correction for mutual unit correlation increases efficiency and makes for higher levels of statistical significance. The reported GLS-ARMA results of equation 2, then, can actually be viewed as a conservative test.

4. For purposes of statistical comparison, OLS, LSDV and GLSE were applied to the variables of equation 1, as noted in the text of the autocorrelation described in n. 2. (Recall that Arat [1988] relied on OLS, Gonick and Rosh [1988] on GLSE estimation.) The OLS results indicated significant autocorrelation remaining, according to the autocorrelation function (back five lags). The autocorrelation function of the LSDV estimator also revealed a slight amount of nonstationarity remaining, with a statistically significant first-order autocorrelation estimate of .06. GLSE makes no correction for serial correlation (or heteroskedasticity) but might be considered more efficient than LSDV for taking into account any intercept differences (which OLS ignores). GLSE in fact yielded a parameter estimate of (2.54) for \(Et\), almost identical in value to that in equation 2. However, it yielded a standard error of .30, clearly indicating the efficiency gains of the GLS-ARMA (heteroskedasticity-corrected) estimate of equation 2, with its lower standard error of .22.

Further, other residual diagnostics for these GLS-ARMA results are favorable. With GLS-ARMA, the standard deviation of the residual means is only .34, compared to .37 for GLSE. (The residual variance ratios show essentially the same performance level with, for example, only 9 of 131 scores exceeding 3.0 in GLSE or GLS-ARMA.) These residual departures seem small and bolster our confidence in the GLS-ARMA model estimation and specification. (A healthy pattern of residual diagnostics from a GLS-ARMA analysis led Stimson to a similar conclusion regarding his party issue polarization model [1985, 945–44, tbl. 7]. However, pooled diagnostics provide no sharp boundaries for deciding whether a model is "right" or "wrong." As Stimson goes on to note, at some point it becomes "a judgment call" (p. 943). We experimented with treating the few departing cases as outliers, either eliminating them from the estimation equation or introducing dummy independent variables. We could not improve on the rather robust diagnosis already obtained with the GLS-ARMA estimates of equation 2.

5. Ideally, we would have a fully specified model, including the additive terms for world-system position (\(M\) and \(P\)) in equation 2 along with the interaction terms. We estimated this fully specified model, but severe collinearity (e.g., \(Et\) regressed on the other independent variables) yields an R-squared = .994) rendered the coefficients nonsensical. Because this problem sometimes occurs in the presence of interaction terms, we applied the possible solution of centering the variables, but the results failed to improve (Jaccard, Turrisi, and Wan 1990, 31). As an alternative, we estimated a main-effects model, for comparison to the interaction effects model of equation 2. Using the same variable definitions, statistics, corrections, and estimation methods (GLS-ARMA, "force homoscedastic model") option as in eq. 2, the results are as follows:

\[
D't = .97' + .08'Dt - 1 - 1.89'E't - 4.86'M' - 6.02'P',
\]

(3)

(.10) (.02) (.21) (.69) (.68)

[9.54] [3.80] [4.29] [7.06] [8.80]

with pseudo-R = .70 and N = 2,096.

In comparing these results to equation 2, we observe that the coefficient for the key economics variable, \(Et\), is much less significant than under the interaction specification. To appreciate this, first use the standard errors to construct confidence intervals (95%, two-tailed) around the \(Et\) coefficient in each equation. For equation 2, this confidence interval equals 2.49 +/- (1.96 \(\times\) .22) = 2.06–2.92; for equation 3, this confidence interval equals .89 +/- (1.96 \(\times\) .21) = .48–1.30. One observes that for the interaction specification of equation 2, the lower bound of the confidence interval is much farther from zero, giving much more confidence in rejecting the null hypothesis. The greater statistical security of the other t-ratios show, the precision of the other coefficients is not lessened.

6. The Granger approach, by only allowing an X to account for Y after past values of Y itself have accounted for all they can be a very conservative test. Hence, the survival of economic effects in such a test gives us some confidence in their existence. Because its purpose is to offer a demanding "test for causality," rather than to estimate a structural model, comparisons of Granger results to structural model estimates such as those of equation 2 can be misleading. However, it is
Economic Development and Democracy

possible that even strong Granger findings, such as these, are spurious because of the operation of a third variable (Pindyck and Rubinfeld 1991, 217). The obvious 'third variable' candidate is world-system position. Therefore, as a further test, we included the world-system interaction terms in the unrestricted Granger regression:

\[ D_t = 0.02 + 0.96 D_{t-1} + 1.46 E_{t-1} - 0.03 (M \times E_t - 1) \]

\[ (.09) \quad (.01) \quad (.03) \quad (.02) \]

\[ [.22][160.35] [4.13] [1.63] \]

\[ -.05(P \times E_t - 1), \quad (4) \]

\[ (.02) \]

\[ [2.04] \]

with adjusted R² = .953, N = 2,227, and Box-Ljung Q (distrib. chi-squared, 5 df) = 5.77, where the variables are defined and measured as in equation 1. The figures in parentheses are standard errors, the figures in brackets are absolute t-ratios, and the Box-Ljung Q is a measure of autocorrelation.

These results make clear how demanding the Granger-test is. We observe that if we impose no theoretical structure on the model and operate by statistical brute force, past values of democracy appear almost totally predictive of democracy. Nevertheless, even under this extreme (and unrealistic) structure, economic development manages statistically significant main effects, as well as statistically significant world-system interaction effects. The t-ratios are helpful here. For a one-tailed test, significance at .05 requires \( t > 1.64 \), and significance at .10 requires \( t > 1.28 \). Also, the block F-test is still significant at .05, \( F(3,2222) = 11.22 \) (a number slightly lower than the original because of different degrees of freedom).

Again, we find that economic development is a determinant of democracy, irrespective of world-system position.

References


Ross E. Burkhart is Doctoral Candidate and Michael S. Lewis-Beck is Professor of Political Science, University of Iowa, Iowa City, IA 52242.