AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Have Your Cake and Eat It Too? Cointegration and Dynamic Inference from Autoregressive Distributed Lag Models

Andrew Q. Philips University of Colorado at Boulder

Abstract: *Although recent articles have stressed the importance of testing for unit roots and cointegration in time-series analysis, practitioners have been left without a straightforward procedure to implement this advice. I propose using the autoregressive distributed lag model and bounds cointegration test as an approach to dealing with some of the most commonly encountered issues in time-series analysis. Through Monte Carlo experiments, I show that this procedure performs better than existing cointegration tests under a variety of situations. I illustrate how to implement this strategy with two step-by-step replication examples. To further aid users, I have designed software programs in order to test and dynamically model the results from this approach.*

Replication Materials: The data, code, and any additional materials required to replicate all analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/MPQQC0.

Recent work in the time-series literature has stressed the importance of testing for unit roots as well as the existence of long-run relationships—or cointegration—between variables.[1] Since the presence or absence of each of these characteristics ultimately determines the appropriate model, failure to perform such pretesting makes spurious inferences more likely. Even with existing tools designed to identify unit roots and test for cointegration, short series, the weak power of statistical tests, and the dangers of overfitting make pretesting time-series data particularly problematic. Although recent articles have helped to identify these issues (Grant and Lebo 2016; Keele, Linn, and Webb 2016), users have been left without a straightforward solution about how to deal with such problems.[2]

I propose using the autoregressive distributed lag model and associated bounds testing procedure (ARDL-bounds) developed by Pesaran, Shin, and Smith (2001) as a comprehensive approach to model specification and cointegration testing. Depending on the results of the cointegration test, this strategy absolves users from having to distinguish between stationary (henceforth I(0)) and first-order nonstationary (I(1)) regressors. This is an advantage since unit root testing is difficult in short series and introduces "a further degree of uncertainty into the analysis" (Pesaran, Shin, and Smith 2001, 289). The ARDL-bounds procedure involves the following:

1. Ensuring the dependent variable is I(1).
2. Ensuring the independent variables are not explosive or higher orders of integration than I(1).
3. Estimating the ARDL model in error correction form, and ensuring there is no autocorrelation.
4. Performing the bounds test for cointegration. Three possibilities result: (a) all regressors are I(1) and cointegrating, (b) all regressors are I(0)—by definition, they cannot cointegrate—or (c) indeterminate. An indeterminate result may

Andrew Q. Philips is assistant professor, Department of Political Science, University of Colorado at Boulder, UCB 333, Boulder, CO 80309-0333 (andrew.philips@colorado.edu).

I would like to thank Lorena Barberia, Allyson Benton, Harold Clarke, Peter Enns, Nathan Favero, Eric Guntermann, Mark Pickup, Joe Ura, B. Dan Wood, and participants of the Texas A&M methodology brownbag lunches. Special thanks go to Soren Jordan, Paul Kellstedt, and Guy D. Whitten. Despite this helpful advice, any errors and omissions remain my own.

[1] Covariance stationary series exhibit constant mean, variance, and covariance. A linear combination of two or more first-order nonstationary series that yields a stationary series is said to be cointegrating.

[2] Grant and Lebo (2016) provide two solutions, including the one discussed herein. However, their discussion is brief.

American Journal of Political Science, Vol. 62, No. 1, January 2018, Pp. 230–244

 DOI: 10.1111/ajps.12318

still find cointegration among some of the independent variables, although further testing and respecification (in Step 3) is required.

Surprisingly, while this method is popular in other fields (over 5,300 cites on Google Scholar as of September 2016), it has been cited and implemented only twice among *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, and *Political Analysis*: Dickinson and Lebo (2007) and Grant and Lebo (2016).

Four contributions stand out in this article. First, I discuss why an additional time-series procedure is necessary, given recent debates about the role of error correction models (Esarey 2016; Grant and Lebo 2016; Helgason 2016; Keele, Linn, and Webb 2016). Second, I use Monte Carlo experiments to compare the performance of the ARDL-bounds cointegration test against existing alternatives, under a variety of scenarios that practitioners typically encounter. I also examine how well the model recovers substantively interesting effects, such as long-run multipliers or adjustment parameters. Third, I demonstrate the utility of the ARDL-bounds approach and the merits of dynamic interpretation through two replications. Finally, I conclude with guidelines for implementing this procedure and introduce software programs designed to help practitioners with cointegration testing and exploring the substantive implications of their results.

Unit Roots and Cointegration in Time-Series

Consider a general autoregressive distributed lag ARDL(p, q) model where a series, y_t, is a function of a constant term, α_0, past values of itself stretching back p periods, contemporaneous and lagged values of an independent variable, x_t, of lag order q, and an independent, identically distributed (i.i.d.) error term:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=0}^{q} \beta_j x_{t-j} + \epsilon_t,$$
$$\epsilon_t \sim N(0, \sigma^2). \quad (1)$$

The data generation process for the dependent and independent variables determines how Equation (1) is estimated. If variables on both the left- and right-hand sides are I(0), they will exhibit constant mean, variance, and covariance, and the ARDL(p, q) shown in Equation (1) may be used.[3] Since additional lags may induce multicollinearity, lag order restrictions are often imposed. A common restriction is the ARDL(1,1) model:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t. \quad (2)$$

The contemporaneous effect of x_t on y_t is given by β_0. The magnitude of α_1 informs us about the "memory" of y_t (De Boef and Keele 2008). Assuming $0 < \alpha_1 < 1$, larger values indicate that movements in y_t take longer to dissipate.[4] The long-run effect (or long-run multiplier) is the total effect that a change in x_t has on y_t. It is given as $\kappa_1 = \frac{(\beta_0+\beta_1)}{(1-\alpha_1)}$, and its variance is typically approximated using the delta method.

The generalized error correction model (GECM) may also be used if all variables are I(0); the most common form is the one-step GECM:

$$\Delta y_t = \alpha_0 + \alpha_1^* y_{t-1} + \beta_0 \Delta x_t + \beta_1^* x_{t-1} + \epsilon_t, \quad (3)$$

where the first difference of y_t is a function of a constant term, α_0, its own lag, y_{t-1}, the first difference of x_t and its lag, x_{t-1}, and an i.i.d. error term, ϵ_t. Although the GECM is algebraically equivalent to the ARDL(1,1) model, interpretation changes. Contemporaneous effects of a change in x_t on y_t are still given by β_0. The rate of adjustment, or the speed at which the total effect of a change of x_t accumulates in y_t, is given by α_1^*. It is used in calculating the long-run multiplier, $\kappa_1 = -\frac{\beta_1^*}{\alpha_1^*}$. Although obtaining variance estimates of the short-run effect is straightforward, the variance around κ_1 must be approximated using the Bewley transformation or the delta method (De Boef and Keele 2008).

The GECM is also ideal for when the dependent and independent variables are I(1) *and* cointegrating. In our bivariate example, if there exists some linear combination of the two I(1) series that results in a stationary series, they are said to be cointegrating. Testing is often performed using the Engle-Granger "two-step" approach (Engle and Granger 1987), which involves regressing y_t on x_t:

$$y_t = \kappa_0 + \kappa_1 x_t + z_t. \quad (4)$$

If both variables are I(1), there exists one cointegrating relationship if the residuals in Equation (4), z_t, are stationary.[5] More generally, a sufficient condition in which to use an error correction model is if all variables are I(1) and cointegrating.[6]

[3]The stationarity condition for y_t is given as $|\sum_{i=1}^{p} \alpha_i| < 1$. Such variables are said to be covariance stationary.

[4]Values of α_1 greater than one suggest an explosive series or a model mis-specification. Values less than zero suggest the series is overcorrecting or oscillating; this is rare in the social sciences.

[5]This is true for any k series, which can have up to $k - 1$ cointegrating relationships.

[6]This condition is sufficient but not necessary; one could use other models (e.g., first differences). I focus on I(1) series since higher orders of integration are rare in political science, although this

Even if both series are I(1), there may not always be an underlying cointegrating relationship between them. Practitioners often conflate re-equilibration with error correction and fail to test for cointegration (Grant and Lebo 2016).[7] Even if x_t and y_t are I(1), without cointegration, there cannot be a long-run relationship between them since (rewriting Equation 4) the linear combination of the series, $z_t = (y_{t-1} - \kappa_0 - \kappa_1 x_{t-1})$, will not be stationary. If all variables are I(1) but *not* cointegrating, the series can only be analyzed in first differences since a short-run relationship may still exist

The recommendations above are straightforward in theory. In practice, identifying the correct model is nontrivial. For one, unit root tests often have size distortions and low power in small samples, making it difficult to determine whether a variable is I(0) or I(1) (Choi 2015; Maddala and Kim 1998). This difficulty is compounded since users must test each variable in order to use models such as the GECM. Series may be so highly autoregressive (near-integrated) that testing procedures cannot distinguish them from an I(1) series (De Boef and Granato 1997). Moreover, series may be fractionally integrated. While some scholars argue that these are common in political science (Box-Steffensmeier and Smith 1998; Grant and Lebo 2016; Lebo, Walker, and Clarke 2000), others remain skeptical (Keele, Linn, and Webb 2016; Pickup 2009).[8] In other words, with short series (less than 100), we are often at the mercy of our tests, and we risk choosing models that are not reflective of the characteristics of our data.

As recent work has shown, many scholars have overlooked the crucial steps of testing for unit roots and cointegration (Grant and Lebo 2016). Others find that complex model specifications tend to overfit and perform poorly in small samples (Esarey 2016; Keele, Linn, and Webb 2016). While these important contributions have identified potential problems, they leave users without a clear and easy-to-implement solution. As I show in the next section, a procedure already exists that greatly eases unit root testing, includes a test for cointegration, and is simple to estimate. Moreover, when combined with dynamic simulations, these models can provide additional substantive interpretations.

excludes the possibility of multi-cointegration (Enders 2010, 380–82).

[7]While cointegrating relationships can be estimated using GECMs, estimating GECMs does not necessarily mean two or more series are cointegrated.

[8]Helgason (2016) and Esarey (2016) investigate treating data as fractionally integrated versus I(1) through Monte Carlo simulations.

A Comprehensive Approach to Time-Series Analysis

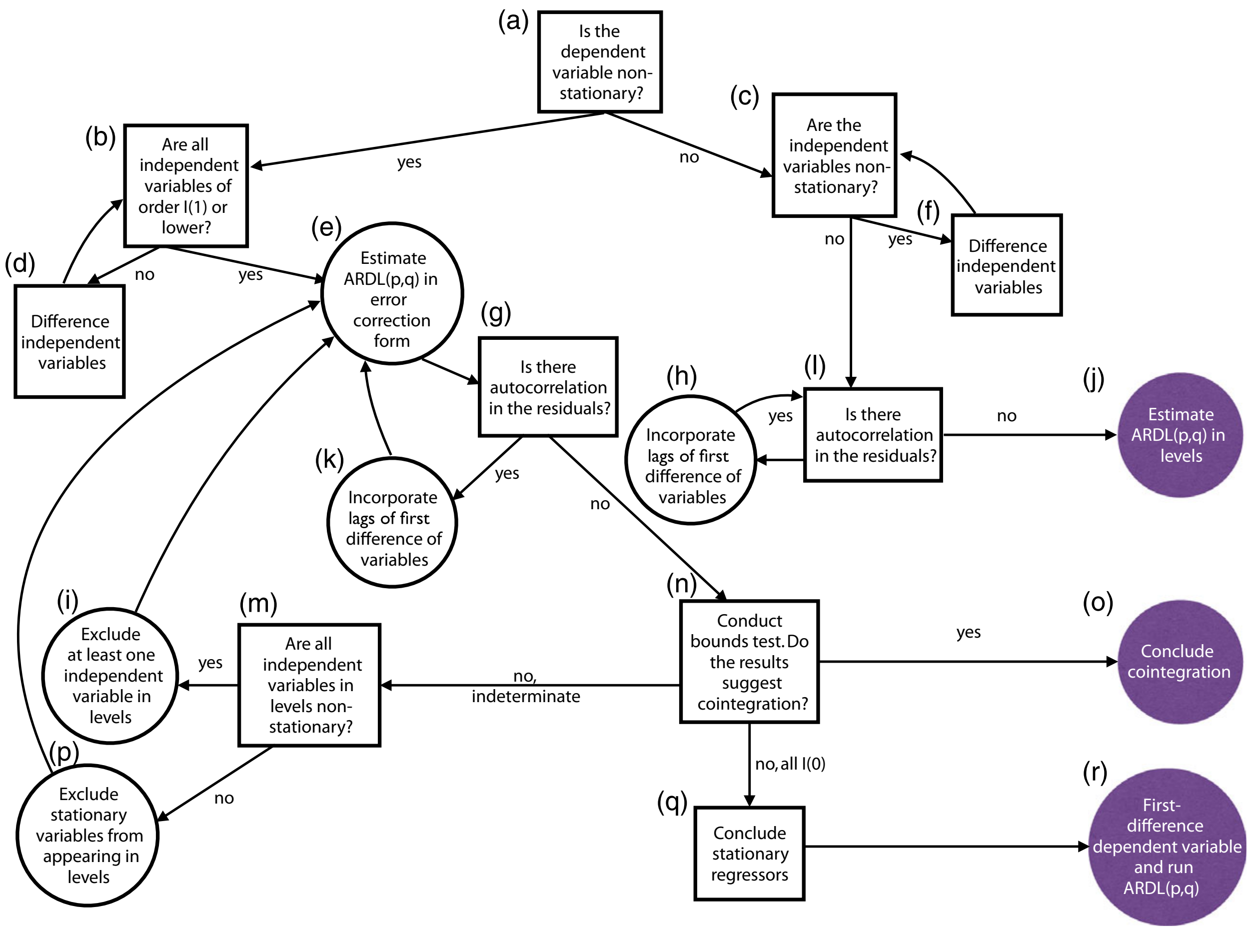
While the autoregressive distributed lag (ARDL) model and associated bounds test of Pesaran, Shin, and Smith (2001) comprise an approach already popular in economics, it remains relatively unknown in political science. It is ideal for four reasons. First, although we may suspect that all regressors are I(1), an initial model can be estimated without having to rely on unit root testing to distinguish between I(0) or I(1) regressors. Restrictions on the independent variables can then be imposed to avoid spurious conclusions of cointegration. Second, the one-step procedure for the initial cointegration test is similar to the GECM, making it easy to estimate. Third, the cointegration test is often straightforward to interpret. Fourth, this framework provides a comprehensive approach for practitioners.

The ARDL-bounds approach is shown in schematic form in Figure 1.[9] As shown in step *a*, users must first establish whether the dependent variable is I(1). To mitigate difficulties with unit root testing, users should employ a suite of unit root tests and account for the possibility of periodicity, drift, and deterministic trends. If the dependent variable is stationary, then cointegration is not possible and any I(1) regressors must be first differenced (step *f*). After ensuring that all independent variables are stationary (step *c*), we must also check that no autocorrelation remains in the residuals (step *i*). As shown by step *h* in Figure 1, if there is autocorrelation, we can incorporate lags of the dependent and independent variables, or lagged first differences if a regressor is I(1). Lag structures are typically chosen based on theoretical expectations about the data generation process, and by minimizing information criteria such as the Akaike Information Criterion (AIC) and Schwarz-Bayesian Information Criterion (SBIC) . If no autocorrelation remains, the resulting ARDL model is one where all variables are I(0), as shown in step *j*, a version of which was shown in Equation (1). There is no need to check for cointegration since all variables are stationary.

If the dependent variable is I(1), there may be cointegration. As shown in step *b* in Figure 1, we do not have to establish whether the regressors are I(0) or I(1); we of course suspect I(1), since we are testing for cointegration. However, we must ensure that there are no explosive series, seasonal unit roots, or series higher than I(1) in *any* of the variables. Violation of these conditions invalidates

[9]For brevity, I do not consider fractionally integrated relationships. I discuss strategies for handling these data in the supporting information.

FIGURE 1 The ARDL-Bounds Procedure's Comprehensive Approach to Time-Series Analysis



the testing procedure. Independent variables that are nonstationary of higher orders than I(1) must be differenced (step *d*) before moving forward.[10]

Next, estimate the ARDL model in error correction form (step *e*). Recall that a cointegrating relationship between an I(1) dependent variable, y_t, and a weakly exogenous I(1) regressor, x_t, can be written as.[11]

$$y_t = \kappa_0 + \kappa_1 x_t + z_t. \tag{5}$$

If the residuals, z_t, are stationary, there is evidence of cointegration.[12] In order to estimate this model, z_{t-1} is included in the following GECM:

$$\Delta y_t = \alpha_0 - \alpha(z_{t-1}) + \beta_0 \Delta x_t + \epsilon_t. \tag{6}$$

Rewritten, it becomes.

$$\Delta y_t = \alpha_0 - \alpha(y_{t-1} - \hat{\kappa}_0 - \hat{\kappa}_1 x_{t-1}) + \beta_0 \Delta x_t + \epsilon_t. \tag{7}$$

The unrestricted error correction model referred to by Pesaran, Shin, and Smith (2001, 293) forms the basis of the ARDL-bounds procedure. It involves multiplying through by $-\alpha$ and collecting terms in Equation (7):

$$\Delta y_t = \alpha_0^* + \theta_0 y_{t-1} + \theta_1 x_{t-1} + \beta_0 \Delta x_t + \epsilon_t, \tag{8}$$

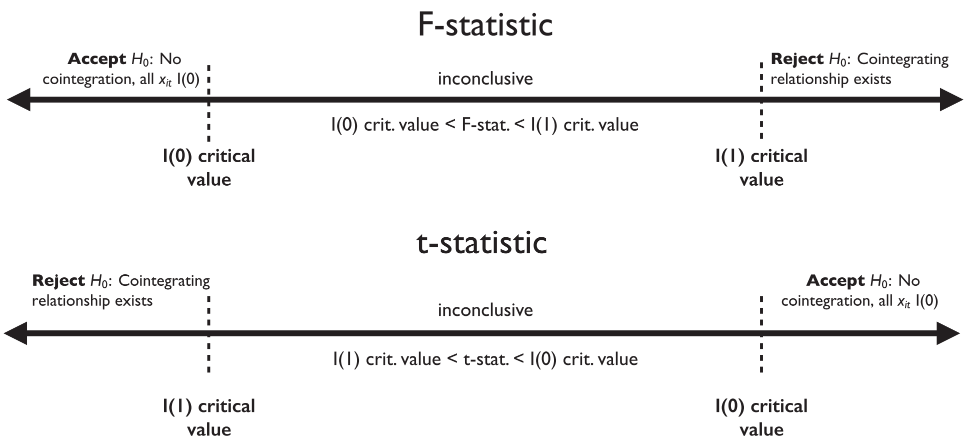
where $\alpha_0^* = (\alpha_0 + \alpha\hat{\kappa}_0)$ and $\theta_0 = -\alpha$. As with the GECM, the coefficient on the lagged value of x_t, $\theta_1 = \alpha\hat{\kappa}_1$, can be combined with the lagged dependent variable to extract the long-run multiplier. The contemporaneous effect is given by β_0. Since residual autocorrelation may be problematic, up to q lags of the first difference of the independent variables, and up to p lags of the first difference of the dependent variable, may be included in order to purge serial autocorrelation from ϵ_t (steps *g* and *k*; Pesaran, Shin, and Smith 2001, 299). Theory and information criteria should be used to specify lag structure, and autocorrelation tests used to ensure white noise residuals.

[10]This excludes the possibility of multi-cointegration (Enders 2010, 380–82).

[11]In the context of cointegration, a variable is weakly exogenous if it "does not respond to the discrepancy from the long-run equilibrium relationship" (Enders 2010, 407).

[12]If a deterministic trend was suspected in y_t, Equation (5) becomes. $y_t = \hat{\kappa}_0 + \hat{\gamma} T + \hat{\kappa}_1 x_t + z_t$. We could also exclude the drift term, $\hat{\kappa}_0$, or account for a deterministic trend in x_t.

FIGURE 2 Bounds Test Statistics



The resulting model appears as.

$$\Delta y_t = \alpha_0^* + \theta_0 y_{t-1} + \theta_1 x_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \sum_{j=0}^{q} \beta_j \Delta x_{t-j} + \epsilon_t. \quad (9)$$

After estimating the ARDL-bounds model in Equation (9) and ensuring white noise residuals (steps *g* and *k*), the next step is to conduct the bounds test (step *n*). It tests the null hypothesis of *no* cointegration between the dependent variable and any regressors included in the cointegrating equation (Pesaran, Shin, and Smith 2001, 294–95). Only regressors that enter into the equation in levels (e.g., x_{t-1}) in Equation (9) can (potentially) cointegrate with y_t. The bounds F-test consists of running a Wald test or F-test on the following restriction from Equation (9):

$$H_0 : \theta_0 = \theta_1 = 0 \quad (10)$$

under the null hypothesis that no cointegrating relationship exists between x_t and y_t. Rejecting H_0 indicates that there is a cointegrating relationship between the series.

In addition to the F-test, a one-sided t-test may be used to test the null hypothesis that the coefficient on the lagged dependent variable is equal to zero: $H_0 : \theta_0 = 0$. The alternative hypothesis is that $\theta_0 < 0$, which suggests cointegration. This is known as the bounds t-test.

The critical value bounds for the F- and t-statistics are nonstandard and depend on the number of regressors appearing in levels, as well as the restrictions placed on the intercept and trend.[13] Asymptotic critical values for the t- and F-statistics can be found in Pesaran, Shin, and Smith (2001, 300–304), and small-sample critical values for the F-statistic can be found in Narayan (2005, 1987–90). No small-sample critical values are currently available for the t-test, so in small samples it should only be used for confirmatory purposes. Interpretation of the bounds test is illustrated in Figure 2. Three possibilities result.

If the value of the F-statistic is *lower* than the stationary critical value, then we cannot reject the null hypothesis that there is no cointegrating relationship (step *q* in Figure 1); in fact, we can conclude that all independent variables appearing in levels are stationary, without having to conduct any further unit root testing. If this is the case, the final model specification is the first difference of the dependent variable regressed on up to *l* lags of the independent variables appearing in levels, as well as up to *p* and *q* lags of the first differences of the dependent and independent variables necessary to remove autocorrelation (step *r*):

$$\Delta y_t = \alpha_0 + \sum_{k=0}^{l} \delta_k x_{t-k} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \sum_{j=0}^{q} \beta_j \Delta x_{t-j} + \epsilon_t. \quad (11)$$

If the value of the F-statistic is *higher* than the I(1) critical value, not only are all series I(1), but there also exists a cointegrating relationship between them. No further unit root testing of the regressors is required, as shown by step *o* in Figure 1. Evidence suggests that the resulting ARDL

[13] Dummy variables may be included without compromising the asymptotic properties of the tests, as long as they tend toward zero as the series increases (Pesaran, Shin, and Smith 2001, 307). The cointegration test does not account for the possibility of seasonal unit roots (Pesaran, Shin, and Smith 2001, 291) or other forms of periodicity, so these should be prewhitened out accordingly.

model in error correction form is correctly specified, and that cointegration exists between the dependent variable and any independent variables appearing in levels.

If the F-statistic is *between* the stationary and I(1) critical values, the test is inconclusive. There could be a mix of stationary and I(1) regressors, and cointegration among the I(1) variables and the dependent variable may still exist. However, further testing is required. As shown by step *m* in Figure 1, the next step is to conduct unit root tests for each independent variable. Since I(0) variables cannot possibly have a cointegrating relationship with an I(1) dependent variable, they should only enter into the model in first-differenced form.[14] After rerunning the ARDL model in error correction form (step *e*), conduct the bounds test for cointegration (step *n*) on the remaining I(1) regressors. If a conclusive result is reached, no further testing is required. If the test is still inconclusive, the next step is to start excluding combinations of I(1) regressors from the cointegrating equation (having a θ coefficient in Equation 9) and repeat steps *e* and n. If, after iterating through the possible combinations of independent variables, there is still no conclusive result from the bounds test, then we can conclude no cointegration. Since short-run effects between I(1) variables may still exist, the final model can be estimated in first differences.

Evaluating the t-statistic is exactly the opposite of the F-statistic; if the value of the t-statistic is lower then the I(1) critical value, then we can reject the null hypothesis of no cointegrating relationship. If the value of the t-statistic falls above the I(0) critical value, then we cannot reject the null hypothesis. Just as with the F-statistic, if the critical value falls between the bounds, the test is inconclusive, and more precise testing of the regressors is necessary. That is to say, we would next use unit root testing to isolate out only the I(1) variables and iterate through them as needed in order to conclude either cointegration (step *o*) or all I(0) regressors (step *q*).

Monte Carlo Evidence

The key component to the ARDL-bounds procedure is the cointegration test, since it ultimately determines our conclusions about the relationships between variables. How does its performance compare to existing approaches? To evaluate this, I present two Monte Carlo experiments. The first focuses on finding evidence of cointegration when it does not exist (Type I error), whereas the second investigates failing to detect cointegration when it exists (Type II error).

To evaluate the ability of the bounds cointegration test to avoid Type I error, I generated an I(1) dependent variable, y_t, for series of length $T = 35, 50, 80$.[15] Next, four independent variables, x_{kt} (where $k = 1, 2, 3, 4$), were generated. These were completely unrelated to y_t, or to one another:

$$y_t = y_{t-1} + \eta_t. \tag{12}$$

$$x_{kt} = \phi_k x_{kt-1} + \nu_{kt}. \tag{13}$$

The stochastic components η_t and ν_{kt} are i.i.d. and independent from each other. As discussed earlier, detection of stationary variables is difficult in short series. To see the consequences of erroneously including an I(0) regressor when all other variables are I(1), I allow the autoregressive process for x_{1t}, ϕ_1, to vary from 0.0 to 1.0 by increments of 0.10. All other independent variables are I(1) (i.e., $\phi_k = 1 \; \forall k \neq 1$). Next, I ran the ARDL-bounds model:

$$\begin{aligned} \Delta y_t = {} & \alpha_0 + \theta_0 y_{t-1} + \theta_1 x_{1t-1} + \cdots + \theta_k x_{kt-1} \\ & + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \sum_{j=0}^{q_1} \beta_{1j} \Delta x_{1t-j} \\ & + \cdots + \sum_{j=0}^{q_k} \beta_{kj} \Delta x_{kt-j} + \epsilon_t. \end{aligned} \tag{14}$$

The number of lagged first differences of y_t and each x_{kt} to include in Equation (14) was determined via SBIC for each of the 500 simulations conducted for all combinations of T, k, and ϕ_{1x}.[16] After estimating Equation (14), an F-test of the null hypothesis that $\theta_0 = \theta_1 = \cdots = \theta_k = 0$ was conducted for each simulation. The resulting statistic was compared against the associated critical values of the bounds test from Narayan (2005, 1988). Since these series were independently generated, evidence of cointegration (an F-statistic greater than the I(1) critical value) is an incorrect rejection of the null hypothesis and thus a form of Type I error.[17]

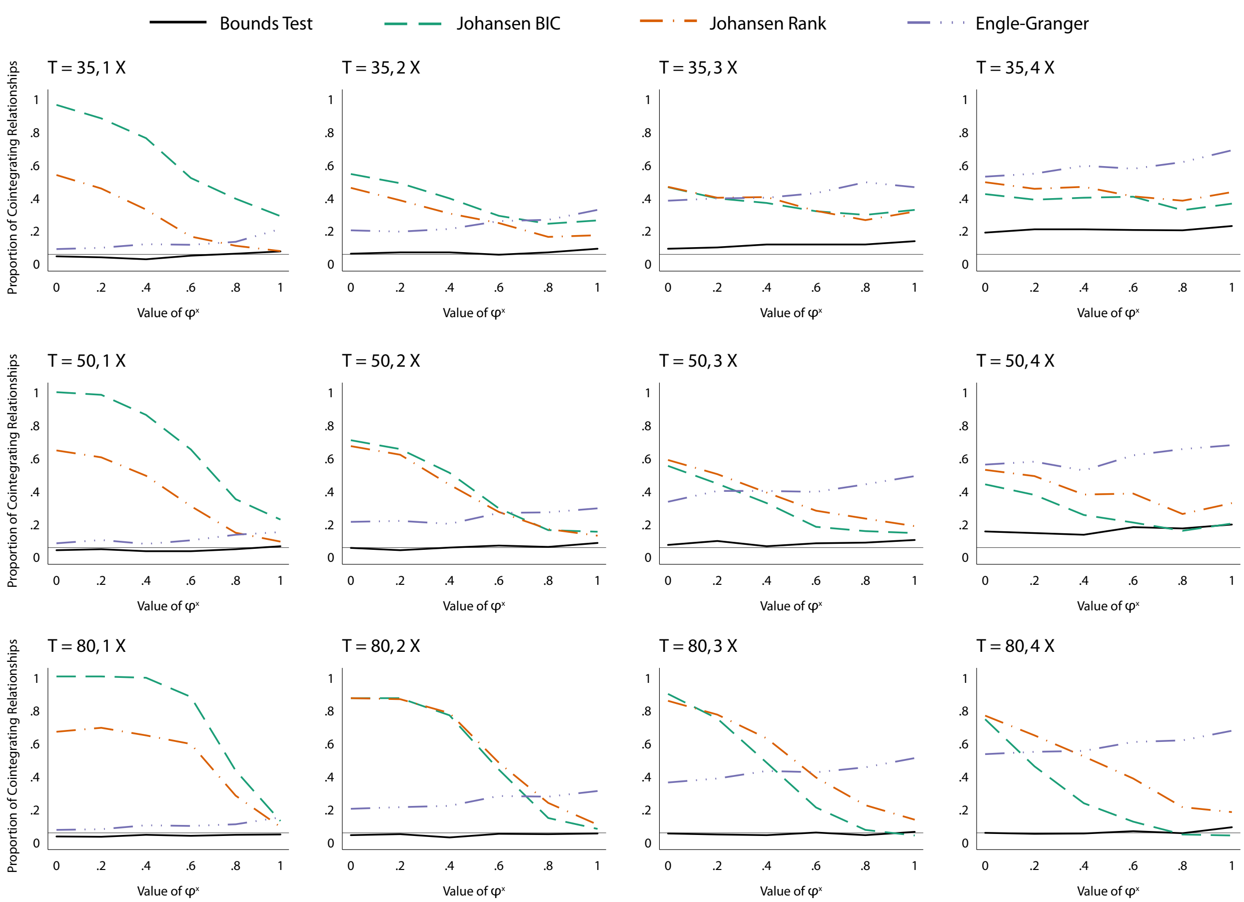
[14]Of course, I(0) series could still appear in levels in the final model specification without risking spurious regression.

[15]To mitigate issues involving initial conditions (Balke and Fomby 1997), I created a burn-in period of $T = 100$ for all simulations.

[16]A restriction of p, $q_k \leq 3$ was placed on the maximum number of lag lengths in Equation (14) for $T = 35$, and 4 for $T = 50, 80$. This restriction appeared to be an ideal trade-off between overfitting and ensuring white noise residuals; I discuss issues regarding overfitting in the supporting information.

[17]F-statistics between the I(0) and I(1) bound, or below the I(0) bound, were treated as avoiding Type I error. Treating them as Type I error does not change the substantive results, as shown in the supporting information.

FIGURE 3 Proportion of Monte Carlo Simulations (Falsely) Detecting Cointegration



Note: Each plot shows the proportion of simulations finding (at $p < .05$) evidence of one cointegrating relationship with up to k regressors and different numbers of observations across varying amounts of autoregression in x_{1t}, using each of the four cointegration testing procedures.

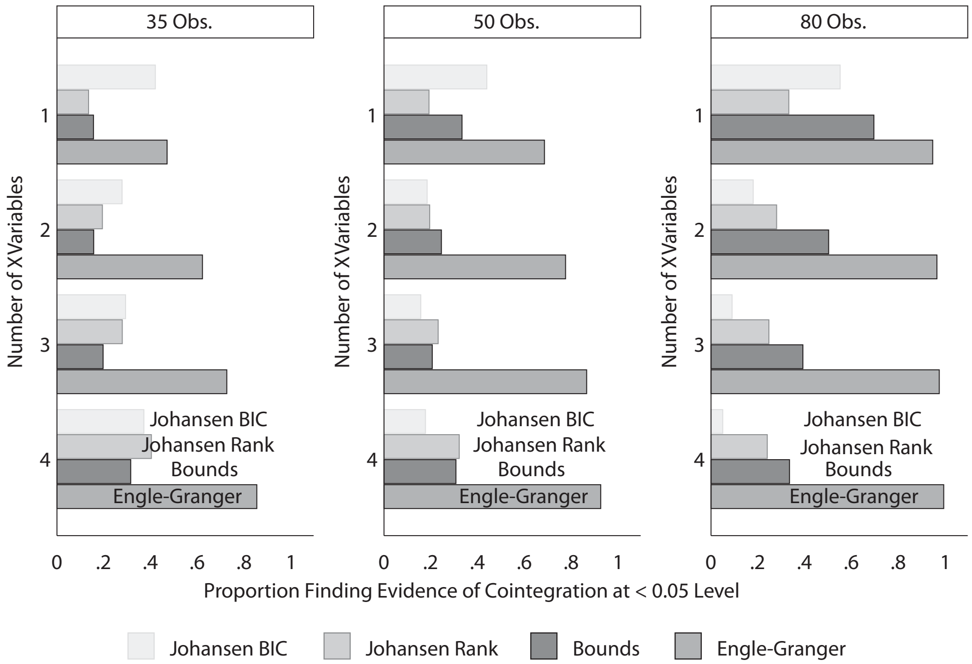
I compare the performance of the bounds test to two other procedures. I included the Engle-Granger two-step procedure by implementing an augmented Dickey-Fuller unit root test on the residual series, z_t, from the cointegrating equation: $y_t = \kappa_0 + \kappa_1 x_{1t} + \cdots + \kappa_k x_{kt} + z_t$.[18] I also used the Johansen procedure for cointegration to test for the existence of a single cointegrating relationship, using both the multiple trace testing procedure as well as the number of cointegrating ranks as chosen by minimizing SBIC (Johansen 1995).[19] Although cointegration tests are only supposed to be run on all-I(1) series, the purpose of this Monte Carlo experiment is to evaluate test performance when a stationary regressor is erroneously included, given that in small series an autoregressive I(0) variable may be indistinguishable from an I(1) series.

The results from the first Monte Carlo experiment are shown in Figure 3. The level of autoregression, ϕ_1, in the single stationary series—x_{1t}—is on the horizontal axis. The proportion of simulations finding evidence of cointegration is on the vertical axis; higher values indicate Type I error. When there are only 35 observations, it is clear that the bounds test is the only cointegration procedure that comes close to the conventional 5% rejection rate (shown by the thin black line). As the number of independent variables increases (each column shows the number of k regressors), all tests tend to have increased Type I error. For instance, when there are four regressors, we find spurious evidence of cointegration about 60% of the time when using the Engle-Granger test; surprisingly, its high rate of Type I error does not change as T increases. This finding underscores recent work on overfitting in short time-series (Helgason 2016; Keele,

[18] The same lag restrictions were placed on the additional augmenting lags of Δy_{t-i} needed to remove autocorrelation, as determined by minimizing SBIC. Critical values are from MacKinnon (1994).

[19] Lag-order selection was the same as the Engle-Granger procedure. Results of $r \neq 1$ were recorded as no evidence of Type I error.

FIGURE 4 Proportion of Monte Carlo Simulations (Correctly) Detecting Cointegration



Linn, and Webb 2016). Despite this, the bounds test excels at successfully failing to reject the null hypothesis of no cointegration under all scenarios. Only the Johansen test appears to have the same low rate of Type I error, but only when the level of autoregression in x_{1t} approaches a unit root process.

The performance of the bounds test is notable in a number of ways. Not surprisingly, I find evidence that it, along with other cointegration tests, performs poorly in small samples. However, this is only when the length of the series is small and the number of regressors large. Even then, the rate of Type I error using the bounds test is often half that of the other cointegration tests, and it remains robust to erroneously including an I(0) regressor. Only the Johansen-BIC test has a similar level of Type I error, but only when all variables are at or near I(1). The fact that the performance of the bounds test is barely affected by autoregression indicates that it is a good test for cointegration in small samples; this is exactly when we might erroneously include an I(0) variable. Finally, while the Engle-Granger procedure is robust to autoregression in a single regressor, it has much larger Type I error as the number of regressors increases. Taken together, this evidence suggests that the bounds cointegration test has lower Type I error than other tests, and it remains robust to short series, multiple regressors, and erroneously including stationary regressors.

I next explore the likelihood that the bounds test fails to detect cointegration when it exists (Type II error). As before, I vary the number of regressors and the number of observations. However, now the independent variables cointegrate with the dependent variable:[20]

$$x_{kt} = x_{kt-1} + \nu_{kt}. \tag{15}$$

$$u_t = 0.75u_{t-1} + \eta_t. \tag{16}$$

$$y_t = 0.25x_{1t} + \cdots + 0.25x_{kt} + u_t. \tag{17}$$

The errors ν_{kt} and η_t are independent. This data generation process yields an adjustment parameter of -0.25 and a long-run multiplier of 0.25 for each of the k independent variables. The cointegration tests are the same as in the previous experiment, and conducted on 1,000 simulations across each combination of observations and regressors.

The results of the second experiment are shown in Figure 4. Each bar depicts the proportion of cointegrating relationships for a particular cointegration test, across each combination of observations and regressors. Higher values correspond with a lower rate of Type II error. For all tests, as the length of the series increases, Type II error decreases. In addition, as the number of cointegrating regressors increases, the Engle-Granger test correctly iden-

[20]A proof of this is in the supporting information.

tifies cointegration at a greater rate than other tests. The bounds test has the largest Type II error rate when $T = 35$, although this improves sharply as the series lengthen. In addition, the proportion of simulations correctly identifying cointegration varies significantly across tests; the Engle-Granger procedure has between one-third and one-half the rate of Type II error as the bounds test, and the bounds test has about one-half the Type II error as the Johansen tests.

A number of important findings stand out from these two experiments on cointegration. The bounds test has the lowest Type I error across *all* scenarios; moderate Type I error (20%) occurs only when there are four regressors and 50 observations or fewer. While the bounds test is largely unaffected, the Johansen test tends to experience a rapid increase in Type I error rates when an I(0) regressor is included. Although the Engle-Granger test has the lowest Type II error rates, the bounds test tends to perform better than the Johansen tests in all scenarios, except for a single regressor or short series.

In the supporting information, I conduct eight additional Monte Carlo experiments. These include varying the adjustment parameter and long-run multiplier, using fractionally (co)integrated series, and examining the percentage of time a given cointegration test correctly or incorrectly diverges from the other three cointegration tests. I also examine the ability of the GECM and ARDL-bounds models to recover substantively interesting effects (e.g., short- and long-run effects, or the adjustment parameter). Many of the findings are consistent with those above; interested readers are directed to the brief summary in Table 1 in the supporting information.

Taken together, the Monte Carlo results suggest that the bounds test offers an ideal compromise between Type I and Type II error. Given calls for more conservative cointegration tests (Grant and Lebo 2016), the bounds test seems the prudent choice since it strongly avoids spurious cointegration, yet can still identify true cointegrating relationships, at least for weakly exogenous regressors.[21] I show two applications of this approach below.

Application I: Kelly and Enns (2010)

Kelly and Enns (2010) examine how income inequality affects public mood liberalism and support for welfare policy. The authors find that in the long-run, increases in inequality are associated with the public becoming more conservative and less supportive of welfare. They find no evidence that policy liberalism, income inequality, unemployment, or inflation has any effect on public mood in the short-run. There are two reasons to believe these results may be suspect. First, the number of observations is small. Second, although Kelly and Enns perform unit root testing on the dependent variable, the authors make no mention of testing the regressors.

I replicated their model of public support for welfare policy.[22] Results from their GECM are shown in Table 1, Model 1. First, I ensured that the dependent variable is I(1) (see step *a* in the Figure 1 schematic). Results from five unit root tests are shown in Table 2. While we can reject the null hypothesis of an I(1) series using the augmented Dickey-Fuller test, more powerful ones such as the Dickey-Fuller Generalized Least Squares (DF-GLS) and Elliott-Rothenberg-Stock (ERS) tests find evidence of a unit root process.[23] Although the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test also provides mixed evidence, we can tentatively confirm that the data-generating process of the dependent variable is I(1).[24]

After ensuring that all regressors are first-order nonstationary or less (step *b* in Figure 1), I then estimated the ARDL model in error correction form (step *e*).[25] Using SBIC, I found that the lag structure in the original model used by Kelly and Enns (2010) was optimal, given the data. This specification produced white noise residuals, as evidenced by a battery of post-estimation diagnostics. Thus, the ARDL-bounds model shown in Model 2 in Table 1 is identical to the original ECM in Model 1.

Since the model appears to be dynamically stable, we next use the bounds test to identify whether a cointegrating relationship exists between support toward welfare policy, policy liberalism, and income inequality (step *n* in Figure 1). An F-test that the parameters on the variables appearing in lagged levels—$Welfare_{t-1}$, $Policy\ Liberalism_{t-1}$, and $Income\ Inequality_{t-1}$—are jointly equal to zero yields an F-statistic of 4.15. Although Narayan (2005) provides the small-sample critical values necessary to evaluate this statistic, these are also available in Stata and R using the programs `pssbounds` and `pss`,

[21]Were the regressors endogenous, methods such as the Johansen approach should be used.

[22]See Table 1, Model 4, on page 864 in their article.

[23]The augmented Dickey-Fuller and Phillips-Perron tests suffer from size distortions and weak power, and they are often outperformed by the ERS and DF-GLS tests (Choi 2015, 37–54; Enders 2010, 234–37; Maddala and Kim 1998, 98–103).

[24]I examine the consequences of concluding stationarity in the supporting information. Although the final model differs, the substantive results remain unchanged.

[25]Unit root tests of the first difference of policy liberalism and inequality rejected the I(2) null hypothesis; results are in the supporting information.

TABLE 1 Results of the ARDL-Bounds Model for Welfare Policy Mood (Kelly and Enns 2010)

	(1) Original GECM	(2) ARDL-Bounds	(3) ARDL-Bounds Excluding Policy Liberalism$_{t-1}$	(4) ARDL-Bounds Excluding Income Inequality$_{t-1}$	(5) Final Model in First Differences
Welfare$_{t-1}$	-0.55^{**}	-0.55^{**}	−0.23	-0.37^{*}	
	(0.16)	(0.16)	(0.12)	(0.14)	
Δ Policy Liberalism$_t$	−0.24	−0.24	−0.33	−0.05	−0.43
	(0.42)	(0.42)	(0.46)	(0.46)	(0.45)
Δ Policy Liberalism$_{t-1}$			−0.76		−0.73
			(0.45)		(0.45)
Policy Liberalism$_{t-1}$	-0.65^{**}	-0.65^{**}		−0.21	
	(0.22)	(0.22)		(0.11)	
ΔIncome Inequality$_t$	−175.70	−175.70	−136.99	−183.79	−155.62
	(122.56)	(122.56)	(132.23)	(134.45)	(133.85)
ΔIncome Inequality$_{t-1}$				−202.02	
				(137.35)	
Income Inequality$_{t-1}$	-152.17^{*}	-152.17^{*}	7.89		
	(65.71)	(65.71)	(33.50)		
Constant	80.03^{*}	80.03^{*}	3.18	13.11^{*}	−0.54
	(29.92)	(29.92)	(12.92)	(4.79)	(1.20)
Observations	33	33	33	33	33
Adjusted R^2	0.26	0.26	0.12	0.17	0.07
Breusch-Godfrey χ^2 of AR(1)	1.41	1.41	0.06	1.10	0.63
AR(2)	1.43	1.43	1.09	1.64	2.05
AR(3)	1.50	1.50	1.37	2.10	2.37
Durbin's Alternative χ^2 of AR(1)	1.16	1.16	0.05	0.90	0.55
AR(2)	1.13	1.13	0.85	1.31	1.79
AR(3)	1.14	1.14	1.04	1.63	2.02
Cumby-Huizinga χ^2 of AR(1)–AR(3)	1.06	1.06	1.48	1.26	1.76
Shapiro-Wilk z	−0.44	−0.44	1.66	0.78	2.70

Note: Dependent variable is welfare policy mood. Model 1 shows results from Kelly and Enns (2010), Model 2 shows results using ARDL-bounds procedure, Models 3 and 4 show the results when testing down using ARDL-bounds, and Model 5 is the final model in first differences. Lag structures are determined by SBIC. Standard errors are in parentheses. $^{*}p < .05$, $^{**}p < .01$.

respectively (Jordan and Philips 2016; Philips 2016b). The critical values for 33 observations and two regressors are a lower stationary bound of 4.183 and an upper I(1) bound of 5.333. Strictly speaking, the F-statistic is below the stationary lower bound, so we might conclude that all regressors are stationary (step q in Figure 1). However, given that the test result was so close to the I(0) lower bound of the test, we may want to treat the result as inconclusive, which means that further testing is needed.[26]

[26]Moreover, the one-sided bounds t-test on the significance of the lagged dependent variable, −3.46, falls between the asymptotic upper I(0) and lower I(1) critical bounds of −2.86 and −3.53, respectively; this supports the "inconclusive" decision.

Although the results of the cointegration test were borderline inconclusive with both policy liberalism and income inequality, a single regressor may still cointegrate with welfare policy mood. The next step is to test that the regressors are I(1), since any I(0) regressor can easily be excluded from the cointegrating equation (step m in Figure 1). Unit root testing (available in the supporting information) indicated that both policy liberalism and income inequality are I(1).

Since unit root testing did not narrow down which series should not appear in the cointegrating equation, I estimated two different models (step n). In Model 3, I test to see whether *only* income inequality has a cointegrating relationship with public mood toward welfare.

TABLE 2 Public Mood Toward Welfare Is I(1) (Kelly and Enns 2010)

Unit Root Test	Welfare
Augmented Dickey-Fuller (with drift)	-2.05^{*}
Phillips-Perron	-1.94
Dickey-Fuller GLS (with trend)	-2.55
Elliott-Rothenberg-Stock	-2.55
Kwiatkowski-Phillips-Schmidt-Shin (H_0 = stationary)	0.49^{*} (no lag), 0.29 (1 lag)
Conclusion	I(1)

Note: Thirty-three observations with 1-year lag are included for all tests unless otherwise noted. H_0 = series contains a unit root for all tests except KPSS $^{*}p < .05$.

Therefore, policy liberalism does not appear in levels in Model 3. In order to produce white noise residuals (steps g and k), the lagged first difference of policy liberalism was included. Because Model 3 reflects a data-generating process where only income inequality is cointegrating, evidence of cointegration in Model 3 would indicate that income inequality, not policy liberalism, is cointegrating with public mood toward welfare. An F-test of the significance of the lagged variables in Model 3 yields an F-statistic of 1.72. Since this is below the critical value of 5.290 for the I(0) lower bound and 6.175 for the I(1) upper bound, we can conclude that income inequality and public mood toward welfare are not cointegrating.

Next, I test to see whether *only* policy liberalism has a cointegrating relationship with public mood toward welfare. Therefore, in Model 4, income inequality does not appear in levels. To produce white noise residuals, one lag of the first difference of income inequality was included. For Model 4, a rejection of the null hypothesis using the bounds test would suggest that policy liberalism, not income inequality, is cointegrating with public mood toward welfare. An F-test of the significance of the lagged variables yields an F-statistic of 3.57. Since this falls below the I(0) critical value of 5.290 (as well as the upper I(1) critical value, 6.175), we can conclude that policy liberalism and public mood toward welfare are not cointegrating.

Since neither income inequality nor policy liberalism on their own appear to have a cointegrating relationship with welfare policy mood—nor do the three variables all together, as found in Model 2—we can conclude that there is no cointegration (step q). Since the two independent variables may still affect public mood toward welfare in the short run, we may run a model of first differences (step r). This is shown in Model 5 in Table 1. The results indicate that income inequality and policy liberalism do not have a statistically significant effect on the public's feelings toward welfare policy in the short run, a similar conclusion to what Kelly and Enns (2010) find.[27]

This replication is informative since it shows how one should proceed, given an inconclusive bounds test result. After finding that all regressors were I(1), I proceeded to iterate through two different models, excluding one of the regressors from the cointegrating equation in Models 3 and 4. Since there was no evidence for cointegration when isolating out income inequality and policy liberalism, the final model was one of first differences since the error correction framework is no longer appropriate.

While suggestive, this replication does not completely overturn the findings of Kelly and Enns (2010). Short series introduce a large amount of uncertainty into cointegration tests, so it seems reasonable that different researchers might come to different conclusions.[28] Overall, given the best available methods, there appear to be null findings in their model of public mood toward welfare.[29]

Application II: Volscho and Kelly (2012)

Volscho and Kelly (2012) use a GECM to probe the determinants of the rise in top income shares in the United States from 1949 to 2008. I examine their power resource model, which investigates whether the share of income of the top 1% is determined by political and institutional factors. Results from their original model are shown in Table 3, Model 1. As Volscho and Kelly find, increases in Democratic strength in Congress, union membership, and the presence of divided government tend to decrease the share of income held by the superrich, but only in the long run. In contrast, Democratic presidents have no effect.

To implement the ARDL-bounds procedure, I first ensured that the dependent variable, *Top 1% Share*, was I(1), as shown in Table 4 (step a in Figure 1). After confirming that the regressors are I(1) or less (step b), I used

[27] What differs is that the authors find evidence of a long-run effect, whereas the ARDL-bounds approach does not.

[28] The Monte Carlo results show that while the bounds test tends to avoid spurious conclusions of cointegration in small samples, it also tends to have a high rate of false negatives; thus, it is hard to ascertain whether their result holds.

[29] However, I find evidence of cointegration using this same approach when examining Kelly and Enns's other dependent variable—public mood liberalism—as detailed in the supporting information.

TABLE 3 Results of the ARDL-Bounds Model (Volscho and Kelly 2012)

	(1) Original GECM	(2) ARDL-Bounds
Top 1% Share$_{t-1}$	-0.36^{**} (0.09)	-0.30^{**} (0.07)
Δ Democratic President$_t$		1.47^{**} (0.53)
Δ Democratic President$_{t-1}$	0.14 (0.56)	
Democratic President$_{t-1}$	−0.20 (0.36)	0.11 (0.34)
Δ % Congressional Democrat$_t$		−0.03 (0.04)
Δ % Congressional Democrat$_{t-1}$	0.05 (0.04)	
% Congressional Democrat$_{t-1}$	-0.12^{**} (0.04)	-0.12^{**} (0.03)
Δ Divided Government$_t$		0.37 (0.46)
Δ Divided Government$_{t-1}$	−0.11 (0.50)	
Divided Government$_{t-1}$	-0.93^{*} (0.42)	-0.83^{*} (0.37)
Δ Union Membership$_t$	0.29 (0.28)	0.04 (0.28)
Union Membership$_{t-1}$	-0.11^{**} (0.03)	-0.09^{**} (0.02)
Constant	15.05^{**} (3.83)	13.30^{**} (2.81)
Observations	60	61
Adjusted R^2	0.20	0.29
Breusch-Godfrey χ^2 of AR(1)	1.39	3.19
AR(2)	1.39	3.21
AR(3)	2.79	5.03
Durbin's Alternative χ^2 of AR(1)	1.16	2.76
AR(2)	1.14	2.72
AR(3)	2.29	4.31
Cumby-Huizinga χ^2 of AR(1)–AR(3)	4.41	5.09
Shapiro-Wilk z	0.17	0.99

Note: Dependent variable is the share of income of the top 1%. Model 1 shows results from Volscho and Kelly (2012), and Model 2 shows results using ARDL-bounds procedure, with lag structure determined by minimizing SBIC. Standard errors are in parentheses. $^{*}p < .05$, $^{**}p < .01$.

TABLE 4 Top 1% Share Is I(1) (Volscho and Kelly 2012)

Unit Root Test	Top 1% Share
Augmented Dickey-Fuller (with drift)	0.02
Phillips-Perron	−0.21
Dickey-Fuller GLS (with trend)	−1.35
Elliott-Rothenberg-Stock	−1.35
Kwiatkowski-Phillips-Schmidt-Shin (H_0 = stationary)	2.20^{*}
Conclusion	I(1)

Note: $T = 60$ with 1-year lag included for all tests. H_0 = series contains a unit root for all tests except KPSS. $^{*}p < .05$.

SBIC to assist in lag selection for the ARDL model in error correction form, the result of which is shown in Model 2 (step *e*). Although the authors may have had theoretical reasons to use the "dead-start" GECM, I find instead that a model of contemporaneous short-run effects has a lower SBIC. While theory should always guide model specification, users *must* ensure that the residuals are white noise in order to run the bounds test; in this example, both the dead-start and standard GECM yielded white noise residuals.[30]

Since Model 2 contains white noise residuals, we can move onto cointegration testing using the bounds test (step *n* in Figure 1). An F-test of the joint significance of the five lagged variables (the four regressors plus the dependent variable) yields an F-statistic of 5.02. Critical values for 61 observations and four regressors are 3.068 and 4.274 for the lower and upper bounds, respectively. Since the F-statistic is greater than the I(1) upper bound, we can conclude that there is a cointegrating relationship (step *o*). As further confirmation, we can use the bounds t-test; the t-statistic on the lagged dependent variable is −4.01, which is below the critical value of the I(1) lower

[30]Therefore, one could use the bounds test on either model.

bound (−3.99). Thus, there is strong evidence that all four regressors are cointegrating with the dependent variable.

The largest difference between Volscho and Kelly's (2012) original model and the ARDL-bounds model is the significance of the short-run effect of a Democratic president. To see whether this leads to different conclusions than the ones made by the authors, in the supporting information I use dynamic simulations to help interpret how changes in one regressor affect the dependent variable over time. Model-based dynamic simulations are growing in popularity in political science (King, Tomz, and Wittenberg 2000; Williams and Whitten 2012), and they are especially valuable for examining complex model specifications such as autoregressive relationships with interactions (Williams and Whitten 2011) or dynamic compositional dependent variables (Philips, Rutherford, and Whitten 2015, 2016). The ARDL-bounds procedure's lag structure makes it a prime candidate for dynamic simulations. Using the program `dynpss` to create dynamic simulations of the ARDL-bounds model (Philips 2016a), I find that in the short run, moving from a Republican to a Democratic president increases the income concentration of the top 1%. However, this effect loses statistical significance after 4 years, it is not statistically significantly different from the predictions using Volscho and Kelly's (2012) GECM, and the long-run effect is nearly zero.[31] These results are available in the supporting information.

In summary, I find evidence for cointegration in the power resources model of Volscho and Kelly (2012). While the ARDL-bounds model had slight specification differences, the substantive findings do not change, as evidenced by dynamic simulations. Institutional and political factors may affect the income share of the top 1%, but only in the long-run.

Discussion and Conclusion

The two examples above represent a variety of situations that the ARDL-bounds approach is designed to handle. For the Kelly and Enns (2010) replication, I find no evidence of cointegration. Using the steps outlined in Figure 1, I find no evidence that policy liberalism and income inequality affect welfare policy mood in the long- or short-run. For the Volscho and Kelly (2012) replication, I find evidence of cointegration; these findings support the authors' conclusions about the long-run effect of institutions and politics on the concentration of income of the top 1%. In the supporting information, I also replicate Ura (2014) and find evidence of cointegration.

Although the examples above are representative of most situations practitioners are likely to encounter, I briefly review how users should proceed, given their own theoretically specified model:

1. *Unit root testing of the dependent variable.* If the dependent variable is I(1), proceed with the ARDL in error correction form.[32]
2. *Ensure that no independent variables are of an order of integration higher than I(1).* The main advantage of the bounds approach is that users do not have to make difficult decisions between I(0) and I(1) regressors; the results of the bounds test inform us of these characteristics. However, users must ensure that no variables are integrated more than I(1), are explosive, or contain seasonal unit roots.[33]
3. *Estimate the ARDL in error correction form.* Since the bounds testing procedure relies on white noise residuals, add lags of the first differences of the dependent variable and regressors as needed. Use theory and information criteria to aid in lag specification. Ensure that the residuals are white noise.
4. *Test the joint significance of all lagged variables appearing in levels using a Wald/F-test.* Use small-sample critical values of the bounds test in Narayan (2005). As an auxiliary test, use the one-sided t-test of the lagged dependent variable using asymptotic critical values in Pesaran, Shin, and Smith (2001).
5. If the results of the bounds test.
   (a) *Suggest cointegration*: All variables appearing in levels appear to be I(1) and have a cointegrating relationship with the dependent variable.
   (b) *Suggest stationarity*: All regressors appearing in levels are I(0) and cannot possibly be in a cointegrating relationship. A model of first differences must be estimated since the variables may still affect the dependent variable in the short run.
   (c) *Are inconclusive*: Each regressor should be tested for a unit root. Only I(1) variables can

[31] This is confirmed analytically by calculating the long-run multiplier, which is 0.36 and is not statistically significantly different from zero.

[32] If the dependent variable is I(0), it is not first differenced, leading to a lagged dependent variable model as shown in the Figure 1 schematic.

[33] While the test statistics can be adjusted to account for deterministic trends in the dependent variable, it is advisable to identify and detrend instead.

appear in levels in the error correction model. Stationary variables may still appear in first differences.[34] Repeat Steps 3 and 4. If the resulting statistic is still inconclusive, combinations of variables appearing in levels may need to be tested. Continue testing until (5a) or (5b) is reached.

6. *Interpretation.* Use dynamic simulations and analytical calculations for hypothesis testing.

While the ARDL-bounds procedure provides a comprehensive approach to modeling time-series and testing for cointegration, it is not a remedy for all problems. First, like all time-series models, it tends to perform poorly in small samples. As a precaution against overfitting, Keele, Linn, and Webb (2016, 40) suggest a minimum of between 10 and 20 observations per parameter.[35] However, as shown by Monte Carlo simulations, the bounds cointegration test tends to perform at least as well as other cointegration tests in small samples. Second, this single-equation model imposes a causal ordering and assumes weak exogeneity of the regressors (Pesaran, Shin, and Smith 2001, 293), a disadvantage shared with GECMs. Users unwilling to impose a causal ordering should consider alternative methods such as vector error correction models, which can account for multiple cointegrating relationships. Third, the cointegration test serves as a substitute for unit root testing to distinguish between I(0) and I(1) regressors *only* when the test results fall outside of the critical bounds. Given an inconclusive test result, users must use unit root tests on all regressors and identify the stationary, I(1), and I(1)-and-cointegrating variables through an iterative process, as shown in the Kelly and Enns (2010) replication. Last, this procedure still requires balanced equations (Grant and Lebo 2016; Keele, Linn, and Webb 2016); although stationary regressors can appear in levels in the final model, I(1) regressors that are not cointegrating cannot appear in levels in the final model without risk of spurious regression.

To aid in the use of this approach, this article has provided a step-by-step guide for practitioners that can be used with any software package that contains unit root, autocorrelation, and the F- and t-tests necessary for the bounds test (e.g., R, Stata, or EViews). In addition, in the supporting information, I discuss software programs in Stata and R designed to help users test for cointegration and create dynamic simulations.[36]

This article was motivated by a series of recent articles in the time-series literature that stress the importance of careful unit root and cointegration testing. To achieve this, I have advocated for the autoregressive distributed lag bounds approach. I have shown that the ARDL-bounds procedure starts with a theoretically specified model and moves step-by-step to arrive at an informed conclusion. Through careful testing and model specification, the ARDL-bounds procedure is a powerful approach to a difficult problem in applied time-series analysis.

References

Balke, Nathan S., and Thomas B. Fomby. 1997. "Threshold Cointegration." *International Economic Review* 38(3): 627–45.

Box-Steffensmeier, Janet M., and Renee M. Smith. 1998. "Investigating Political Dynamics Using Fractional Integration Methods." *American Journal of Political Science* 42(2): 661–89.

Choi, In. 2015. *Almost all about unit roots: Foundations, developments, and applications.* Cambridge: Cambridge University Press.

De Boef, Suzanna, and Jim Granato. 1997. "Near-Integrated Data and the Analysis of Political Relationships." *American Journal of Political Science* 41(2): 619–40.

De Boef, Suzanna, and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52(1): 184–200.

Dickinson, Matthew J., and Matthew J. Lebo. 2007. "Reexamining the Growth of the Institutional Presidency, 1940–2000." *Journal of Politics* 69(1): 206–19.

Enders, Walter. 2010. *Applied econometric time series.* 3rd ed. New York: John Wiley and Sons.

Engle, Robert F., and Clive W. J. Granger. 1987. "Co-integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55(2): 251–76.

Esarey, Justin. 2016. "Fractionally Integrated Data and the Autodistributed Lag Model: Results from a Simulation Study." *Political Analysis* 24(1): 42–49.

Grant, Taylor, and Matthew J. Lebo. 2016. "Error Correction Methods with Political Time Series." *Political Analysis* 24(1): 3–30.

Helgason, Agnar Freyr. 2016. "Fractional Integration Methods and Short Time Series: Evidence from a Simulation Study." *Political Analysis* 24(1): 59–68.

[34] I(0) variables could appear in levels in the final model without risking spurious regression.

[35] I address concerns about overfitting in the supporting information.

[36] In Stata, these are `pssbounds` for displaying critical values of the bounds test and `dynpss` for creating dynamic simulations of the ARDL-bounds model (Philips 2016a, 2016b). The `pss` package implements these commands in R (Jordan and Philips 2016).

Johansen, Soren. 1995. *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford: Oxford University Press.

Jordan, Soren, and Andrew Q. Philips. 2016. "pss: R Package to Perform the Bounds Test for Cointegration and Create Dynamic Simulations." https://github.com/andyphilips/pss. R package version 1.3.9.

Keele, Luke, Suzanna Linn, and Clayton M. Webb. 2016. "Treating Time with All Due Seriousness." *Political Analysis* 24(1): 31–41.

Kelly, Nathan J., and Peter K. Enns. 2010. "Inequality and the Dynamics of Public Opinion: The Self-Reinforcing Link between Economic Inequality and Mass Preferences." *American Journal of Political Science* 54(4): 855–70.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44: 347–61.

Lebo, Matthew J., Robert W. Walker, and Harold D. Clarke. 2000. "You Must Remember This: Dealing with Long Memory in Political Analyses." *Electoral Studies* 19(1): 31–48.

MacKinnon, James G. 1994. "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests." *Journal of Business and Economic Statistics* 12(2): 167–76.

Maddala, Gangadharrao S., and In-Moo Kim. 1998. *Unit roots, cointegration, and structural change.* Cambridge: Cambridge University Press.

Narayan, Paresh Kumar. 2005. "The Saving and Investment Nexus for China: Evidence from Cointegration Tests." *Applied Economics* 37(17): 1979–90.

Pesaran, M. Hashem, Yongcheol Shin, and Richard J. Smith. 2001. "Bounds Testing Approaches to the Analysis of Level Relationships." *Journal of Applied Econometrics* 16(3): 289–326.

Philips, Andrew Q. 2016a. "dynpss: Stata Module to Dynamically Simulate Autoregressive Distributed Lag (ARDL) Models." https://andyphilips.github.io/dynpss/.

Philips, Andrew Q. 2016b. "pssbounds: Stata Module to Conduct the Pesaran, Shin, and Smith (2001) Bounds Test for Cointegration." http://andyphilips.github.io/pssbounds/.

Philips, Andrew Q., Amanda Rutherford, and Guy D. Whitten. 2015. "The Dynamic Battle for Pieces Of Pie—Modeling Party Support in Multi-Party Nations." *Electoral Studies* 39: 264–74.

Philips, Andrew Q., Amanda Rutherford, and Guy D. Whitten. 2016. "Dynamic Pie: A Strategy for Modeling Trade-Offs in Compositional Variables over Time." *American Journal of Political Science* 60(1): 268–83.

Pickup, Mark. 2009. "Testing for Fractional Integration in Public Opinion in the Presence of Structural Breaks: A Comment on Lebo and Young." *Journal of Elections, Public Opinion and Parties* 19(1): 105–16.

Ura, Joseph Daniel. 2014. "Backlash and Legitimation: Macro Political Responses to Supreme Court Decisions." *American Journal of Political Science* 58(1): 110–26.

Volscho, Thomas W., and Nathan J. Kelly. 2012. "The Rise of the Super-Rich: Power Resources, Taxes, Financial Markets, and the Dynamics of the Top 1 Percent, 1949 to 2008." *American Sociological Review* 77(5): 679–99.

Williams, Laron K., and Guy D. Whitten. 2011. "Dynamic Simulations of Autoregressive Relationships." *Stata Journal* 11(4): 577–88.

Williams, Laron K., and Guy D. Whitten. 2012. "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models." *Journal of Politics* 74(3): 685–93.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

1. Programs to Assist in Implementing the Pesaran, Shin and Smith (2001) ARDL Procedure
2. Summary of Monte Carlo Results
3. Additional Monte Carlo Results
4. Proof of the Equivalence of the Triangular Error-Correction Representation to the Standard Representation
5. Three Replications