Taking Variance Seriously: Visualizing the Statistical and Substantive Significance of ARCH-GARCH Models

Accepted for publication in The Journal of Politics.

Taking Variance Seriously: Visualizing the Statistical and Substantive Significance of ARCH-GARCH Models*

Allyson L. Benton
Soren Jordan
Andrew Q. Philips

Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized ARCH (GARCH) models allow users to estimate the conditional mean and conditional error variance of a time series process. While simulation methods exist to disaggregate the short- and long-run effects of covariate shocks to the conditional mean, scholars' inferences about the conditional error variance are currently limited to tabular interpretation. We propose a novel method of interpretation that moves beyond these tabular inferences. First, we show how changes in ARCH-GARCH processes are conditional on starting values, other covariates, and dynamics, which has led to incomplete or even incorrect inferences. We then develop three bootstrapping techniques to simulate conditional error variance model results and showcase the usefulness of each through replication of prominent studies. Our techniques demonstrate the crucial role of simulation and prediction for drawing statistical and substantive inferences about the volatility of dynamic time series processes.

Keywords: ARCH; GARCH; dynamic simulations; volatility
Short title: "Taking Variance Seriously"

---

*Allyson L. Benton is a Reader, Department of Government, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, allyson.benton@essex.ac.uk. Soren Jordan is an Associate Professor, Department of Political Science, Auburn University, Auburn, AL 36849, sorenjordanpols@gmail.com. Andrew Q. Philips is an Associate Professor and the Henry W. Ehrmann Professor in Law and Jurisprudence, Department of Political Science, University of Colorado Boulder, UCB 333, Boulder, CO 80309-0333, andrew.philips@colorado.edu. Authors listed alphabetically.

Replication files are available in the *JOP* Data Archive on Dataverse (https://dataverse.harvard.edu/dataverse/jop). The empirical analysis has been successfully replicated by the JOP replication analyst.

Supplementary material for this article is available in the appendix in the online edition.

Volatility is everywhere. While the bulk of time series work in political science has focused on modeling the average value of a series, many concepts central to the discipline exhibit considerable and explainable variance in their average value. Political economy data—such as economic indicators (i.e., inflation rates, oil prices, trade flows) and financial market data (i.e., stock market returns, sovereign bond premiums, and currency prices)—are well-known for this (e.g., Bernhard and Leblang 2006*b*; Füss and Bechtel 2008; Schneider and Troeger 2006). A myriad of political indicators—such as partisan identification, party polarization, government approval, vote intention, and electoral support (e.g., Bechtel 2012; Box-Steffensmeier and Smith 1996; Boef 2000; Hellwig 2007) and governmental budgets, revenues and spending, fiscal policy, and financing needs and costs (e.g., Brooks, Cunha and Mosley 2022; Wibbels 2000)—also exhibit considerable variance over time. Politicians care deeply about the stability of key linkages to the public, especially sudden shifts in approval or party support (c.f., Kriner 2006). Theories of budgeting, such as punctuated equilibrium theory, recognize that the variance around an average expenditure is of equal or greater political interest to policymakers than the average value itself, which is assumed to be in equilibrium and relatively stable (Flink 2017). Research investigating volatility has thus made fundamental contributions to the discipline, especially in the fields of comparative politics and political economy. Consider how expectations about the relationship between government ideology and election processes matter to portfolio investors and financial market behavior. This area of study has generated thousands of citations, many of which revolve around research on volatility by Bernhard and Leblang (2006*b*).

Standard tools exist to model such time-varying phenomena. Engle's (1982) Autoregressive Conditional Heteroskedasticity (ARCH) model and Bollerslev's (1986) Generalized ARCH (GARCH) extension allow examination of "conditionally heteroskedastic" time series processes. Such data are characterized by alternating short-run periods of high and low volatility (Enders 2015). ARCH-GARCH models consist of two equations: a conventional time series equation estimating the conditional mean of a time series process and another estimating its conditional variance. This allows examination of whether, how much, and how long covariates affect the conditional mean as well as the volatility around it. ARCH-GARCH models have been a powerful addition to the time series toolkit, prompting extensions to accommodate different conditional error variance structures (c.f., Enders 2015).

Yet despite the availability of these tools, and even with a variety of phenomena to explore, political scientists rarely deploy ARCH-GARCH models. Among articles published in the top 125 political science journals (2000–2020)—ranked by the *Journal Citation Reports, Social Sciences Edition, Clarivate Analytics* (2018)—we found only 31 that used this approach. We believe the disconnect between the potential for the study of volatility and the actual utilization of ARCH-GARCH tools stems from a dearth of practical advice, tailored to political scientists, on the estimation and interpretation of ARCH-GARCH models. Most econometric texts focus on the uses of ARCH-GARCH models for economic forecasting rather than hypothesis testing, with most attention given to identifying optimal lag and error variance structures (e.g., Enders 2015; Engle 2001). There is little guidance on how to draw statistical and substantive inferences from ARCH-GARCH conditional error variance equation results, especially for any covariates essential for hypothesis testing. This has left ARCH-GARCH users—especially in political science—to draw inferences about the impact of covariates on the conditional error variance from tables of parameter estimates that capture only short-run effects. We argue this leads to incomplete or even incorrect inferences.

In this study, we seek to raise interest in ARCH-GARCH models in political science by facilitating their interpretation. We do two things. First, we review the ARCH-GARCH approach, emphasizing differences between how these models are described in econometric texts and how they are estimated statistically. In the process, we show why tabular ARCH-GARCH conditional error variance equation results are insufficient for statistical and substantive inference. Second, we argue that ARCH-GARCH conditional error variance results should be interpreted using graphical visualizations. We present three techniques for this purpose. The first formalizes existing parametric bootstrapping methods—currently used for interpreting conditional mean equations (e.g., Williams and Whitten 2011, 2012; Philips, Rutherford and Whitten 2016; Gandrud, Williams and Whitten 2016; Jordan and Philips 2018*a*,*b*, 2020)—to ARCH-GARCH conditional error variance equations, an approach first utilized by Benton and Philips (2020). The second and third introduce two novel bootstrap-based techniques—one residual based and the other maximum entropy based—for simulating expected values of the conditional error variance and associated measures of uncertainty.

To demonstrate the usefulness of our techniques, we conduct replications of two prominent studies. This exercise reveals the limitations of inferences drawn from tabular ARCH-GARCH conditional error variance equation results. It also demonstrates how stochastic simulation uncovers additional information about the expected conditional error variance, how it differs across substantively interesting scenarios, how exogenous shocks to these scenarios change expected volatility, and by how much and for how long this volatility would be expected to persist. Our aim is to show how time-series scholars can gain better substantive insights into the volatility of their series.

ARCH-GARCH Models in Theory and Practice

Engle's (1982) ARCH model and Bollerslev's (1986) GARCH extension consist of the simultaneous estimation of two equations: a mean equation describing the evolution of the conditional mean of y_t and a variance equation describing the evolution of the conditional variance of y_t, which is equivalent to the conditional error variance. Despite this straightforward intuition, ARCH-GARCH models are less straightforward in their implementation. This complicates model interpretation for reasons we describe below.

For Engle (1982) and Bollerslev (1986), the conditional mean equation is usually expressed as a simple (conventional) time series model, such as (but not limited to) a first-order autoregressive (i.e., AR(1)) process (Enders 2015):

$$y_t = \beta_0 + \phi y_{t-1} + \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \tag{1}$$

where the stationary dependent variable y observed at time t is modeled by a constant β_0, its lag y_{t-1}, a vector of exogenous covariates $\mathbf{x}_t$, and an error term ε_t. The conditional mean of y_t refers to the mean value of y_t conditional on all current and past information, that is, on all past values of y_t $(y_{t-1}, y_{t-2}, ...)$, current and past values of ε_t $(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, ...)$, and current and past values of $\mathbf{x}_t\boldsymbol{\beta}$ $(\mathbf{x}_t\boldsymbol{\beta}, \mathbf{x}_{t-1}\boldsymbol{\beta}, \mathbf{x}_{t-2}\boldsymbol{\beta}, ...)$. The conditional variance of y_t is $Var[y_t \mid \mathcal{M}_{t-1}, \mathbf{x}_t] = E[(y_t - \beta_0 - \phi y_{t-1} - \mathbf{x}_t\boldsymbol{\beta})^2] = E[\varepsilon_t]^2$ (Enders 2015, p.124)[1] and equal to the conditional error variance σ_t^2, since $Var[\varepsilon_t] = E[(\varepsilon_t - E[\varepsilon_t])^2] = E[(\varepsilon_t - 0)^2] = E[\varepsilon_t]^2$. When the conditional error variance is constant at σ^2 (with no subscript t), the conditional variance of y_t will be constant at ε^2. In this case, the error term is mean-zero and its variance is independent of past history (therefore unconditional and constant (homoskedastic) such that $\varepsilon_t \sim i.i.d.\ N(0, \sigma^2)$), and Equation 1 can be estimated using Ordinary Least Squares (OLS) regression (Enders 2015).[2]

However, when the error term ε_t is mean-zero and its variance σ_t^2 is dependent on past history (as indicated by the subscript t), and therefore conditional and non-constant (heteroskedastic) such that $\varepsilon_t \sim i.i.d.\ N(0, \sigma_t^2)$, OLS estimation of Equation 1 may produce biased standard errors (Enders 2015, p.124). In this case, Engle (1982) and Bollerslev (1986) recommend modeling the conditional mean of y_t (Equation 1) alongside the conditional variance of y_t, which is equivalent to the conditional error variance σ_t^2. The simplest approach is to model the conditional error variance σ_t^2 as a first-order autoregressive AR(1) process, using squares of the estimated residuals, such that $\hat{\varepsilon}_t^2 = \omega_0 + \omega_1\hat{\varepsilon}_{t-1}^2 + \nu_t$, where ν_t is a white-noise error term (Enders 2015). The expected conditional error variance at t is $E[\varepsilon_t^2 \mid \varepsilon_{t-1}] = \omega_0 + \omega_1\varepsilon_{t-1}^2$ and is rewritten as:

$$\sigma_t^2 = \omega_0 + \omega_1\varepsilon_{t-1}^2 \tag{2}$$

where the error variance σ^2 at time t is a function of a constant ω_0 and the previous error squared ε_{t-1}^2. The ω_1 parameter represents the ARCH(1) term and captures how

[1] This is distinct from the unconditional variance of y_t, the long-run variance of y regardless of current or past values of y_t and x_t, $Var[y_t] = \frac{\sigma^2}{1-\phi^2}$, where σ^2 is the error-term variance. See Enders (2015, p.124).

[2] This *i.i.d.* assumption also assumes no autocorrelation in ε_t.

unanticipated shocks in the prior period $t-1$ affect the subsequent error variance σ_t^2.

Because the conditional variance of y_t, and thus the conditional error variance σ_t^2, cannot be negative, ω_0 is restricted to $\omega_0 > 0$, while ω_1 is restricted to $0 \leq \omega_1 < 1$ (Enders 2015, p.126).[3] If $\omega_1 = 0$, no ARCH effects are present and $\sigma_t^2 = \omega_0$ (a constant). In the presence of ARCH effects, the closer the ARCH term is to one, the greater the impact of past unanticipated shocks on the conditional error variance. The closer the ARCH term is to zero, the smaller the impact of past unanticipated shocks on the conditional error variance.[4] Additional lags capturing previous unanticipated shocks can be added to create an ARCH process up to order q:

$$\sigma_t^2 = \omega_0 + \omega_1 \varepsilon_{t-1}^2 + \omega_2 \varepsilon_{t-2}^2 + \ldots + \omega_q \varepsilon_{t-q}^2 \tag{3}$$

where $\omega_0 > 0$, $0 \leq \omega_i < 1\ \forall i$, and $\sum_{i=1}^{q} \omega_i < 1$ (to ensure stationarity) (Engle 1982, p.993).

The difficulty in identifying the optimal lag length of the ARCH(q) model led Bollerslev (1986) to generalize it through the addition of the lagged error variance σ_{t-1}^2 (Enders 2015). Today, Bollerslev's (1986) parsimonious generalized ARCH—or GARCH—model is the most widely used, with its most common specification the GARCH(1,1) model (with one ARCH term and one GARCH term):

$$\sigma_t^2 = \omega_0 + \omega_1 \varepsilon_{t-1}^2 + \alpha \sigma_{t-1}^2 \tag{4}$$

where the conditional error variance σ_t^2 is modelled as a function of a constant ω_0, the previous error term squared ε_{t-1}^2, and the lagged error variance σ_{t-1}^2. Continuous recursive substitution reveals that the GARCH(1,1) model is equivalent to an infinite-order ARCH(q) model: $\sigma_t^2 = \hat{\omega}_0 + \sum_{i=0}^{\infty} \hat{\omega}_i \varepsilon_{t-i}^2$. As above, $\omega_0 > 0$ and $0 \leq \omega_1 < 1$, while $0 \leq \alpha < 1$ and $0 \leq (\omega_1 + \alpha) < 1$ to ensure stationarity (Bollerslev 1986).

The advantage of the GARCH approach lies in its fewer terms and coefficient re-

[3]This is not to say volatility only increases, as the left-hand side of Equation 2 is also strictly positive. If the combination of previous-period error ε_{t-1}^2 and its persistence ω_1 are less than the previous-period volatility σ_{t-1}^2, volatility will fall.

[4]And, $\omega_1 < 1$ to ensure stationarity (Engle 1982, p.993).

strictions (Enders 2015, p.129). As above, the ARCH term ω_1 captures how previous unanticipated shocks affect subsequent error variance. The GARCH term α captures the level of persistence in volatility across t. The greater the GARCH term, the more persistent any change in the conditional error variance. The lower the GARCH term, the less persistent any change in the conditional error variance. Similar to the ARCH(1) model above, the GARCH(1,1) model can be expanded to include GARCH(p,q) lags:

$$\sigma_t^2 = \omega_0 + \sum_{i=1}^{q} \omega_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \alpha_j \sigma_{t-j}^2 \tag{5}$$

where $\omega_0 > 0$, $0 \leq \omega_i < 1$ $\forall i$, $0 \leq \alpha_j < 1$ $\forall j$, $\sum_{i=1}^{q} \omega_i < 1$, $\sum_{j=1}^{p} \alpha_j < 1$, and $(\sum_{i=1}^{q} \omega_i + \sum_{j=1}^{p} \alpha_j) < 1$. These restrictions ensure σ_t^2 is positive and remains stationary (Bollerslev 1986). We focus on the GARCH(1,1) below, as it prevails in political science.

Adding (Conditional) Covariate Shocks

One crucial feature of ARCH-GARCH models for political scientists is that additional covariates can be added to the conditional mean and conditional error variance equations. The conditional mean equation with covariates $\mathbf{x}_t$ is shown in Equation 1 and operates as in any conventional time series model. Because y_t can take negative or positive values, the vector of coefficients $\boldsymbol{\beta}$ can include positive and negative estimates.

The augmented GARCH(1,1) conditional error variance equation including covariates $\mathbf{z}_t$ is usually expressed in econometric texts as:

$$\sigma_t^2 = \omega_0 + \omega_1 \varepsilon_{t-1}^2 + \alpha \sigma_{t-1}^2 + \mathbf{z}_t \boldsymbol{\gamma} \tag{6}$$

where $\mathbf{z}_t$ is a vector of covariates which may or may not be the same as $\mathbf{x}_t$.[5] However, while some covariates might raise the conditional error variance σ_t^2, others might lower it, such that $\boldsymbol{\gamma}$ might include positive or negative values. This means that covariates in the conditional mean equation cannot be modelled as typically expressed in economet-

[5]Covariates could also include interactions with time, if a time-varying effect were suspected, or period dummies (c.f., Hellwig 2007; Benton and Philips 2020).

ric texts and must be transformed to ensure that they can take negative values, since the conditional error variance σ_t^2 must remain strictly positive. The simplest way to allow negative values in $\boldsymbol{\gamma}$ is by including both ω_0 and any covariate shocks $\mathbf{z}_t$ through an exponential link function, with this *applied* GARCH(1,1) model specified as:[6]

$$\sigma_t^2 = \omega_1 \varepsilon_{t-1}^2 + \alpha \sigma_{t-1}^2 + \exp(\omega_0 + \mathbf{z}_t \boldsymbol{\gamma}) \quad (7)$$

Most statistical packages automatically apply this link function (e.g., Stata) or recommend it (e.g., EViews). However, because most econometric texts do not discuss this transformation, we think scholars may have sometimes misinterpreted the tabular coefficient results. Even for those aware of this transformation, it is difficult to draw conclusions about the impact of covariate shocks on the conditional error variance from tabular results alone. Thanks to the exponential link function, their statistical and substantive impact depends on their starting values and those of other covariates, as well as on the GARCH α parameter defining their temporal persistence.

A synthetic example illustrates the conditionality of covariate effects on the conditional error variance and the difficulty of drawing inferences from tabular results. We generated data for a GARCH(1,1) process following Equation 7 with two standard normal variables X and Z, varying the GARCH parameter α from 0.2 (small persistence in the error variance) to 0.8 (large persistence in the error variance).[7] We are interested in the response of the series to a one standard deviation ("SD") "shock" to the X series at different baseline levels of X and at different levels of the other covariate Z. These effects are shown in Figure 1.

In Figure 1, each row represents a different level of the GARCH parameter α, each column represents a different shock to X holding Z at different values (mean or one SD above/below), each line color represents the effect of holding X at different val-

[6]The GARCH(p,q) model would be specified as: $\sigma_t^2 = \sum_{i=1}^{q} \omega_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2 + \exp(\omega_0 + \mathbf{z}_t \boldsymbol{\gamma})$.

[7]The constant $\omega_0 = 0.5$ and the vector of covariates $\mathbf{z}_t$ include two standard normal variables, X and Z, where $T = 400$ for both series, and where the ARCH parameter $\omega_1 = 0.2$ in all simulations. This follows a 100-period burn in process, described below. The effect of both X and Z on the conditional volatility (that is, their value of γ in Equation 7) is 1.

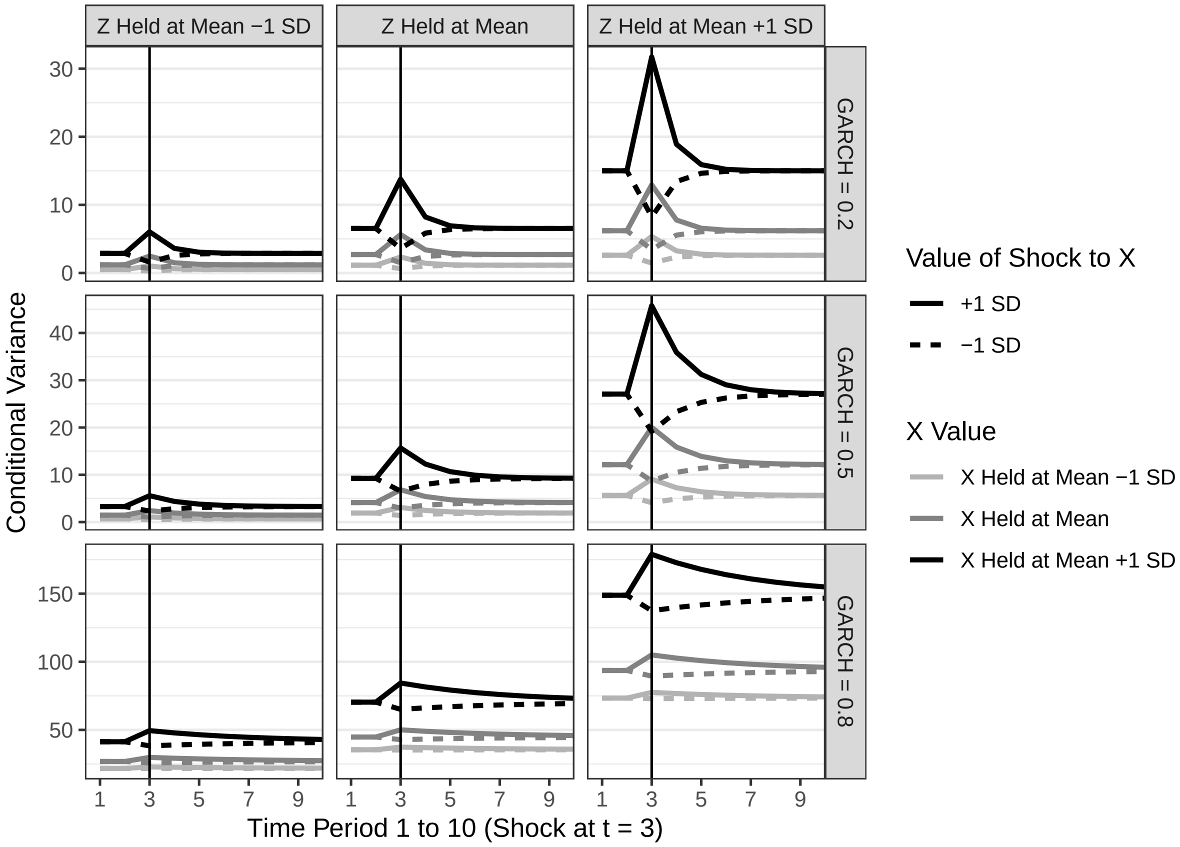


Figure 1: Synthetic GARCH data responding to simulated shocks to X in the conditional error variance (Equation 7)

Note: Dependent variable follows a GARCH(1,1) process; both X and Z are standard normal variables.

ues (mean, one SD above/below), and each type of line represents a positive/negative shock to X (solid a one-SD increase in X, dashed a one-SD decrease in X). The asymmetries are stark. Because of the exponential link in Equation 7, even seemingly marginal changes in the value of any additional control variables in the conditional error variance equation (Z across the columns) can produce large differences in the observed effect of X. For instance, a one-SD increase in X when Z is held at one SD below its mean (the top-left panel) produces a five-unit increase in the conditional error variance. When Z is held at one SD above its mean, the same shock (the top-right panel) produces a 15-unit increase in the conditional error variance, three times the original effect. This asymmetry grows as the magnitude of the GARCH coefficient increases as one moves down the rows. The figure also illustrates the asymmetry between positive and negative shocks to X as well as across different values at which X can be held.

Finally, Figure 1 shows the asymmetric effect of the GARCH parameter α on the impact of X on the conditional error variance. Similar to the coefficient on a lagged dependent variable in an autoregressive distributed lag model, α determines the temporal persistence in the effect of an increase in X on the conditional error variance. When α is relatively low (the top row), shocks to X produce changes in the conditional error variance, but these effects do not persist. Note the top-right panel: when Z is held at its mean and X is held at one SD above its mean (the black line), a one SD shock to X (the solid line) at $t = 3$ produces a 15-unit increase in the conditional error variance of y in the next time period. However, this effect fully dissipates by $t = 6$. When $\alpha = 0.8$, the same shock to X does not dissipate until $t = 10$.

These dynamics would be obscured by relying solely on a table of coefficients. However, that is exactly how GARCH models are typically interpreted in political science. Despite the conditionality of covariate effects in the conditional error variance equation, most scholars rely on tabular results to interpret their statistical and substantive impact. Among the articles mentioned including covariates in the conditional error variance equation, 78 percent interpret their impact by only referencing the sign and statistical significance of tabular coefficients. The remaining 22 percent calculate the predicted marginal effects of their main covariates on the conditional error variance. Reference to the persistence of effects is based on separate discussion of the magnitude of the GARCH term. There is no examination of whether volatility remains statistically and substantively significant across different baseline values of key covariates or any controls (if included) or over time. While this approach is not wrong, it is incomplete; it only allows for understanding the short-term change in volatility for one isolated predictor, without accounting for the previous level of volatility or the levels of other predictors. Yet, Figure 1 demonstrates that this is essential.[8] We found only two articles—Bernhard and Leblang (2006*a*) and Bechtel (2012)—which report associated estimates of uncertainty. In calculating marginal effects, both consider the baseline values at which they set their main covariates and control covariates. But

[8]The GARCH model's exponentiated link function gives more weight to increases in X at higher values. This distorts inferences from other typical quantities of interest, like average marginal effects.

neither article examines the marginal effects of their main covariates over time. To the extent that temporal dynamics are discussed, the authors note whether GARCH parameter estimates are close to one. They do not use GARCH parameter estimates to adjust aggregate covariate marginal effects calculations to show how they evolve or whether remain statistically significant over time. Far more interesting dynamics remain to be explored. De Boef and Keele (2008) taught us to "take time seriously," which we argue should include volatility over time.

Our Approach

We advocate that scholars move beyond tabular results by creating expected values of the conditional variance—along with associated measures of uncertainty—to improve statistical inference and take full advantage of the ARCH-GARCH model.[9] We present three different bootstrapping techniques for generating graphical visualizations of the expected conditional error variance, as well as its temporal evolution in response to hypothetical changes to a covariate(s). These techniques differ in their computational complexity and assumptions underlying measures of uncertainty. We show the feasibility of our approach using the delta method as an alternative technique to bootstrapping when approximating standard errors in the Supplemental Materials (SM).

Our techniques provide five benefits for ARCH-GARCH model interpretation. First, they can be used to produce estimates and graphical depictions of the expected conditional error variance under different substantively interesting covariate scenarios, along with measures of uncertainty to show whether these scenarios produce estimates that are statistically significantly different from one another, whether they undergo any statistically significant changes over time, and whether the rates of decay (due to the GARCH term) are fast or slow. Second, our techniques can be used to produce es-

[9] In the following, basic time series requirements apply during model estimation, such as that the conditional mean and conditional error variance models are correctly specified, stationarity conditions are met, and the errors ε_t are not autoregressive and assumed normally distributed, although non-constant variance is assumed since we are working with GARCH models. We explore consequences of violating these assumptions in the Supplemental Materials (SM).

timates and graphical depictions of the impact of covariate shocks on the expected conditional error variance, along with confidence intervals that demonstrate whether the effects of these shocks are statistically significant in magnitude, whether they are statistically significantly different over time, and whether their rates of decay are fast or slow. Third, they can be used to compare the relative impact of positive vs. negative covariate shocks and any asymmetries in their effects, including whether their magnitudes and temporal evolutions are significantly different from each other, across different periods, and with different rates of decay. Fourth, they can be used to help interpret any interactions in the model; "best practices" recommend graphical representations for interpreting interaction effects (e.g., Brambor, Clark and Golder 2006; Kam and Franzese 2007). Fifth, they can be used to estimate the impact of covariates (across levels, positive/negative values, and/or interactions), while considering these effects at different substantively interesting levels of any additional covariate controls (in the conditional error variance equation). This allows scholars to vary the "all else equal" setting and probe the effects of a chosen covariate, while setting controls at their means or alternatives (medians, modes, or hypothetical levels of interest). In fact, we suggest that GARCH model users generate these graphic visualizations at different levels of their covariate of interest, different shock values (especially positive and negative) to the covariate of interest,[10] and different combinations of values of other predictors to help explore asymmetries in estimated effects, so that they are not simply understanding the "average case" (Hanmer and Kalkan 2013).

Our bootstrapping techniques also improve upon existing ARCH-GARCH post-estimation tools available in most statistical software. In Stata, for instance, users can generate expected values of the conditional mean, the conditional error variance, and the multiplicative heteroskedasticity component (the $\exp(\omega_0 + \mathbf{z}_t \boldsymbol{\gamma})$ term), but they cannot calculate the level of uncertainty around these estimates, limiting their useful-

[10]Our baseline "shock value" typically changes a covariate by one standard deviation, which is standard in the counterfactual simulation literature (e.g. Jordan and Philips 2018*a*; Adolph, Breunig and Koski 2020). Scholars are free to choose whatever counterfactual shock value they prefer, although they should ensure that it is observed in the empirical data, being attentive to realized period-to-period changes in the series (for instance, Lipsmeyer, Philips and Whitten 2023).

ness for inference (which King, Tomz and Wittenberg (2000) note is a general problem plaguing similar estimates). In contrast, our bootstrapping techniques allow researchers to probe ARCH-GARCH results by automating the estimation of the predicted conditional error variance and measures of uncertainty around it over time, across different main covariate values of interest, across different control covariate values, after different main covariate shocks, and under different combinations or interactions of each. They are also an improvement over current practice and available post-estimation tools in another way: they explicitly account for temporal dynamics. They estimate the impact of chosen covariates (under any of the different scenarios above) on the conditional error variance, not just in the contemporaneous period but—by taking into account lagged values of the conditional error variance—dynamically as well.

Parametric-Based Bootstrap

The simplest of our techniques is a parametric bootstrap, first introduced in the context of ARCH-GARCH by Benton and Philips (2020). While it is by far the least computationally intensive of the techniques we present, it comes at a cost of restrictive assumptions about ARCH-GARCH models (discussed below). Parametric bootstrapping creates a distribution of parameter values centered around the original ARCH-GARCH parameter estimates and uses it to generate a large number of expected conditional error variances for each time period. Due to the nature of ARCH-GARCH models, this is a recursive process which takes into account the lagged error variance and the values of any covariates, the latter of which may be changed to substantively interesting values at different points in time. This process consists of the following steps:

1. *Estimate the ARCH-GARCH model.* Obtain the parameter estimates, $\hat{\boldsymbol{\gamma}}$, and corresponding estimated variance-covariance matrix, $\hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}})$.

2. *Generate the parametric bootstrap of the parameter estimates.* Generate a sufficiently large number B of bootstrapped parameters in the ARCH-GARCH equation, $\tilde{\boldsymbol{\gamma}}_b$, for each parameter estimate in $\hat{\boldsymbol{\gamma}}$, drawn from a multivariate normal distribution

with mean equal to the original parameter estimates and variance taken from the estimated variance-covariance matrix: $\tilde{\boldsymbol{\gamma}}_b \sim MVN(\hat{\boldsymbol{\gamma}}, \hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}}))$.[11]

3. *Generate the expected conditional error variance*. For each bootstrap replicate b, generate $\mathrm{E}[\tilde{\sigma}^2_{b,m}]$ for the number of desired time points $m = 1, 2, \cdots, M$, including a time $m = s$ in which a counterfactual "shock" is given to a covariate of interest. This is done in the following steps:

    (a) *Determine the pre-shock baseline values of the covariates in the model*. A usual choice is means (for continuous series) or modes (for categorical series).

    (b) *Conduct a "burn in" process for $m = 1$*. The value of the expected conditional error variance at time $m = 1$ in a GARCH(1,1) requires $\mathrm{E}[\tilde{\sigma}^2_{b,0}]$, which is unknown. However, using the baseline values from the above step, assuming no ARCH effect (i.e., $\varepsilon^2_{m-1} = 0$) or GARCH effect (i.e., $\sigma^2_{b,m-1} = 0$), we can calculate b expected values for the conditional error variance, put this expected value into the GARCH term, calculate a new conditional error variance, and so on, across a number of burn-in periods until the value is stable. The final conditional error variance at $m = 1$ (calculated across all boostrapped estimates, although the value will vary across the bootstraps given uncertainty created by the bootstrapping draws) is $\tilde{\sigma}^2_{b,1}$.

    (c) *Simulate the conditional error variance in the pre-shock period.* For $m = 1, 2, \cdots m = s - 1$, continue to hold all covariates in the GARCH equation at their baseline values, including the the ARCH term $\varepsilon^2_{b,m-1}$ which should remain at zero.[12] The exception is the GARCH term, $\tilde{\sigma}^2_{b,m-1}$, which we set to $\mathrm{E}[\tilde{\sigma}^2_{b,m-1}]$ by using the mean of the conditional error variances calculated in the prior time period. Combine these values with bootstrapped parameter estimates $\tilde{\boldsymbol{\gamma}}_b$ in order to obtain $\tilde{\sigma}^2_{b,m}$.

[11]These simulated parameters should remain consistent with the underlying ARCH-GARCH model and require no additional assumptions beyond those needed to estimate it (see King, Tomz and Wittenberg 2000). A "sufficiently large" number of b is somewhat arbitrary, but $B = 1000$ is conventional.

[12]We assume $\varepsilon^2_{b,m-1} = 0$ since we are interested in the effect of covariates in the GARCH equation, not the effect of unanticipated past shocks, although our approach could easily be adapted to the latter.

(d) *Simulate the impact of the covariate shock on the conditional error variance.* At $m = s$, set the covariate of interest to the value of substantive interest—for instance its mean $\pm$ one standard deviation—keeping all other variables constant as noted above except the GARCH term which should be incorporated as $E[\tilde{\sigma}^2_{b,s-1}]$. Generate b expected values of the conditional error variance $\tilde{\sigma}^2_{b,s}$ using the b simulated parameter estimates. In dynamic simulations, this is argued to be the easiest way of observing a counterfactual "shock" in a variable while keeping close to the "all else equal" interpretation as done in standard regression (c.f., Williams and Whitten 2011, 2012; Philips, Rutherford and Whitten 2016; Gandrud, Williams and Whitten 2016; Jordan and Philips 2018*a*,*b*, 2020); the *only* thing being changed is one of the covariates, which allows us to view the corresponding response in the conditional error variance across both contemporaneous as well as future time periods.

(e) *Simulate the future evolution on the conditional error variance*. For all future periods (e.g., $m = s+1,\ s+2, \cdots, M$) predict future conditional error variances across all b. Note that depending on the context, two future period scenarios may be of interest: either returning the shock to its baseline value (analogous to a temporary "impulse") or keeping it at its new value (akin to a permanent "shift"). We provide examples of both below. Regardless, given the autoregressive nature of the conditional error variance, any movement in response to a counterfactual shock will continue to persist across time (as determined by the magnitude of the GARCH term).

4. *Graph the predictions*. Having calculated $\tilde{\sigma}^2_{b,m}$, plot the predictions. This can be done two ways:

- *Percentile Method:* At each point m in time, plot the median value of $\tilde{\sigma}^2_{b,m}$ and take percentiles of $\tilde{\sigma}^2_{b,m}$ to create upper and lower confidence intervals (e.g., 90, 95%).[13] We prefer the median as opposed to the mean given the asym-

[13]In the SM we also show an alternative plot of the short- and long-run change in the conditional error variance, relative to the pre-shock value.

metric nature of $\tilde{\sigma}^2_{b,m}$ which tends to skew away from zero (since variance cannot be negative). The predictions can be further expressed as a percentage of the "pre-shock" conditional error variance. This provides a much more intuitive interpretation than the original variance metric; for instance, a value of 130% would indicate that volatility has increased by 30 percent.

- *Standard Deviations:* At each point in time, plot $\hat{\sigma}^2_m$, which comes from the original estimates $\hat{\boldsymbol{\gamma}}$ (and user-defined values of covariates). Corresponding 95% confidence intervals are given by $\hat{\sigma}^2_m \pm 1.96 \times \text{SD}(\tilde{\sigma}^2_{b,m})$, where $\text{SD}(\tilde{\sigma}^2_{b,m})$ is the standard deviation of the B bootstrapped conditional error variances calculated at each time point m.

Which to choose? The percentile method has the advantage of showing the skewed nature of the expected conditional error variance. In contrast, the standard deviation approach has the advantage of showing the actual expected conditional error variance calculated from the underlying model, although its confidence intervals might at times overlap with or fall below zero, which is impossible given the nature of error variance.

Residual-Based Bootstrap

Building on a procedure developed by Pascual, Romo and Ruiz (2006) for forecasting prediction intervals for expected future volatility, we also propose a residual-based bootstrap, which constructs bootstrapped conditional mean series on which we estimate ARCH-GARCH models. This process consists of the following steps:

1. *Estimate the ARCH-GARCH model.*

2. *Calculate and rescale residuals.* We first rewrite the residuals as a composite, $\varepsilon_t = \sigma_t \nu_t$, where $\nu_t \sim N(0,1)$, and σ_t (where $\sigma_t > 0$), which allows for time-varying heteroskedasticity in ε_t. Next, isolate the white-noise error process from the heteroskedastic variance component by rescaling the estimated residuals, $\hat{\nu}_t =$

$\frac{y_t - \mathbf{x}_t\hat{\boldsymbol{\beta}}}{\hat{\sigma}_t} = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t}$.[14] These now variance-one estimated residuals are then recentered in order to make them mean-zero: $\tilde{\nu}_t = (\hat{\nu}_t - \frac{1}{T-P}\sum_{t=P+1}^{T}\hat{\nu}_t)$, where P is the highest order of ARCH-GARCH terms (c.f., Chen et al. 2011) and T is the total number of time periods. $\tilde{\nu}_t$ now forms the empirical distribution $\hat{\mathcal{F}}$.

3. *Draw B sets of bootstrap errors.* From $\hat{\mathcal{F}}$, draw a sufficiently large number B sets of bootstrap errors $\tilde{\nu}_t$ over length $T - P$.

4. *Recursively create the bootstrap time series.* Using $\tilde{\nu}_t$, recreate the bootstrap series $\tilde{y}_t$ and bootstrap variance $\tilde{\sigma}_t^2$. This is done in two steps:

   (a) *Initialize the series at $t = 1$.* The value of both the variance and rescaled residuals are unknown at $t = 0$, but are necessary for creating $\tilde{\sigma}_1^2$. We therefore initialize the series at $t = 1$ by taking a single bootstrap draw of both the rescaled residuals as well as the residual variance (call these $\tilde{\nu}_{b,0}$ and $\tilde{\sigma}_{b,0}$, respectively) in order to create the heteroskedastic variance at time $t = 1$, for each bootstrap series b:[15]

$$\tilde{\sigma}_{b,1}^2 = \hat{\omega}_1(\tilde{\nu}_{b,0}^2\tilde{\sigma}_{b,0}^2) + \hat{\alpha}\tilde{\sigma}_{b,0}^2 + \exp(\hat{\omega}_0 + \mathbf{z}_1\hat{\boldsymbol{\gamma}}) \tag{8}$$

   where $\hat{\omega}_1$, $\hat{\alpha}$, $\hat{\omega}_0$, and $\hat{\boldsymbol{\gamma}}$ are estimated parameters from Step 1, $\mathbf{z}_1$ are actual values of the regressors observed at $t = 1$, and $\tilde{\nu}_{b,0}^2$ and $\tilde{\sigma}_{b,0}^2$ are from a bootstrap draw as discussed above.

   (b) *Recursively create the bootstrap time series for $t = 2, \cdots, T - P$* (where $T - P$ is the total length of the original time series after taking into account lagged terms) as follows:

$$\tilde{\sigma}_{b,t}^2 = \hat{\omega}_1(\tilde{\nu}_{b,t-1}^2\tilde{\sigma}_{b,t-1}^2) + \hat{\alpha}\tilde{\sigma}_{b,t-1}^2 + \exp(\hat{\omega}_0 + \mathbf{z}_t\hat{\boldsymbol{\gamma}}) \tag{9}$$

[14] To obtain $\hat{\sigma}_t$ we alter the variance equation, i.e., $\hat{\sigma}_t = \sqrt{\hat{\omega}_1\hat{\varepsilon}_{t-1}^2 + \hat{\alpha}\hat{\sigma}_{t-1}^2 + \exp(\hat{\omega}_0 + \mathbf{z}_t\hat{\boldsymbol{\gamma}})}$.

[15] If a lagged dependent variable is included in the mean equation, a similar bootstrap draw must be made for $y_{b,0}$.

$$\tilde{y}_{b,t} = \mathbf{x}_t \hat{\boldsymbol{\beta}} + \tilde{\nu}_{b,t} \tilde{\sigma}_{b,t} \tag{10}$$

Keeping all regressors ($\mathbf{z}_t$,$\mathbf{x}_t$) fixed at their observed values, recursively create $\tilde{\sigma}^2_{b,t}$ and $\tilde{y}_{b,t}$, respectively (since $\tilde{\sigma}_{b,t}$ is needed in the latter equation). Note that $\tilde{\nu}^2_{b,t}$ are the bootstrapped residuals created in Step 3 above.

5. *Re-estimate the GARCH model*, now using the B new series $\tilde{y}_{b,t}$, and keeping $\mathbf{z}_t$ and $\mathbf{x}_t$ fixed across b. Save the estimated parameters from the variance equation for each b. Once B sets of parameter estimates are obtained, the whole simulation and plotting process works exactly the same as described under the parametric bootstrapping technique, Steps 3-4.

Maximum Entropy Bootstrap

Last, we propose using a maximum entropy bootstrap for creating synthetic $\tilde{y}_t$. Building on a procedure introduced by Vinod and Lopez-de Lacalle (2009) for drawing inferences from time series models, our maximum entropy bootstrap technique resamples the mean series y_t instead of its errors. This process consists of the following steps:

1. *Create random realizations of the mean series.* Create a sufficiently large number B random realizations of the y_t series, following the process described by Vinod (2006) and Vinod and Lopez-de Lacalle (2009):

   (a) Create y^*_t by ordering y_t from its lowest to highest values. Order the original time indicator as well to identify where these newly ordered values fall in the original time series.

   (b) Compute the $T-1$ intermediate points in the newly ordered data y^*_t. Intermediate points are the means of each value and its subsequent value in the newly ordered y^*_t series, specifically, $z_t = (y^*_t + y^*_{t+1})/2$.

   (c) Construct point intervals I_t from the newly ordered y^*_t series. First, identify plausible lower and upper bound limits of the newly ordered y^*_t series. To do this, compute the mean $\bar{z}_t$ of the intermediate point series z_t. Vinod and Lopez-de Lacalle (2009) suggests trimming this mean, say, by 10% to reduce the impact of extreme values. Subtract the trimmed mean $\bar{z}_{t\,\text{trim}}$ from the lowest value of the newly ordered y_t^* series to get $z_0 = y_1^* - \bar{z}_{t\,\text{trim}}$. Add the trimmed mean from the highest value of the newly ordered y_t^* series to get $z_T = y_T^* + \bar{z}_{t\,\text{trim}}$. Second, identify the intervals I_t. To do this, use the lower bound z_0, upper bound z_T, and the intermediate points z_t. The first interval I_1 lies between the lower-bound z_0 and the first observation in the z_1 series, the second lies between the first z_1 and second observations in the z_2 series, and so forth until the final interval I_T has upper bound z_T. There will be T intervals in total.

(d) Compute means for the maximum entropy density for each of the intermediate point intervals I_t. For each of the T intervals, compute the means m_t for the uniform density. For the lowest interval I_1, $m_1 = 0.75y_1^* + 0.25y_2^*$. For the highest interval I_T, $m_T = 0.25y_{T-1}^* + 0.75y_T^*$. For each of the intermediate intervals, $m_t = 0.25y_{t-1}^* + 0.50y_t^* + 0.25y_{t+1}^*$.

(e) Draw T pseudorandom numbers p_t from the $[0, 1]$ uniform interval. Define the range in which each p_t falls, relative to

$$R_t = \left(\frac{t}{T}, \frac{t+1}{T}\right], t = 1, \cdots, T-1 \tag{11}$$

(f) Match R_t to I_t and use these to compute the maximum entropy density. These are the quantiles from the inverse CDF of the maximum entropy density, where the mean of the tth uniform density equals the correct mean. This is the key element of the bootstrap: for each draw of pseudorandom numbers, different elements of I_t will be assigned to different weights from the ME density, allowing variation in the movement between observations.

(g) Reorder the sorted sample. Reorder the sample to its original time series sequence using the original ordering index above. This recovers the temporal dependence of the original data.

2. *Estimate the ARCH-GARCH model.* Using these $\tilde{y}_{b,t}$ series, estimate B ARCH-GARCH models (note that $\mathbf{z}_t$ and $\mathbf{x}_t$ are not bootstrapped). Save the estimated parameters from each variance equation each time. Follow the plotting process as described under the parametric bootstrapping technique, Steps 3-4.

Comparing the Techniques

Table 1 summarizes the three bootstrapping techniques, which we further compare in the SM. The parametric bootstrap is the computationally fastest, requiring the estimation of a single ARCH-GARCH model. However, the sampling distribution of ARCH-GARCH parameters (and the error term) may face greater curvature, heavier tails, and/or greater skew than that underlying the standard normal distribution (Hall and Yao 2003), potentially violating the distributional assumptions behind the parametric bootstrap. In contrast, neither the residual nor the maximum entropy bootstrap require distributional assumptions. Advocates of the maximum entropy bootstrap note its superior performance over other types of time series bootstrap techniques. Unlike the block bootstrap, for example, it can be used to resample time series data with strong temporal dependence, non-stationarity, and residual heteroskedasticity (Yalta 2016).[16] However, the residual bootstrap should be able to handle data with such characteristics, provided that the ARCH-GARCH model is adequately specified; recall that this technique requires recursively creating the dependent variable, conditional on the estimated model. In contrast, the maximum entropy bootstrap resamples y_t using only the maximum entropy density (i.e., no underlying ARCH-GARCH model is required).

In addition to differences in distributions and computational speeds, we evaluate the performance of each of our bootstrapping techniques using a defined data-generating process in the SM. Monte Carlo simulations show that when the residuals are normally distributed (but time-varying heteroskedastic due to the ARCH-GARCH

[16]But see Bergamelli, Novotný and Urga (2015) for a critique.

Table 1: Comparison of Boostrapping Techniques

	Parametric	Residual	Maximum Entropy
Approach to variation	Simulate coefficients	Recreate y series	Recreate y series
Location of bootstrap resampling	Distribution of coefficients $\tilde{\boldsymbol{\gamma}}_b \sim MVN(\hat{\boldsymbol{\gamma}}, \hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}}))$	Distribution of rescaled and recentered residuals $\hat{\mathcal{F}}$	Maximum entropy density of y
Number of models estimated	1	$B+1$	B
Distributional assumption	Yes	No	No

features), the residual bootstrap technique typically outperforms the others in creating *correctly sized* (i.e., not too small, not too large) standard deviations and confidence intervals (around the simulated expected conditional error variance) in both small ($T = 250$) and large ($T = 1000$) series. All techniques improve as T grows larger. However, when the residual component is Student-t distributed, we find that all techniques perform poorly. We thus stress the importance of carefully specifying the model and testing that the residuals are approximately normal. Last, while the residual bootstrap outperforms the the parametric bootstrap, our simulations suggest that the parametric bootstrap performs only slightly worse. This means that the parametric bootstrap's gains in computational speed may make it the preferred choice for some users.

One alternative to bootstrapping is to calculate the conditional error variance and use the delta method approximation to construct the surrounding confidence intervals, which is required since a non-linear combination of regression parameter estimates are being used to calculate the error variance. Such a technique, like the parametric bootstrap, is less computationally intensive, since it only involves estimating a single ARCH-GARCH model. Our Monte Carlo experiments in the SM show that the delta method performs similarly to the parametric bootstrap in terms of producing correctly-sized confidence intervals.[17] However, one disadvantage of the delta method

[17] It is also typically performs best in the Student-t distributed error scenarios.

is the potential to produce negative confidence intervals (similar to the "Standard Deviations" approach when using the bootstrapping techniques discussed above), which is impossible given that conditional error variances are strictly positive. Nevertheless, we discuss the delta method and provide applied examples in the SM.

Replications

To demonstrate the benefits of our proposed bootstrapping techniques as well as their relative performance in an applied setting, we replicate two articles published in top political science journals. Our replications reveal key additional inferences that could have been made had our bootstrapping techniques been used.

Hellwig (2007)

Hellwig (2007) examines how globalization affects support for national governments. He argues that, because globalization raises the exposure of governments to international market dynamics—resulting in a diffusion of policy control away from them and toward private national and international actors—the capacity of voters to evaluate government responsibility for the economy will decline. As a result, rising exposure to the world economy will raise variance in evaluations of government popularity (p. 775). Hellwig tests his argument using quarterly data on trade exposure, capital flows, and government approval across four countries: Denmark, France, the UK, and the US. Since he is interested in explaining variance in government support, Hellwig deploys GARCH models, where the conditional error variance model follows Equation 7 above. His main covariates of interest are the level of trade exposure or capital flows (in separate models), as well as key political covariates specific to each country. After considering covariate coefficients, Hellwig concludes that rising market exposure raises volatility in government support. To the extent that he discusses the temporal dynamics of covariate effects, he notes that the sum of the ARCH and GARCH term results was less than one and thus that the "conditional variance is mean-reverting and not long-memoried" (p. 780).

While the results provide support for his argument, additional inferences are possible using our proposed approach. Take, for example, the results for the impact of trade exposure (*Trade*) in the UK in Hellwig's Model 1, Table 3, which we replicate in Table 2.[18] Based on the sign and significance of the covariate for *Trade* in the conditional error variance equation, he concludes that rising trade exposure is associated with greater volatility in governing party support across all prime ministers. Even so, he notes that there appears to be greater volatility during earlier UK governments (Harold Wilson/James Callaghan, Margaret Thatcher) than the John Major (the omitted category) or Tony Blair periods.

Table 2: Trade Exposure and UK Government Support, Hellwig (2007)

	Coefficient	Standard Error
Mean Equation		
Unemployment_t	0.351	(0.278)
Inflation_t	0.118	(0.134)
$\text{Prospective Evaluations}_t$	0.056***	(0.017)
Political Events	2.167***	(0.371)
Trade_t	0.058	(0.040)
Wilson/Callaghan	1.648	(1.334)
Thatcher	0.526	(0.670)
Blair	-0.752	(0.640)
Constant	-1.794	(1.567)
Variance Equation		
ARCH(1)	0.263**	(0.119)
GARCH(1)	0.333**	(0.143)
Trade_t	0.058*	(0.034)
Wilson/Callaghan	2.661***	(0.992)
Thatcher	1.634**	(0.697)
Blair	-0.219	(0.690)
Constant	-1.765	(1.388)
N	303	
Log Lik.	-698.854	

Note: Dependent variable is governing party support in the UK. GARCH model with robust standard errors in parentheses, replicating Hellwig (2007), Table 3, Model 1. $T = 303$. Additional control variables included in the mean equation but not shown. Two-tailed tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Given different volatility effects observed across governments, we were also cu-

[18]These are almost identical to Hellwig's original results, with small discrepancies due to estimating GARCH models across operating systems (see Benton and Philips 2020, replication files).

rious to see how a shock to *Trade* might affect each. We extend Hellwig's results by examining the impact of a four point increase in *Trade*—a substantively plausible trade exposure shock given that it represents the single largest year-to-year change observed in the data—on the expected conditional error variance of governing party support for each prime minister.[19] To do this, we hold *Trade* at its mean for days 1-2, raise it four points on day 3, and return it to its mean thereafter, for each prime minister.

We present the results for each of our bootstrap procedures using the percentile methods in Figure 2, expressing expected volatility relative to the pre-shock value of volatility (e.g., 100 means that volatility remains the same as the pre-shock value, 120 would represent a 20 percent increase over the pre-shock value, and so on). Each row shows the results of a trade shock under a given Prime Minister, and each column (from left to right) shows the predictions calculated using our parametric, maximum entropy, and residual bootstrap procedures. We present alternative visualizations (such as using standard deviations for confidence intervals, the delta method, and unscaled results) in the SM. As shown in Figure 2, even though the coefficient for *Trade* was statistically significant in Table 2, there appears to be no meaningful increase in the conditional error variance of government support in response to a positive shock to mean trade exposure, no matter the prime minister. Moreover, negative shocks to mean trade exposure do not lead to statistically significant reductions in the conditional error variance of government support either, as shown in Figure 3, which shows a negative four-point shock to *Trade*.

Although Hellwig's tabular results suggests that *Trade* has a (weakly) statistically significant positive effect on the conditional error variance in government support in the UK example, our approach reveals something different: realistic four-point trade exposure shocks do not produce substantively significant effects on the conditional error variance in government support for any of the UK prime ministers.[20] Moreover, we reach this same substantive conclusion no matter the bootstrapping method utilized. It

[19]We chose a four point rise because a one standard deviation change in trade exposure is over three times the size of any year-to-year change in trade that occurred in the sample (13.62 vs. 3.93).

[20]Nor does this result change if we set the pre-shock value of trade to values other than the mean.

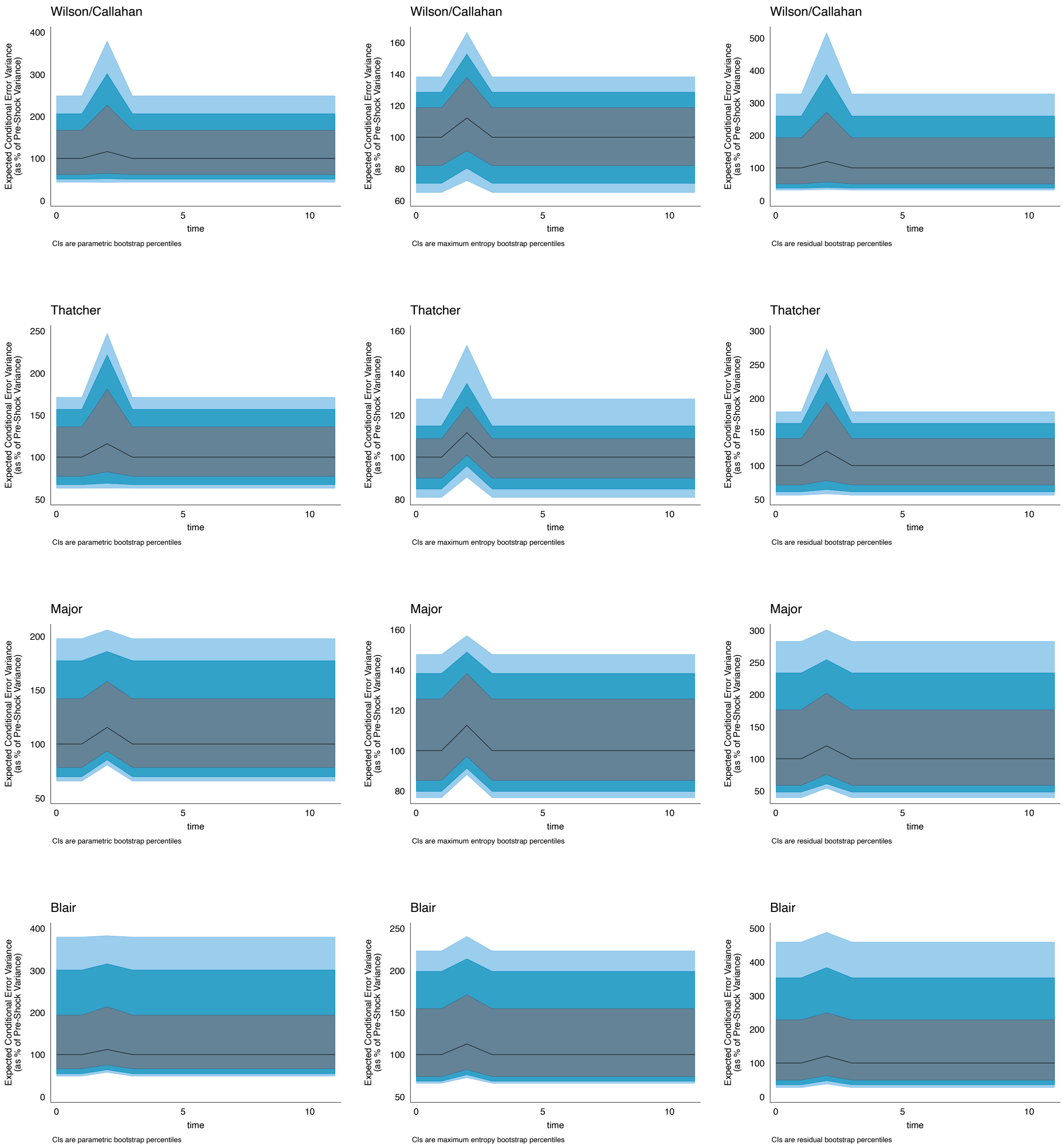


Figure 2: A positive four-point trade shock has little effect on governing party support in the UK

Note: For each prime minister, all other prime minister dummy variables set to 0. From left to right: parametric bootstrap, maximum entropy bootstrap, residual bootstrap. Black line shows median expected conditional error variance as a percentage of the pre-shock variance. Grey: 75% confidence interval, medium blue: 90% confidence interval, light blue: 95% confidence interval.

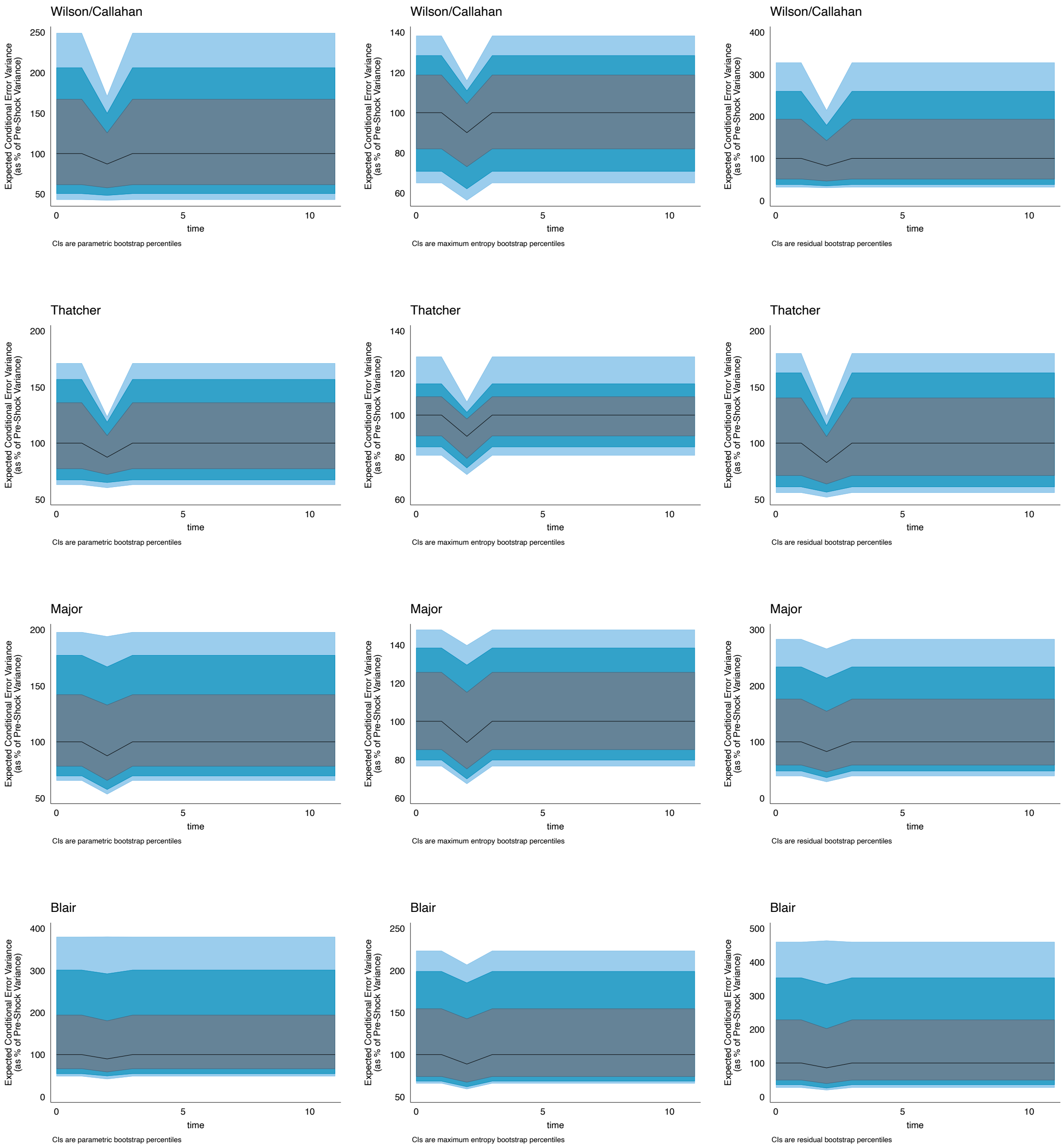


Figure 3: A negative four-point trade shock has little effect on governing party support in the UK

Note: For each prime minister, all other prime minister dummy variables set to 0. From left to right: parametric bootstrap, maximum entropy bootstrap, residual bootstrap. Black line shows median expected conditional error variance as a percentage of the pre-shock variance. Grey: 75% confidence interval, medium blue: 90% confidence interval, light blue: 95% confidence interval.

thus appears that perhaps only very atypical shocks to trade exposure would produce substantively significant effects on volatility in government support.

Schneider and Troeger (2006)

Our second replication shows how our technique travels to ARCH-GARCH extensions. In their influential article, Schneider and Troeger (2006, henceforth ST) examine how international conflict affects financial markets. They argue that international conflicts are characterized by both conflictive and cooperative events, with conflictive events having a greater impact on stock market prices—lowering prices, raising volatility—than cooperative ones. To test their argument, ST examine events during three militarized conflicts over the 1990-2000 period—the conflict between Iraq and several United Nations member states, the conflict between Israel and Palestine, and the Ex-Yugoslavian civil wars—on daily stock market prices in three stock markets: the Dow Jones in New York, the CAC in Paris, and the FTSE in London. To measure these effects, ST use the Goldstein (1992) transformation of the World Event Interaction Survey (WEIS), where numeric values are assigned to conflictive (-10 being the most conflictive) and cooperative (8.3 being the most cooperative) events. Three series are created: a continuous measure of the daily sum of the positive event codings, a continuous measure of the daily sum of negative event codings (absolute value), and a dummy measure noting days with an event coded at the most negative point on the Goldstein transformation of the WEIS scale. Positive (negative) coefficients indicate a positive (negative) impact of cooperative or conflictive events on the stock market.

ST examine their data using an extension of the GARCH approach known as the threshold-GARCH (T-GARCH) model (Zakoian 1994) of the following form:

$$\sigma_t^2 = \omega_1 \varepsilon_{t-1}^2 + \omega_2 \varepsilon_{t-1}^2 D_{t-1} + \alpha \sigma_{t-1}^2 + \exp(\mathbf{z}_t \boldsymbol{\gamma}) \tag{12}$$

where D_{t-1} is a dichotomous indicator equal to one if $\varepsilon_{t-1} > 0$, and zero otherwise. In this specification, if the previous day's error is negative, the ARCH effect is given by

ω_1. If it is positive, the ARCH effect is $\omega_1 + \omega_2$. This allows for asymmetric effects on volatility based on whether unexpected shocks are positive or negative. After examining their results for such asymmetric ARCH effects, ST conclude that "negative shocks have a greater impact on volatility than positive events," as "indicated by a positive and significant γ in the T-GARCH case" and by the fact that "[p]ositive shocks are not even significant in the T-GARCH models" (p.639). "Thus, even though the predictions are not that clear for the expected values, we can detect a significant and stable pattern for the volatility of the three stocks" (p. 639).

We reproduce the results of the T-GARCH model from ST's Table 1 for the first-difference of the daily value of the Dow Jones Industrial Average (DJIA) and report it in Table 3.[21] For the variance equation, severity of the Gulf war (*Gulf Severity*) did not appear to increase volatility in the DJIA. In contrast, increased severity in both the Israeli-Palestinian (*Israel-Palestine Severity*) conflict and Ex-Yugoslavian (*Ex-Yugolslavia Severity*) civil wars tend to raise DJIA volatility. Also, many of the year dummy variables included in the variance equation are statistically significant, and smaller in magnitude at the beginning of the series and larger in magnitude towards the end of the series. This indicates that DJIA volatility increased between 1990 and 2000. The GARCH term is positive and statistically significant, indicating that increased volatility tends to be persistent across trading days. Both the ARCH and T-ARCH terms are positive and statistically significant—albeit in opposite directions. This suggests that when the previous error is negative, ARCH effects are present and volatility persists. In contrast, when the previous error is positive, ARCH effects are effectively zero.

The results in Table 3 also suggest that regional conflicts—especially *Israel-Palestine Severity*—increase volatility in the DJIA. However, it is difficult to picture how severe conflict events affect volatility, especially over time. Our proposed procedure can help visualize how this covariate in the error variance equation affects volatility. In Figure 4, we show the effect of a conflict (coded as most severe on the Goldstein transformation

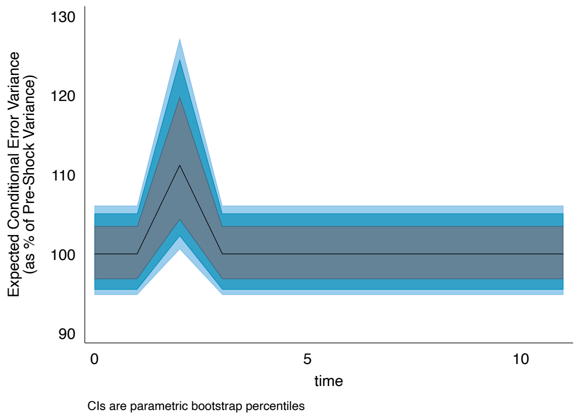
[21]Results differ slightly due to estimation differences between the program they used (EViews) and Stata, although the log-likelihoods are near-identical.

Table 3: International Conflict and Stock Market Returns, Schneider and Troeger (2006)

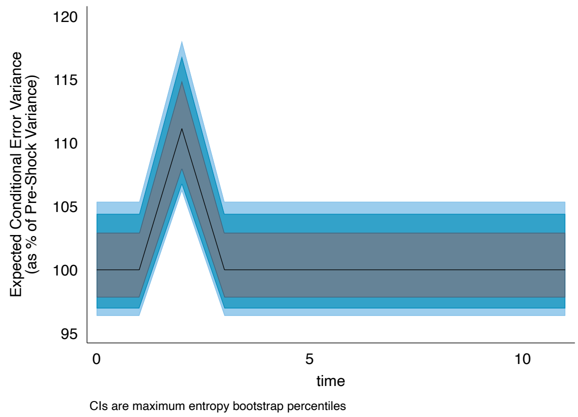
	Coefficient	Standard Error
Mean Equation		
Gulf Severity	-2.584	(4.143)
Gulf Sum of Daily Cooperation	-0.301	(0.223)
Gulf Sum of Daily Conflict	0.261**	(0.109)
Israel-Palestine Severity	2.034	(2.144)
Israel-Palestine Sum of Daily Cooperation	0.065	(0.141)
Israel-Palestine Sum of Daily Conflict	0.015	(0.130)
Ex-Yugoslavia Severity	-3.620*	(1.997)
Ex-Yugoslavia Sum of Daily Cooperation	-0.143	(0.096)
Ex-Yugoslavia Sum of Daily Conflict	-0.006	(0.096)
Differenced Dow Jones-1	-0.019	(0.020)
Differenced CAC	0.203***	(0.028)
Differenced FTSE	0.307***	(0.026)
Constant	1.403*	(0.751)
Variance Equation		
ARCH(1)	0.165***	(0.021)
T-ARCH(1)	-0.192***	(0.022)
GARCH(1)	0.696***	(0.040)
Gulf Severity	0.005	(0.292)
Israel-Palestine Severity	0.319**	(0.142)
Ex-Yugoslavia Severity	0.242*	(0.133)
1991	0.063	(0.153)
1992	-0.346**	(0.162)
1993	-0.492***	(0.159)
1994	0.040	(0.146)
1995	-0.004	(0.147)
1996	1.152***	(0.145)
1997	2.297***	(0.150)
1998	2.574***	(0.139)
1999	2.915***	(0.170)
2000	3.400***	(0.147)
Constant	4.785***	(0.197)
N	2845	
Log Lik.	-14470.190	

Note: Dependent variable is the first-difference of the daily returns of the Dow Jones Industrial Average. $T = 2845$. GARCH model with standard errors in parentheses, replicating Schneider and Troeger (2006), their Table 1. Two-tailed tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

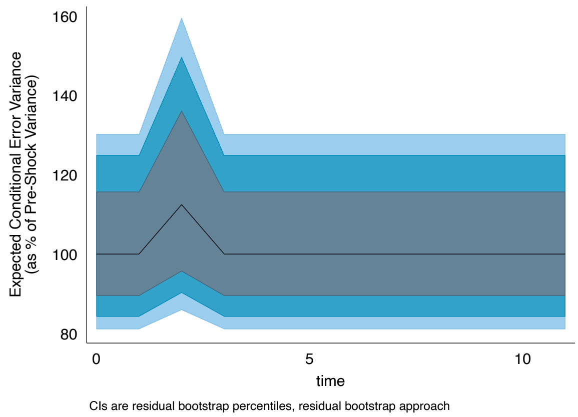
of the WEIS scale) during the Israeli-Palestinian conflict on DJIA volatility. To produce this simulation, we hold the *Israel-Palestine Severity* variable at 0 for days 1 and 2, raise it to 1 on day 3, and return it to 0 on day 4 and thereafter. All other covariates are held at 0. Our simulation shows that volatility increases by around 12 percent but—with the exception of the maximum entropy bootstrap—this short-lived increase is not statistically significantly different from the pre-shock level of volatility. Moreover, the expected conditional error variance moves back to its original value following this one-day conflict. In other words, while a 12 percent increase might be substantively meaningful, we find little evidence that such an increase is statistically significant.



(a) Parametric bootstrap



(b) Maximum entropy bootstrap

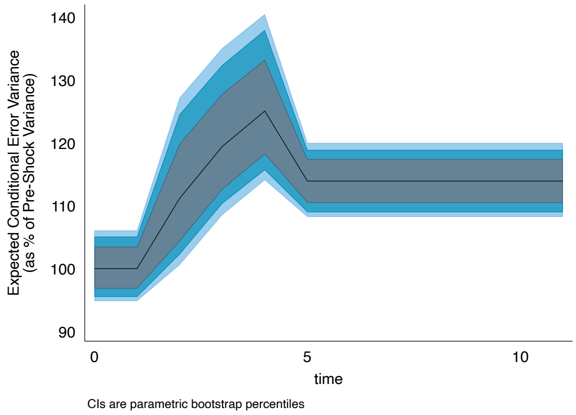


(c) Residual bootstrap

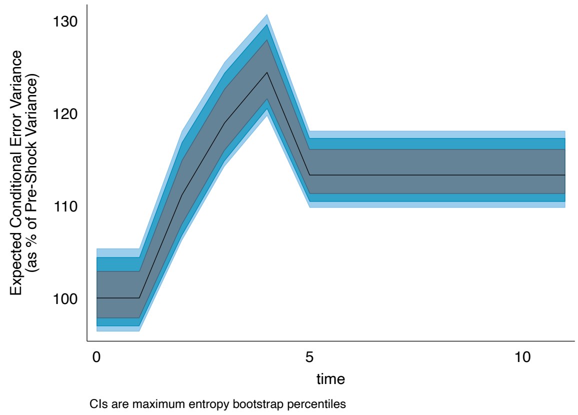
Figure 4: Effect of Palestinian/Israeli conflict severity.

Note: The black line shows median expected conditional error variance as a percentage of the pre-shock variance. Grey: 75% confidence interval, medium blue: 90% confidence interval, light blue: 95% confidence interval.

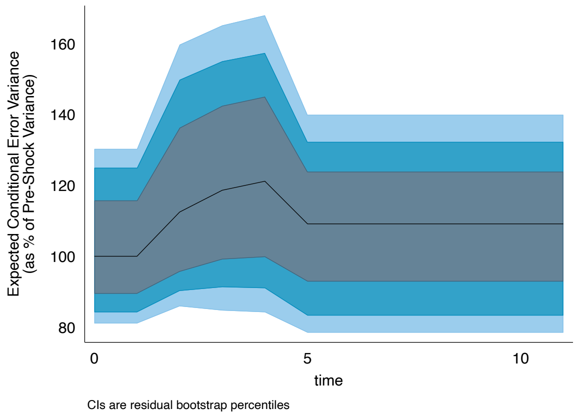
Our graphical approach also allows us to view the effects of more complex relationships, for instance, whether repeated instances of severe conflict lead to larger changes in volatility. In the sample examined by ST, consecutive days where conflict severity reached maximum values on the Goldstein transformation of the WEIS scale are relatively uncommon, but do occur (two successive days occur 19 percent of the time a conflict occurs, and three successive days about 7 percent of the time). In Figure 5, we show the effect of a three-day Israeli-Palestinian conflict (*Israel-Palestine Severity*) coded as most severe on DJIA volatility (i.e., time points $t = 3$ through $t = 5$). In this scenario, we see that a large and long-run permanent increase in volatility occurs, with volatility rising by about 20 percent at its peak over the original volatility level and returning to a steady value of around 15 percent over the pre-shock conditional error variance.



(a) Parametric bootstrap



(b) Maximum entropy bootstrap



(c) Residual bootstrap

Figure 5: Effect of Palestinian/Israeli conflict: Three days of ongoing severity.

Note: The black line shows median expected conditional error variance as a percentage of the pre-shock variance. Grey: 75% confidence interval, medium blue: 90% confidence interval, light blue: 95% confidence interval.

The increase in Figure 5 is statistically significantly different from the pre-shock values for both the parametric and maximum entropy bootstraps, although not for the residual bootstrap. Which bootstrap should we rely on, given these discrepancies? First, we stress that our Monte Carlo experiments in the SM tend to find that the residual bootstrap constructs confidence intervals of the correct size relative to other techniques. Second, users might examine whether results differ largely when other, non-shocked covariates are set to other values as a form of "graphical robustness." Third, users can instead rely on short- and long-run effects plotted against the expected pre-shock value as easier to interpret, if sharp hypothesis tests are of interest, as shown in the SM. Last, users might choose one technique but still report the others, showing that while in terms of statistical significance our conclusions may differ given different bootstrapping techniques, our substantive conclusions (here, that volatility rises sharply in the short-run and persists in the long-run) would remain unchanged.

Overall ST's results provide convincing support for their argument, but our graphical approach produces additional inferences that would have gone unnoticed. It may be that severe events only matter in a substantively significant way to stock market volatility if they occur over consecutive days, at least during the Israeli-Palestinian conflict and the US DJIA. In this case, severe one-off events appear not to be enough to raise US stock market volatility in a meaningful, long-lasting way. Only consecutive severe events achieve this effect, permanently increasing in market volatility.

Conclusion

Despite the usefulness of the ARCH-GARCH approach as a way to model volatility, these models remain difficult to interpret. With few exceptions, scholars have relied on parameter-based interpretations of the statistical and substantive significance of covariate effects in this equation, limiting the quality and range of inferences they can make about the volatility dynamics of their series. In this study, we shed light into this ARCH-GARCH "black box." Using synthetic data, we show that conclusions about expected volatility depend on a number of factors, such as the persistence of the GARCH term, the magnitude and direction of any shocks to the conditional error variance, and the baseline values at which all covariates—including the main covariates of interest, any interaction terms, and any controls—are held. This asymmetry strongly suggests scholars find a way to move past relying on tables alone for interpretation.

To improve ARCH-GARCH model interpretation, we have proposed a series of bootstrapping and graphical techniques to visualize the statistical and substantive effects of covariates on the conditional error variance of a series. Replicating two prominent studies, we show that our approach allows users to better understand and interpret volatility and its temporal dynamics. It also shows that covariate coefficients in conditional error variance equations are often statistically significant in the table of results but not substantively significant when it comes to expected volatility. This is important, as it supports findings in prior time series research (Williams and Whitten 2011, 2012; Philips, Rutherford and Whitten 2016; Gandrud, Williams and Whitten 2016; Jordan and Philips 2018*a*,*b*) as well as in other types of models (King, Tomz and Wittenberg 2000; Kropko and Harden 2020) which show that quantities of interest such as expected values often provide more reliable substantive answers to our research questions—and are far easier to communicate to readers—than tables of coefficient results. While we are not the first to examine and explain the benefits of simulation for improving inferences from (conventional) time series models—and take inspiration from much of the work listed above—we are the first to fully explore the importance of this technique's application to those who have theories about variance.

In undertaking this study, we aim to help scholars to take full advantage of the powerful ARCH-GARCH modeling approach in asking and answering questions about the volatility dynamics of their time series data. In doing so, we contribute to research in time series political methodology dedicated to identifying best practices in model selection and specification (e.g., Lebo, Walker and Clarke 2000; De Boef and Keele 2008; Esarey 2016; Enns et al. 2016; Keele, Linn and Webb 2016; Lebo and Grant 2016; Freeman 2016; Philips 2018; Wilkins 2018; Webb, Linn and Lebo 2019, 2020) by raising awareness of the intricacies of ARCH-GARCH model specification, estimation, and interpretation. To this end, we have created a Stata program to allow practitioners the ability to visualize results from ARCH-GARCH models. These programs will facilitate model interpretation and allow scholars to improve the quality and range of inferences they are able to make.

Acknowledgements

We would like to thank Howard Liu, Vera Troeger, and Clayton Webb for comments on an earlier version of this paper, as well as members of the Dynamic Pie Group at Texas A&M University. A prior version of this paper was presented at PolMeth Europe 2021.

References

Adolph, Christopher, Christian Breunig and Chris Koski. 2020. "The political economy of budget trade-offs." *Journal of Public Policy* 40:25–50.

Bechtel, Michael M. 2012. "Not always second order: Subnational elections, national-level vote intentions, and volatility spillovers in a multi-level electoral system." *Electoral Studies* 31(1):170–183.

Benton, Allyson L and Andrew Q Philips. 2020. "Does the @realDonaldTrump Really Matter to Financial Markets?" *American Journal of Political Science* 64(1):169–190.

Bergamelli, Michele, Jan Novotný and Giovanni Urga. 2015. "Maximum Non-Extensive Entropy Block Bootstrap For Non-Stationary Processes." *L'Actualité Economique* 91(1-2):115–139.

Bernhard, William and David Leblang. 2006*a*. "Polls and Pounds: Public Opinion and Exchange Rate Behavior in Britian." *Quarterly Journal of Political Science* 1:12–47.

Bernhard, William and David Leblang. 2006*b*. *Pricing Politics: Democratic Processes and Financial Markets*. Cambridge: Cambridge University Press.

Boef, Suzanna De. 2000. "Persistence and aggregations of survey data over time: from microfoundations to macropersistence." *Electoral Studies* 19(1):9–29.

Bollerslev, Tim. 1986. "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* 31(3):307–327.

Box-Steffensmeier, Janet M. and Renée M. Smith. 1996. "The Dynamics of Aggregate Partisanship." *American Political Science Review* 90(3):567–580.

Brambor, Thomas, William Roberts Clark and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1):63–82.

Brooks, Sarah M., Raphael Cunha and Layna Mosley. 2022. "Sovereign Risk and Government Change: Elections, Ideology and Experience." *Comparative Political Studies* 55(9):1501–1538.

Chen, Bei, Yulia R Gel, Narayana Balakrishna and Bovas Abraham. 2011. "Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes." *Journal of Forecasting* 30(1):51–71.

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

Enders, Walter. 2015. *Applied econometric time series*. 4 ed. John Wiley and Sons.

Engle, Robert. 2001. "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics." *Journal of Economic Perspectives* 15(4):157–168.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50(4):987–1007.

Enns, Peter K, Nathan J Kelly, Takaaki Masaki and Patrick C Wohlfarth. 2016. "Don't jettison the general error correction model just yet: A practical guide to avoiding spurious regression with the GECM." *Research & Politics* 3(2):1–13.

Esarey, Justin. 2016. "Fractionally integrated data and the autodistributed lag model: Results from a simulation study." *Political Analysis* 24:42–49.

Flink, Carla M. 2017. "Rethinking punctuated equilibrium theory: A public administration approach to budgetary changes." *Policy Studies Journal* 45(1):101–120.

Freeman, John R. 2016. "Progress in the study of nonstationary political time series: A Comment." *Political Analysis* 24(1):50–58.

Füss, Roland and Michael M. Bechtel. 2008. "Partisan politics and stock market performance: The effect of expected government partisanship on stock returns in the 2002 German federal election." *Public Choice* 135(3):131–150.

Gandrud, Christopher, Laron Williams and Guy Whitten. 2016. "Visualize Dynamic Simulations of Autoregressive Relationships in R." *The Political Methodologist* 23(2):6–10.

Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *The Journal of Conflict Resolution* 36(2):369–385.

Hall, Peter and Qiwei Yao. 2003. "Inference in ARCH and GARCH models with heavy–tailed errors." *Econometrica* 71(1):285–317.

Hanmer, Michael J. and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57(1):263–277.

Hellwig, Timothy. 2007. "Economic openness, policy uncertainty, and the dynamics of government support." *Electoral Studies* 26(4):772–786.

Jordan, Soren and Andrew Q Philips. 2018*a*. "Cointegration testing and dynamic simulations of autoregressive distributed lag models." *The Stata Journal* 18(4):902–923.

Jordan, Soren and Andrew Q Philips. 2018*b*. "Dynamic Simulation and Testing for Single-Equation Cointegrating and Stationary Autoregressive Distributed Lag Models." *The R Journal* 10(2):469–488.

Jordan, Soren and Andrew Q. Philips. 2020. "Exploring meaningful visual effects and quantities of interest from dynamic models through dynamac." *Journal of Open Source Software* 5(54):2528.

Journal Citation Reports, Social Sciences Edition, Clarivate Analytics. 2018.

Kam, Cindy D. and Robert J. Franzese. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor: University of Michigan Press.

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

King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2):347–361.

Kriner, Douglas L. 2006. "Examining variance in presidential approval: The case of FDR in World War II." *Public Opinion Quarterly* 70(1):23–47.

Kropko, Jonathan and Jeffrey J. Harden. 2020. "Beyond the Hazard Ratio: Generating Expected Durations from the Cox Proportional Hazards Model." *British Journal of Political Science* 50(1):303–320.

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.

Lebo, Matthew J and Taylor Grant. 2016. "Equation balance and dynamic political modeling." *Political Analysis* 24(1):69–82.

Lipsmeyer, Christine S., Andrew Q. Philips and Guy D. Whitten. 2023. *The Politics of Budgets: Getting a Piece of the Pie*. Cambridge University Press.

Pascual, Lorenzo, Juan Romo and Esther Ruiz. 2006. "Bootstrap prediction for returns and volatilities in GARCH models." *Computational Statistics & Data Analysis* 50(9):2293–2312.

Philips, Andrew Q. 2018. "Have your cake and eat it too? Cointegration and dynamic inference from autoregressive distributed lag models." *American Journal of Political Science* 62(1):230–244.

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):268–283.

Schneider, Gerald and Vera E. Troeger. 2006. "War and the World Economy: Stock Market Reactions to International Conflicts." *Journal of Conflict Resolution* 50(5):623–645.

Vinod, Hrishikesh D. 2006. "Maximum entropy ensembles for time series inference in economics." *Journal of Asian Economics* 17(6):955–978.

Vinod, Hrishikesh D. and Javier Lopez-de Lacalle. 2009. "Maximum Entropy Bootstrap for Time Series: The meboot R Package." *Journal of Statistical Software* 29(5):1–19.

Webb, Clayton, Suzanna Linn and Matthew J. Lebo. 2020. "Beyond the Unit Root Question: Uncertainty and Inference." *American Journal of Political Science* 64(2):275–292.

Webb, Clayton, Suzanna Linn and Matthew Lebo. 2019. "A Bounds Approach to Inference Using the Long Run Multiplier." *Political Analysis* 27(3):281–301.

Wibbels, Erik. 2000. "Federalism and the politics of macroeconomic policy and performance." *American Journal of Political Science* 44(4):687–702.

Wilkins, Arjun S. 2018. "To Lag or Not to Lag?: Re-Evaluating the Use of Lagged Dependent Variables in Regression Analysis." *Political Science Research and Methods* 6(2):393–411.

Williams, Laron K and Guy D Whitten. 2011. "Dynamic simulations of autoregressive relationships." *Stata Journal* 11(4):577–588.

Williams, Laron K and Guy D Whitten. 2012. "But wait, there's more! Maximizing substantive inferences from TSCS models." *The Journal of Politics* 74(03):685–693.

Yalta, A. Talha. 2016. "Bootstrap Inference of Level Relationships in the Presence of Serially Correlated Errors: A Large Scale Simulation Study and an Application in Energy Demand." *Computational Economics* 48(2):339–366.

Zakoian, Jean-Michel. 1994. "Threshold heteroskedastic models." *Journal of Economic Dynamics and Control* 18(5):931–955.

Biographical Statements

Allyson L. Benton is a Reader, Department of Government, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, allyson.benton@essex.ac.uk.

Soren Jordan is an Associate Professor, Department of Political Science, Auburn University, Auburn, AL 36849, sorenjordanpols@gmail.com.

Andrew Q. Philips is an Associate Professor and the Henry W. Ehrmann Professor in Law and Jurisprudence, Department of Political Science, University of Colorado Boulder, UCB 333, Boulder, CO 80309-0333, andrew.philips@colorado.edu.