

How to Improve the Substantive Interpretation of Regression Results when the Dependent Variable is Logged

Oliver Rittmann^{*1}, Marcel Neunhoeffer^{2,3}, and Thomas Gschwend¹

¹University of Mannheim

²Boston University

³LMU Munich

Date: June 15, 2023

Accepted for publication in Political Science Research and Methods

Abstract

Regression models with log-transformed dependent variables are widely used by social scientists to investigate nonlinear relationships between variables. Unfortunately, this transformation complicates the substantive interpretation of estimation results and often leads to incomplete and sometimes even misleading interpretations. We focus on one valuable but underused method, the presentation of quantities of interest such as expected values or first differences on the original scale of the dependent variable. The procedure to derive these quantities differs in seemingly minor but critical aspects from the well-known procedure based on standard linear models. To improve empirical practice, we explain the underlying problem and develop guidelines that help researchers to derive meaningful interpretations from regression results of models with log-transformed dependent variables.

*Corresponding Author: orittman@mail.uni-mannheim.de We thank Marie-Lou Sohnus for excellent research assistance. We are grateful to Daniel Stegmueller, Lukas Stoetzer, Patrick Kraft, Andreas Murr, Marie-Lou Sohnus and the anonymous reviewers feedback on earlier versions of this manuscript. Moreover, we thank the students of our class “Advanced Quantitative Methods” at the University of Mannheim for inspiring us to write this paper. An earlier version of the manuscripts was presented at the first Meeting of the Society for Political Methodology in Europe 2021. Competing interests: The author(s) declare none.

Abstract: 117 words

Manuscript: 3,515 words

Regression models with log-transformed dependent variables are widely used by social scientists to investigate nonlinear relationships between variables. Unfortunately, this transformation complicates the substantive interpretation of respective estimation results. In an effort to improve empirical practice, we clarify one popular strategy for the substantive interpretation of such regression results—the presentation of quantities of interest such as predicted values, expected values, or first differences on the original scale of the dependent variable (King, Tomz and Wittenberg, 2000). We show that calculating such quantities together with their associated uncertainty is different from well-known procedures that work in the case of linear regression models without log-transformed dependent variables. Ignoring this difference can lead to erroneous communication of regression results when the dependent variable is log-transformed.

The key point of confusion is this: A regression with a logged dependent variable estimates $E[\ln(y|X)]$. For a substantive interpretation we want to calculate quantities of interest and their associated uncertainty on the original scale of the dependent variable rather than the logged scale. However, scholars cannot simply exponentiate expected values, standard errors, or upper and lower bounds of the estimated confidence intervals on the logged scale in order to transform them to the original scale. While $\ln(y|X)$ is normally distributed, its transformation $y|X$ back to the original scale is skewed. The consequence is that $e^{E[\ln(y|X)]} \neq e^{\ln(E[y|X])} = E[y|X]$. This is well known among methodologists (e.g. Manning, 1998), but often neglected by substantive scholars. To derive the desired quantities on the original scale together with the associated uncertainty, scholars need to carefully apply appropriate transformation formulas and simulate their respective confidence intervals correctly.

Popular methods textbooks also acknowledge this “retransformation problem” (e.g., Cameron and Trivedi, 2022, Section 4.2.3; see also Cameron and Trivedi, 2005, Section 20.5.2). In this letter, we add two important aspects. Cameron and Trivedi (2022) declare that the prediction $e^{E(\ln(y|X))}$ is a “very poor” and incorrect prediction for $E[y|X]$. We show that $e^{E(\ln(y|X))}$ is indeed not a valid prediction for the mean on the original scale. However, it is still an interesting quantity as it represents the conditional median of the log-normal distribution of $y|X$. Furthermore, the solutions described in Cameron and Trivedi (2022) focus on the point estimate of the conditional mean, but offer little guidance on the uncertainty that comes with those estimates.

Here, we integrate solutions to the retransformation problem of point estimates with the simulation algorithm by [King, Tomz and Wittenberg \(2000\)](#). In doing so, our solution also provides accurate confidence intervals for $E[y|X]$ and other quantities of interest on the original scale of the dependent variable.

Even though the presentation of meaningful quantities of interest became best practice for the interpretation of a wide range of statistical models, our review of current practice shows that substantive scholars make little use of this approach when interpreting results from regression models with logged dependent variables. We base this conclusion on a content analysis of all research articles published in the *American Political Science Review* and *American Journal of Political Science* between 2015 and 2020. In total, we identify 39 articles in which scholars report at least one statistical model with a log-transformed dependent variable.

We identify three main styles of interpretation.¹ First, an “old school” strategy that uses the mere direction (positive/negative) and statistical significance of regression coefficients for interpretation (in 6 out of 39 articles). Those articles provide no substantive interpretation of the respective results and their uncertainty. Second, the most popular practice is the interpretation of regression coefficients as “percent increase” of the dependent variable (used in 31 out of 39 articles). While this practice is not incorrect, we see an important shortcoming. Any concrete interpretation of a “percent increase” does not provide us with an adequate sense of the effect’s absolute magnitude. Furthermore, scholars rarely present the uncertainty associated with a “percent increase”. Finally, there is an interpretation of regression results through the presentation of quantities of interest such as predicted values, expected values, or first differences (in 13 out of 39 articles). This is the one we recommend if done correctly and effectively. Surprisingly, few authors provide uncertainty assessments for their quantities of interest (only 3 of 13 quantities are presented with uncertainty estimates). Because there is no reason why scholars should not be interested in communicating uncertainty on the original scale of the dependent variable, we interpret this result as stemming from a lack of guidance in how to correctly derive quantities of interest together with appropriate confidence intervals on

¹The categories are not mutually exclusive. A list of all articles can be found in the supporting information ([SI.1](#)).

the original scale when the dependent variable is log-transformed.

In this letter, we provide guidance on how substantive scholars can improve their interpretation of regression results when the dependent variable is logged. We show how to calculate quantities of interest on the original scale even when the dependent variable is log-transformed, and how to derive respective confidence intervals using simulations. Furthermore, we highlight how the nonlinear nature of the log-transformation has important consequences for the calculation of first differences and how the presentation of first differences is especially useful for the interpretation of estimation results from models that include interaction terms. We illustrate the utility of our approach with a reanalysis of a recent study on executive appointment processes in the US.

Calculating Quantities of Interest when the Dependent Variable is logged

To calculate quantities of interest on the original scale for models with logged dependent variables, the workflow is to log-transform the dependent variable y , estimate the regression model with $\ln(y)$ as dependent variable, and then use estimation results to calculate quantities of interest, such as $E[\ln(y)]$, although substantively rarely meaningful. To get substantively meaningful quantities, one needs to transform those quantities back to the variable's original scale to get $E[y]$. While the transformation in the first step is fairly simple—we take the natural log of each value of y —the back transformation requires careful thinking.

The back transformation is not straightforward because it maps $\ln(y)$ back to y , which is skewed log-normally distributed conditional on the model. For $y = e^{\ln(y)} = e^{X\beta + \epsilon}$ one can show that $E[y] = E[e^{\ln(y)}] = e^{E[\ln(y)]} \cdot E[e^{\epsilon}] = e^{E[\ln(y)] + \frac{1}{2}\hat{\sigma}^2} > e^{E[\ln(y)]}$ (e.g., Manning, 1998). Therefore, we cannot simply exponentiate $E[\ln(y)]$ to obtain $E[y]$. Table 1 provides an overview of the correct transformation formulas: If we are interested in $E[y]$, we need to transform the estimates on the log-scale with $E[y] = e^{E[\ln(y)] + \frac{1}{2}\hat{\sigma}^2}$. But $e^{E[\ln(y)]}$ is an interesting quantity as well, as this represents the median of the resulting log-normal distribution. Both the mean and the median of y can be interesting and reasonable quantities to present, but researchers must be aware of the difference and should not confuse one with the other when interpreting results. For unimodal, continuous skewed distributions, such as the log-normal distributed y ,

the median is often considered to be a more typical value than the mean (von Hippel, 2005).

To develop a deeper intuition and to illustrate the consequences of choosing the mean or the median as a typical value in the context of linear models, we walk through a motivating example introduced by Rainey (2017). Consider the following data generating process (DGP):

$$\ln(\text{INCOME}) = \beta_{\text{cons}} + \beta_{\text{edu}}\text{EDUCATION} + \epsilon, \text{ and } \epsilon \sim N(0, \sigma^2) \quad (1)$$

The true values of the coefficients are given by $\beta_{\text{cons}} = 2.5$, $\beta_{\text{edu}} = 0.1$, and $\sigma^2 = 1$. The challenge with this DGP is that the dependent variable is log-transformed. Scholars, however, are usually interested in interpreting the results on the original scale of the dependent variable because we are interested in the results in Dollars rather than $\ln(\text{Dollar})$. Suppose that we are interested in a typical income for a person with 20 years of education given our model. How can we calculate such a typical value of income in Dollars, even if the dependent variable is income in $\ln(\text{Dollar})$?

First, we need to choose whether we want to present the median or mean as our typical value. Both can be interesting. To calculate a point estimate of the median income conditional on our scenario of 20 years of education, we get $\text{Med}(Y_c) = e^{2.5+0.1 \times 20} = e^{4.5} \approx 90.06$. If we are interested in a point estimate of the mean income conditional on 20 years of education, we get $E(Y_c) = e^{2.5+0.1 \times 20 + \frac{1}{2} \times 1} = e^5 \approx 148.41$. This shows that the mean and median are two very distinct quantities.²

Simulating Confidence Intervals when the Dependent Variable is logged

Every estimation entails uncertainty. Transparent communication of this uncertainty is fundamental to scientific practice. In this section we show how the simulation approach by King, Tomz and Wittenberg (2000) can be used to get confidence intervals for both, the median and mean of y on the original scale. Figure 1 provides the algorithm to simulate confidence

²The sampling distributions of these estimators for $\widehat{\text{Med}}(Y_c) = e^{X_c \hat{\beta}}$ and $\widehat{E}(Y_c) = e^{X_c \hat{\beta} + \frac{1}{2} \hat{\sigma}^2}$ are right skewed as well and thus the estimators are biased. Rainey (2017) describes this bias as transformation induced bias.

intervals for the mean when the dependent variable is not transformed (left column), the median when the dependent variable is log-transformed (center column), and the mean when the dependent variable is log-transformed (right column). Consider the workflow for models with untransformed dependent variables first.

The procedure consists of four steps. In *step 1* we approximate the distributions of the estimated coefficients to account for *estimation uncertainty*.³ We start by drawing $\tilde{\sigma}^2$ from an inverse-gamma distribution, $\text{Inv-}\Gamma\left(\frac{N-k}{2}, \frac{\hat{\sigma}^2(N-k)}{2}\right)$, where $\hat{\sigma}^2$ is the estimated variance, N is the number of observations, and k is the number of coefficients ($\hat{\beta}$). Next, we approximate the distribution of $\hat{\beta}$ by drawing simulations $\tilde{\beta}$ from the multivariate normal distribution $MVN(\hat{\beta}, \tilde{\sigma}^2(X'X)^{-1})$. In *step 2* we choose our scenario of interest, i.e. we specify covariate values X_c that are held constant during simulation. In *step 3* we calculate $X_c\tilde{\beta}$, the linear combination of the simulations $\tilde{\beta}$ and the chosen values of the covariates (X_c) that define the scenario of interest. This results in a simulated distribution $\tilde{E}(Y_c)$ of expected values of Y conditional on the specified scenario X_c . In *step 4* we get a $(1 - \alpha) \times 100\%$ -confidence interval by summarizing the distribution with the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ percentiles.

Now consider the center column in figure 1 where we outline the simulation procedure for a confidence interval for the conditional median of a model with a logged dependent variable. Using the example in equation 1, suppose that we are interested in the median income in Dollars of a person with 20 years of education. Steps 1 to 3 do not differ from the standard procedure: We draw simulations of the model coefficients ($\tilde{\beta}$ and $\tilde{\sigma}^2$), we set EDUCATION = 20, and compute the linear combination of the simulated coefficients and the scenario of interest. This yields $\tilde{E}(\ln(Y_c))$, a simulated distribution of $\ln(\text{INCOME})$ conditional on 20 years of education. To get confidence intervals for the median in Dollars, we exponentiate the distribution of $\tilde{E}(\ln(Y_c))$ as shown in the transformation step in figure 1. Step 4 then is the same as before. Note that the resulting distribution is skewed and the confidence intervals will not be symmetric around the

³The simulation approach follows an informal Bayesian logic. When we write that we simulate the distribution of a quantity, then this means that we draw from an informal posterior distribution of that quantity.

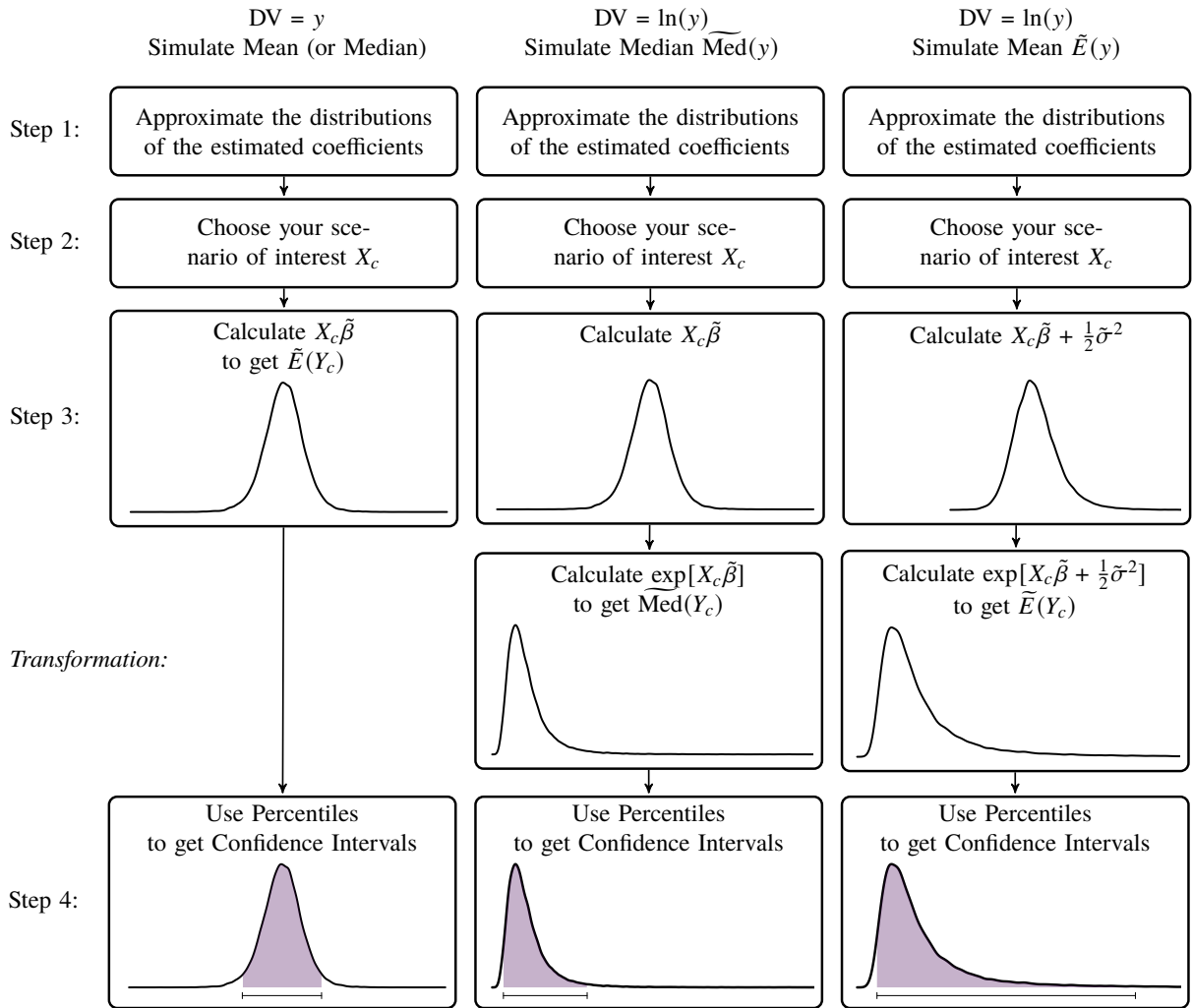


Figure 1: Algorithm to simulate confidence intervals (King, Tomz and Wittenberg, 2000) when the dependent variable is untransformed (left column), for the median when the dependent variable is logged (center column) and for the mean when the dependent variable is logged (right column).

point estimate from the previous section.⁴

In the right column of figure 1, we show how to get confidence intervals for the mean, e.g. the mean income in Dollars of a person with 20 years of education. The procedure mostly remains the same with the exception of step 3 where we need to add $\frac{1}{2}\tilde{\sigma}^2$ to $X_c\tilde{\beta}$. The derivation of confidence intervals for first differences follows equivalently (see table 1 for an overview).⁵

⁴This is also the reason why taking the average of the simulations is not a good strategy to get a point estimate. As the distribution is right skewed, continuous and unimodal, the mean of this distribution will be biased upwards.

⁵We show that the confidence intervals calculated with this procedure have the correct coverage through a Monte Carlo simulation we present in the supporting information (SI.2).

Quantity of Interest		Point Estimate	Simulated Confidence Interval Use $Q_{\frac{\alpha}{2}, 1-\frac{\alpha}{2}}(\cdot)$
Variance	σ^2	$\hat{\sigma}^2$	$\tilde{\sigma}^2 \sim \text{Inv-}\Gamma\left(\frac{N-k}{2}, \frac{\hat{\sigma}^2(N-k)}{2}\right)$
Regression Coefficients	β	$\hat{\beta}$	$\tilde{\beta} \sim \text{MVN}\left(\hat{\beta}, \tilde{\sigma}^2(X'X)^{-1}\right)$
Expected value	$E(Y_c)$	$e^{X_c\hat{\beta} + \frac{1}{2}\hat{\sigma}^2}$	$e^{X_c\tilde{\beta} + \frac{1}{2}\tilde{\sigma}^2}$
Median value	$\text{Med}(Y_c)$	$e^{X_c\hat{\beta}}$	$e^{X_c\tilde{\beta}}$
First Difference	$E(Y_{c_1}) - E(Y_{c_2})$	$e^{X_{c_1}\hat{\beta} + \frac{1}{2}\hat{\sigma}^2} - e^{X_{c_2}\hat{\beta} + \frac{1}{2}\hat{\sigma}^2}$	$e^{X_{c_1}\tilde{\beta} + \frac{1}{2}\tilde{\sigma}^2} - e^{X_{c_2}\tilde{\beta} + \frac{1}{2}\tilde{\sigma}^2}$
First Difference	$\text{Med}(Y_{c_1}) - \text{Med}(Y_{c_2})$	$e^{X_{c_1}\hat{\beta}} - e^{X_{c_2}\hat{\beta}}$	$e^{X_{c_1}\tilde{\beta}} - e^{X_{c_2}\tilde{\beta}}$

Table 1: Transformation formulas for point estimates of common Quantities of Interest, and their approximated distributions to construct correct confidence intervals by using $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ percentiles.

First-Differences and interaction terms with log-transformed dependent variables

We point out two additional issues that arise when presenting quantities of interest on the original scale from regression models with log-transformed dependent variables. First, the magnitude of a first difference based on regression models with log-transformed dependent variables depends on *all* covariates in the scenario, even those that are held constant. This is different for regular linear regression models where the point estimate of the first difference does not depend on the values of covariates that are held constant across scenarios.

Second, if an interaction term is present in the regression model, e.g. $\ln(y) = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 (D \times X) + \epsilon$, then the first difference $E(y|D = 1, X) - E(y|D = 0, X)$ can *increase* at higher levels of X , even if β_3 is *negative* (and vice versa). This is different for regular linear models where the first difference is given by the marginal effect of D , i.e. $\beta_1 + \beta_3 X$, and changes as a linear function of X at a rate of β_3 . Figure 2 demonstrates this notion.

For applied work, these peculiarities show the benefit of presenting quantities of interest of regression models with log-transformed dependent variables on the original scale. At the same time they alert researchers to choose and justify all values in their scenarios carefully. An insensitive selection of any of a model's covariate values may artificially inflate or deflate the size of a first difference.

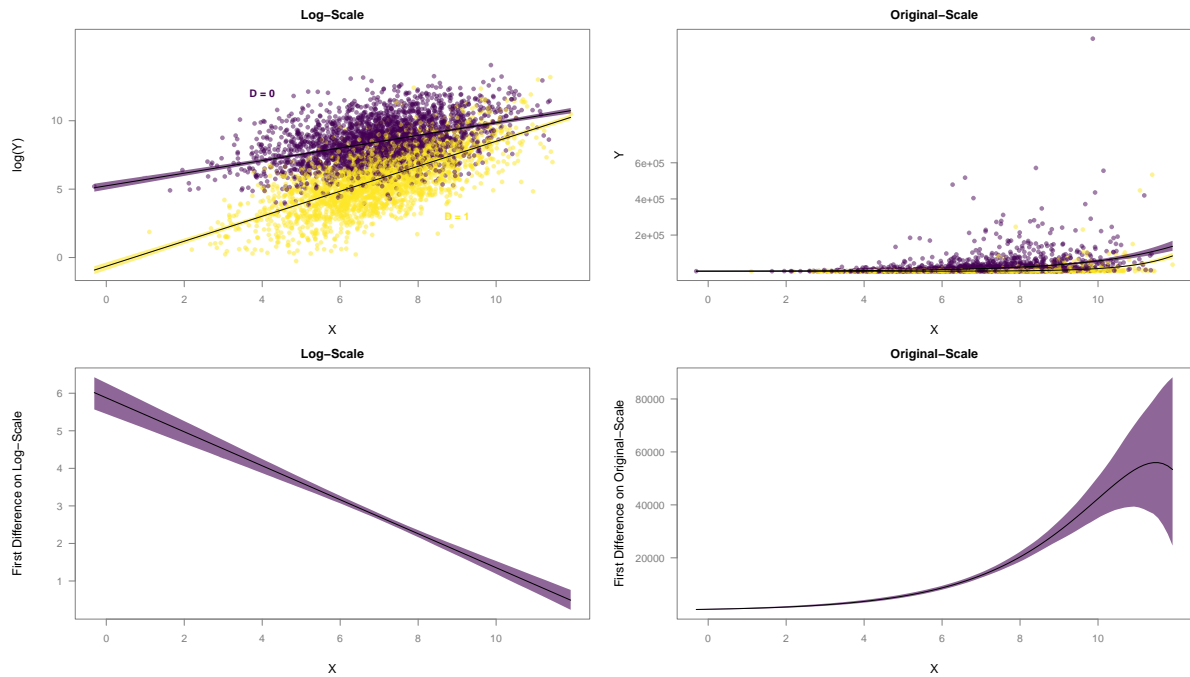


Figure 2: Quantities of interest with an interaction effect. The data generating process follows $\ln(Y) = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 (D \times X) + \epsilon$ with $\beta_0 = 5$, $\beta_1 = -5.5$, $\beta_2 = 0.5$, $\beta_3 = 0.4$ and $\epsilon \sim N(0, 1.5^2)$. As X increases, the first difference on the log-scale of Y decreases, but it increases on the original scale of Y .

Application—A reanalysis of **Hollibaugh and Rothenberg (2018)**

To demonstrate the utility of our approach, we reanalyse the results of a recently published study by **Hollibaugh and Rothenberg (2018)**. The study investigates factors that influence executive appointment processes in the US context. The authors, among other things, study the relation between agency dependence and appointee ideology. We are specifically interested in one of their hypotheses: The higher the independence of the decision-maker in the targeted agency, the higher the ideological divergence between the president and a nominee.⁶

To test this hypothesis, **Hollibaugh and Rothenberg (2018)** estimate linear models. Their dependent variable is the natural log of the ideological divergence between a nominee and the president (*Nominee-President Divergence*). The key independent variable is *Agency Decision Maker Independence*. In support of the hypothesis, **Hollibaugh and Rothenberg (2018)** find that *Agency Decision Maker Independence* is positively associated with the divergence between the president and the nominee.

⁶This relates to hypotheses 6 in the original article. We refer interested readers to the original article for more details on the theoretical arguments behind this hypothesis.

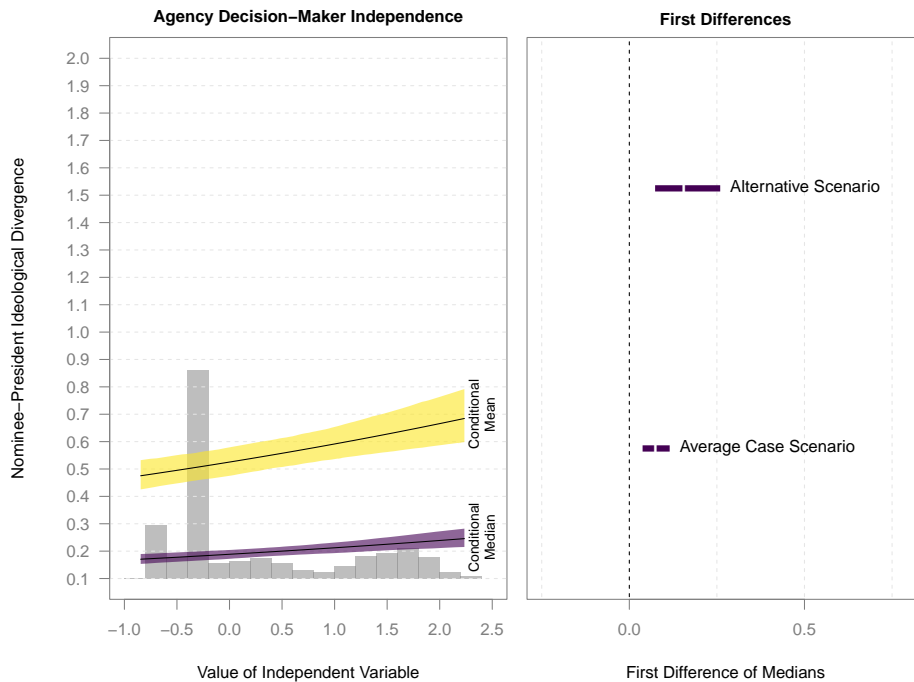


Figure 3: Conditional mean and median values of nominee-president ideological divergence and first difference between minimum and maximum values of the independent variable, conditional on two different sets of covariate values.

To facilitate the interpretation of this effect, [Hollibaugh and Rothenberg \(2018\)](#) report “expected values” of ideological divergence between the president and the nominee from low to high agency decision-maker independence. All binary variables are set to zero, and following an average case approach, all covariate values of continuous variables to their means. We replicate the analysis following our guidelines and present both conditional mean and median values of the dependent variable in figure 3.⁷ Our reanalysis reveals two things: First, researchers are not always aware of the difference between conditional mean and conditional median values of y . What [Hollibaugh and Rothenberg \(2018\)](#) present as expected values are actually conditional median values. Second, this difference is not trivial. Figure 3 demonstrates that the conditional mean and the conditional median of y are two very distinct quantities. The conditional mean values are considerably larger than conditional median values.

Next, we illustrate how the selection of covariate values matters for first differences. The right panel of figure 3 presents two first differences. Both estimates show a first difference of the

⁷[Hollibaugh and Rothenberg \(2018\)](#) z-transform the independent variables and transform simulated values to the empirical percentile scale of the dependent variable.

median between the minimum and maximum values of agency decision-maker independence, but we alternated the values of the covariates that are held constant. One first difference is based on an average case scenario (as in the left panel of figure 3), for the other scenario we set all these covariates to either their minimum or maximum. This would not affect the magnitude of the first difference in regular linear models, but it clearly affects the magnitude in this case where the dependent variable is log-transformed. The first difference amounts to 0.075 for the average case setting, but it roughly doubles to 0.156 if we fix the control variables at more extreme values.

Conclusion

We have shown how to apply appropriate transformation formulas to estimated coefficients of linear regression models with logged dependent variables in order to derive various quantities of interest on the original scale, and how to derive respective confidence intervals using simulations. We conclude with a set of four recommendations that researchers should keep in mind when improving the interpretation of such models.

First, it makes a difference whether conditional mean or median values are presented. Unless there is a special theoretical interest in only one of both quantities, our advice is to present both the conditional mean and the conditional median. Second, point estimates of conditional mean and median values should be calculated directly based on the point estimates of the regression model using appropriate transformation formulas (see table 1). The simulation method in combination with the same formulas allows to derive respective confidence intervals. Third, even values that are held constant across simulations, typically values of control variables, are influential for quantities of interest on the original scale. These values have to be chosen and communicated transparently. Typical strategies are to set those variables to their means, their medians, or to observed values (Hanmer and Kalkan, 2013). Fourth, if the model includes one or more interaction terms, researchers should refrain from interpreting marginal effects on the logged scale. Larger marginal effects on the logged scale do not necessarily reflect larger marginal effects on the original scale. In fact, the opposite may be true as we have shown. To

interpret respective results, our advice is to always calculate first differences on the original scale of the dependent variable.

Software

A software implementation of the proposed method is available as an open-source R package, `simloglm`, at <https://github.com/mneunhoe/simloglm>. We provide a code example for using the package, as well as calculating all mentioned quantities of interest and confidence intervals by hand in R in the supporting information (SI.3).

References

- Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.
- Cameron, A.C. and P.K. Trivedi. 2022. *Microeconometrics Using Stata*. Stata Press.
- 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.
- Hollibaugh, Gary E. and Lawrence S. Rothenberg. 2018. “The Who, When, and Where of Executive Nominations: Integrating Agency Independence and Appointee Ideology.” *American Journal of Political Science* 62(2):296–311.
- 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.
- Manning, Willard G. 1998. “The logged dependent variable, heteroscedasticity, and the re-transformation problem.” *Journal of health economics* 17(3):283–295.
- Rainey, Carlisle. 2017. “Transformation-Induced bias: Unbiased Coefficients Do Not Imply Unbiased Quantities of Interest.” *Political Analysis* 25(3):402–409.
- von Hippel, Paul T. 2005. “Mean, Median, and Skew: Correcting a Textbook Rule.” *Journal of Statistics Education* 13(2).

Supporting Information:

How to Improve the Substantive Interpretation of Regression Results when the Dependent Variable is logged

Oliver Rittmann¹, Marcel Neunhoeffler^{2,3}, and Thomas Gschwend¹.

¹*University of Mannheim.*

²*Boston University.*

³*LMU Munich.*

SI.1 Literature Review

Article	only direction?	% increase?	QoI?	QoI-plot on original scale?	QoI plot on log-scale?
Kriner and Reeves (2015)	X	✓	First Difference	✓	X
Dynes and Huber (2015)	X	✓	First Difference*	X	X
Xu and Yao (2015)	X	✓	X	X	✓
Rogowski (2016)	X	✓	X	X	X
Laitin and Ramachandran (2016)	✓	X	X	X	X
Gulzar and Pasquale (2017)	X	✓	X	X	X
Hainmueller, Hangartner and Pietrantuono (2017)	X	✓	X	X	X
Nellis and Siddiqui (2018)	✓	X	X	X	X
Hyytinen et al. (2018)	X	✓	First Difference*	X	X
Grossman and Michelitch (2018)	✓	X	X	X	X
Szakonyi (2018)	X	✓	X	X	X
Gordon and Simpson (2018)	X	✓	X	X	X
Guardado (2018)	X	✓	First Difference*	X	✓
Ch et al. (2018)	X	✓	X	X	✓
Kim (2018)	✓	X	X	X	X
Li (2018)	X	✓	Predicted Values*	✓*	X
Fouka (2019)	X	✓	X	X	X
Lipsy (2015)	X	✓	X	X	X
Shepherd and You (2020)	X	✓	First Difference*	X	X
Grumbach and Sahn (2020)	X	✓	X	X	X
Gulzar, Haas and Pasquale (2020)	X	✓	First Difference*	X	X
Earle and Gehlbach (2015)	X	✓	X	X	X
Coleman and Mwangi (2015)	X	✓	X	X	X
Beazer and Woo (2016)	X	✓	X	X	X
Berry and Fowler (2016)	X	✓	X	X	X
Zhu (2017)	X	✓	X	X	X
Carnegie and Marinov (2017)	X	✓	First Difference	X	X
Goldstein and You (2017)	X	✓	X	X	X
Fourinaies and Hall (2018)	X	✓	First Difference*	X	X
Fourinaies (2018)	X	✓	X	X	X
Hollibaugh and Rothenberg (2018)	X	X	Expected Values	✓	X
Beazer and Blake (2018)	✓	X	X	X	X
Distelhorst and Locke (2018)	X	✓	First Difference	X	✓
Jiang (2018)	X	X	X	X	X
Paglayan (2019)	X	✓	X	X	X
Mohr et al. (2019)	X	✓	First Difference*	X	✓
Schultz and Mankin (2019)	X	✓	First Difference*	X	X
Pond and Zafeiridou (2020)	X	✓	X	X	X
Jensen, Findley and Nielson (2020)	✓	X	X	X	X

* without uncertainty

Table A1: Results of a content analysis of all research articles published in the *American Political Science Review* and the *American Journal of Political Science* between 2015 and 2020.

SI.2 Coverage Rate: Monte Carlo Evidence

We present Monte Carlo evidence to show that our approach to simulate confidence intervals produces confidence intervals with a proper coverage rate. The Monte Carlo study is set up as follows: We simulate data based on the data generating process described in the main paper, and introduced by [Rainey \(2017\)](#). We restate the DGP here:

$$\ln(\text{INCOME}) = \beta_{\text{cons}} + \beta_{\text{edu}}\text{EDUCATION} + \epsilon, \text{ and } \epsilon \sim N(0, \sigma^2) \quad (2)$$

The true values of the coefficients are given by $\beta_{\text{cons}} = 2.5$, $\beta_{\text{edu}} = 0.1$, and $\sigma^2 = 1$. Further, $N = 10$ with the set of observed values for EDUCATION $\in \{10, 11, 12, 13, 14, 16, 17, 18, 19, 20\}$. We set up a Monte Carlo algorithm that generates data, fits a linear regression, and derives the 95%-confidence intervals for a range of quantities of interest, including conditional median values, conditional mean values, first differences of mean values, first differences of median values, the ratio of median values, and the ratio of mean values. Precisely, the Monte Carlo algorithm takes the following steps:

1. Draw $\epsilon \sim N(0, \sigma^2)$ and compute $\ln(\text{INCOME})$ using equation 2 for all $N = 10$ observations.
2. Regress $\ln(\text{INCOME})$ on EDUCATION via OLS.
3. Use the regression result to simulate the 95%-Confidence Intervals of
 - $\text{Med}(y|\text{edu} = 1)$
 - $\text{Med}(y|\text{edu} = 20)$
 - $E(y|\text{edu} = 1)$
 - $E(y|\text{edu} = 20)$
 - $\text{Med}(y|\text{edu} = 20) - \text{Med}(y|\text{edu} = 1)$
 - $E(y|\text{edu} = 20) - E(y|\text{edu} = 1)$
 - $\text{Med}(y|\text{edu} = 20)/\text{Med}(y|\text{edu} = 1)$
 - $E(y|\text{edu} = 20)/E(y|\text{edu} = 1)$
 following the simulation algorithm presented in the main text with 10,000 simulations.
4. Report whether the true quantity is within the 95%-Confidence Interval.

We repeat this simulation algorithm 10,000 times. Table A2 reports the coverage rate of the quantities of interest. As expected, the coverage rate of all quantities is close to 95%.

Estimand	Coverage (10,000 Monte Carlo Repetitions)
$\text{Med}(y \text{edu} = 1)$	0.9488
$\text{Med}(y \text{edu} = 20)$	0.9500
$E(y \text{edu} = 1)$	0.9484
$E(y \text{edu} = 20)$	0.9448
$\text{Med}(y \text{edu} = 20) - \text{Med}(y \text{edu} = 1)$	0.9498
$E(y \text{edu} = 20) - E(y \text{edu} = 1)$	0.9504
$\text{Med}(y \text{edu} = 20)/\text{Med}(y \text{edu} = 1)$	0.9497
$E(y \text{edu} = 20)/E(y \text{edu} = 1)$	0.9497

Table A2: Monte Carlo Results for the coverage rate of simulated 95%-Confidence Intervals.

SI.3 A simple example with code

```
install_github("mneunhoe/simloglm")
library(simloglm)

# Estimating the model
df <- simloglm::example_df(n = 10)

m1 <- lm(log(income)~educ, data = df)

# Calculating QoI and simulating Confidence Intevals with
  simloglm
set.seed(220609)
res1 <- simloglm(m1, scenario = list(educ = c(1, 20)))

# Summarize results for median
get_summary(res1, which_qoi = "median")

# Summarize results for mean
get_summary(res1, which_qoi = "mean")

# Or get first difference between the two scenarios
get_first_difference(res1, which_qoi = "median")

# Calculating QoI and simulating Confidence Intervals by hand

# Function to sample from inverse gamma distribution
rinvgamma <- function (n,
                        shape,
```

```

        rate = 1,
        scale = 1 / rate)
{
  if (missing(rate) && !missing(scale))
    rate <- 1 / scale
  1 / stats::rgamma(n, shape, rate)
}

# Set up informal posterior of coefficients

# Set number of draws
nsim <- 1000

beta_hat <- coef(m1)
sigma_hat <- summary(m1)$sigma
X_prime_X <- summary(m1)$cov.unscaled

set.seed(220609)
# First  $\sigma^2$ 
sigma2_tilde <- rinvgamma(
  nsim,
  shape = m1$df.residual / 2,
  rate = (sigma_hat ^ 2 * m1$df.residual) / 2
)

# Then the betas

```

```

beta_tilde <- matrix(NA, nrow = nsim, ncol = length(beta_hat)
)

for (sim in 1:nsim) {
  beta_tilde[sim, ] <-
    MASS::mvrnorm(1, beta_hat, X_prime_X * sigma2_tilde[sim
      ])
}

# Set your scenarios as a matrix (don't forget the intercept)
X_c <- rbind(c(1, 1),
             c(1, 20))

# Calculate the linear predictor on the log scale
X_beta <- beta_tilde %*% t(X_c)

# Now transform back to original scale using the appropriate
  formula

# Expected Values/Conditional Mean
# First add the draws of 1/2*sigma2_tilde to each column
X_beta_sigma_tilde <- apply(X_beta, 2, function(x) x + 1/2*
  sigma2_tilde)

# Transform
E_Y_c <- exp(X_beta_sigma_tilde)

# Summarize to get Confidence Intervals
CI_E_Y_c <- apply(E_Y_c, 2, quantile, c(0.025, 0.975))

```

```

# Use beta_hat and sigma_hat for point estimates
X_beta_hat <- beta_hat %*% t(X_c)
X_beta_sigma_hat <- X_beta_hat + 1/2*sigma_hat^2

# Point estimate
E_Y_c_hat <- exp(X_beta_sigma_hat)

# Conditional Median
# First add the draws of 1/2*sigma2_tilde to each column

# Transform
Med_Y_c <- exp(X_beta)

# Summarize to get Confidence Intervals
CI_Med_Y_c <- apply(Med_Y_c, 2, quantile, c(0.025, 0.975))

# Point estimate
Med_Y_c_hat <- exp(X_beta_hat)

# Or get first difference of the medians between the two
  scenarios

# Point estimate
fd_Med_hat <- Med_Y_c_hat[,2] - Med_Y_c_hat[,1]

# Confidence Intevals
fd_Med <- Med_Y_c[,2] - Med_Y_c[,1]

```

```
CI_fd_Med <- quantile(fd_Med, c(0.025, 0.975))
```

SI.4 A Reanalysis of **Shepherd and You (2020)**

We present here a second reanalysis of a prominently published article to show that correctly interpreting effects in regression models is substantively important. **Shepherd and You (2020)** study the influence of career paths of congressional staffers on the legislative output. In particular they are interested in what happens when staffers later become lobbyists. **Shepherd and You (2020, 270)** conclude: “Using comprehensive data on congressional staffers, we find that employing staffers who later become lobbyists is associated with higher legislative productivity for members of Congress, especially in staffers final terms in Congress.”

They run several OLS models with transformed dependent variables.⁸ One of the three dependent variables to measure legislative productivity is the legislative effectiveness score (LES) introduced by **Volden and Wiseman (2014, 2018)**. We focus on the LES since this is the dependent variable that **Shepherd and You (2020)** offer a substantive interpretation for. With the provided replication data we can reproduce the results in the original regression table exactly. Yet, following our guidelines the substantive results using quantities of interest differ. Again, this difference is due to an erroneous transformation of the transformed quantities of interest back to the original scale.

Shepherd and You (2020, Corrigendum 1) report the following substantive effect: “Given that our outcome variables are log-transformed, a one standard deviation increase in the number of future lobbyist staff (0.34) is associated with 1.8% increase in a members Legislative Effectiveness Score (LES) ($\exp(\log(1.7) + 0.0317 \times 0.34) - 1.7 = 0.0184$), if we evaluate the effect [...] at the mean level of LES.” Where 1.7 is the mean of the LES in their sample, 0.0317 is the coefficient for the number of future lobbyists in their model and 0.34 is supposed to be a

⁸The log-transformation of the variables is motivated by “Given that all outcome variables have highly skewed distributions, we use log-transformed variables in the estimation” (**Shepherd and You, 2020, 276**). Unfortunately, this commonly used motivation is wrong, the decision of transforming the dependent variable should only be based on the distribution of $Y|X$, i.e. the error distribution. Moreover, since all of the variables contain 0s, they decide to calculate the log as $\log(x + 1)$. This has to be taken into account when transforming the quantities of interest back to the original scale.

residualized standard deviation, as proposed in [Mummolo and Peterson \(2018\)](#), in the number of future lobbyist staff.

According to our replication analysis the mean of LES is at 1 and the residualized standard deviation of the number of future lobbyists ≈ 0.74 . However, even correcting those two values would still yield an erroneous transformation. Note, that this interpretation neglects the multiplicative nature of a model with a log-transformed dependent variable, the substantive effect depends on a scenario for all other independent variables.⁹

Using the results from the model in [Shepherd and You \(2020, 276, Table 2, Model 4\)](#) and following the steps outlined in section 3 we get that a one standard deviation increase in the number of future lobbyist staff (0.74) is associated with a 5.6% increase in a member's LES for an average member of Congress,¹⁰ with a 95% Confidence Interval from 2.8% to 8.3%. Thus, on average, the effect is more than 3 times as large as reported by [Shepherd and You \(2020\)](#).

Using the 95% Confidence Interval to test the classical two-sided hypothesis whether a one standard deviation increase in future lobbyist staff has an effect (different from 0) on a member's LES, we can reject the Null hypothesis of no effect. This highlights the importance to communicate the uncertainty surrounding quantities of interest and not only relying on a point estimate.

In their paper [Shepherd and You \(2020, 273\)](#) formulate their hypothesis as a directed hypothesis: "Hiring a future revolving-door staffer should be associated with increases in member legislative effectiveness and bill sponsorship activity." Another advantage of the simulation approach is that we can easily calculate the probability that the % Change in Legislative Effectiveness Score that is associated with a one standard deviation increase in future lobbyist staff is greater than 0, a test that is better suited to the directed hypothesis. With our simulation results

⁹The approach of [Shepherd and You \(2020\)](#) would give the same results only if *all* of the other coefficients were exactly 0 or if the intercept is 0 and they chose a scenario where all other covariates are set to 0.

¹⁰Note that this observation does not exist. Other approaches for scenarios based on actual observations have been proposed (e.g. [Hanmer and Kalkan, 2013](#)). Using our R package `simloglm` it is easy to also calculate the percent increase using the observed value approach.

we find that this is the case in all of our 1,000 simulations, thus, yielding a p-value of $< \frac{1}{1000}$ for the one-sided hypothesis test.

Figure A1 displays the correctly calculated results for the scenarios outlined in Shepherd and You (2020). Since the sampling distribution of quantities of interest on the original scale is not necessarily symmetric we advise applied scholars to communicate as much information as possible on the shape of the distribution. This can be achieved, for instance, by reporting multiple confidence intervals.

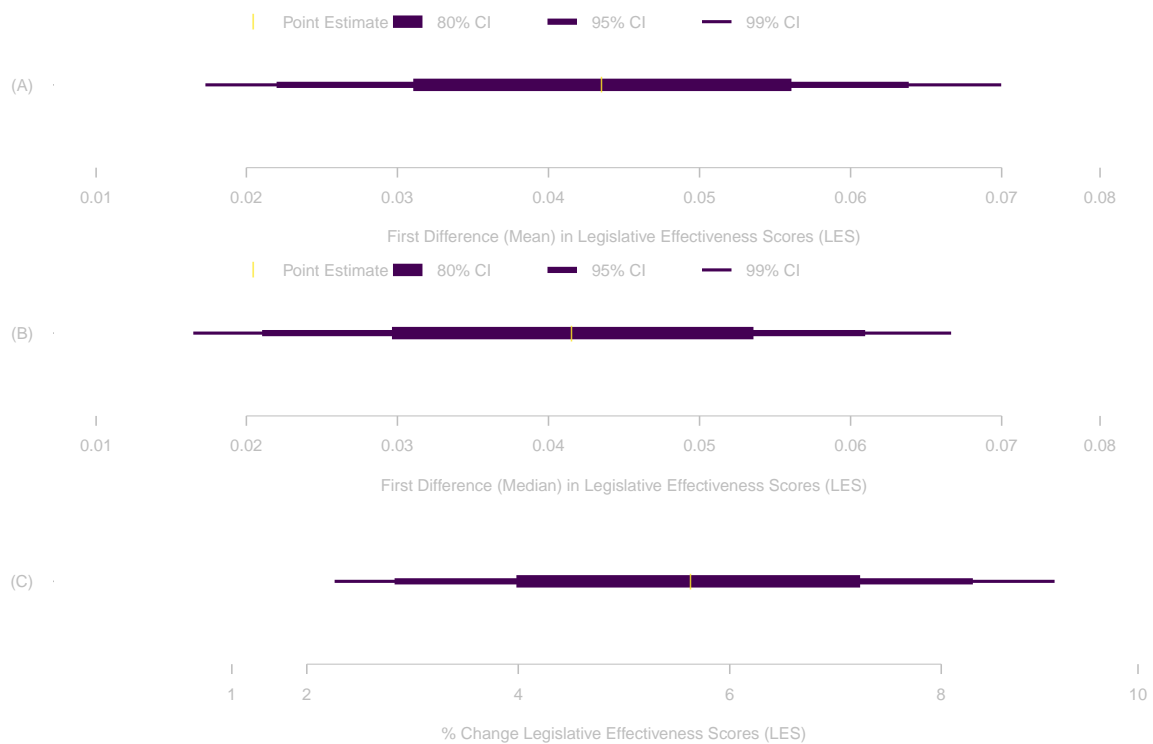


Figure A1: Correct Quantities of Interest based on the results in Shepherd and You (2020). Results for the effect of a one (residualized) standard deviation increase (0.74) in the number of future lobbyists on the Legislative Effectiveness Score (LES), where all other covariate values are set to their means. Panel (A) shows the First-Difference between Expected Values. Panel (B) shows the First-Difference between Medians. Panel (C) shows the % Change.

References

- Beazer, Quintin H. and Byungwon Woo. 2016. "IMF Conditionality, Government Partisanship, and the Progress of Economic Reforms." *American Journal of Political Science* 60(2):304–321.
- Beazer, Quintin H. and Daniel J. Blake. 2018. "The Conditional Nature of Political Risk: How Home Institutions Influence the Location of Foreign Direct Investment." *American Journal of Political Science* 62(2):470–485.
- Berry, Christopher R. and Anthony Fowler. 2016. "Cardinals or Clerics? Congressional Committees and the Distribution of Pork." *American Journal of Political Science* 60(3):692–708.
- Carnegie, Allison and Nikolay Marinov. 2017. "Foreign Aid, Human Rights, and Democracy Promotion: Evidence from a Natural Experiment." *American Journal of Political Science* 61(3):671–683.
- Ch, Rafael, Jacob Shapiro, Abbey Steele and Juan F. Vargas. 2018. "Endogenous Taxation in Ongoing Internal Conflict: The Case of Colombia." *American Political Science Review* 112(4):905–917.
- Coleman, Eric A. and Esther Mwangi. 2015. "Conflict, Cooperation, and Institutional Change on the Commons." *American Journal of Political Science* 59(4):855–865.
- Distelhorst, Greg and Richard M. Locke. 2018. "Does Compliance Pay? Social Standards and Firm-Level Trade." *American Journal of Political Science* 62(3):695–711.
- Dynes, Adam M. and Gregory A. Huber. 2015. "Partisanship and the Allocation of Federal Spending: Do Same-Party Legislators or Voters Benefit from Shared Party Affiliation with the President and House Majority?" *American Political Science Review* 109(1):172–186.
- Earle, John S. and Scott Gehlbach. 2015. "The Productivity Consequences of Political Turnover: Firm-Level Evidence from Ukraine's Orange Revolution." *American Journal of Political Science* 59(3):708–723.
- Fournaies, Alexander. 2018. "When Are Agenda Setters Valuable?" *American Journal of Political Science* 62(1):176–191.
- Fournaies, Alexander and Andrew B. Hall. 2018. "How Do Interest Groups Seek Access to

- Committees?" *American Journal of Political Science* 62(1):132–147.
- Fouka, Vasiliki. 2019. "How do immigrants respond to discrimination? the case of Germans in the US during World War I." *American Political Science Review* 113(2):405–422.
- Goldstein, Rebecca and Hye Young You. 2017. "Cities as Lobbyists." *American Journal of Political Science* 61(4):864–876.
- Gordon, Sanford C. and Hannah K. Simpson. 2018. "The birth of pork: Local appropriations in America's first century." *American Political Science Review* 112(3):564–579.
- Grossman, Guy and Kristin Michelitch. 2018. "Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda." *American Political Science Review* 112(2):280–301.
- Grumbach, Jacob M. and Alexander Sahn. 2020. "Race and representation in campaign finance." *American Political Science Review* 114(1):206–221.
- Guardado, Jenny. 2018. "Office-Selling, Corruption, and Long-Term Development in Peru." *American Political Science Review* 112(4):1104–1110.
- Gulzar, Saad and Benjamin J. Pasquale. 2017. "Politicians, Bureaucrats, and Development: Evidence from India." *American Political Science Review* 111(1):162–183.
- Gulzar, Saad, Nicholas Haas and Benjamin Pasquale. 2020. "Does Political Affirmative Action Work, and for Whom? Theory and Evidence on India's Scheduled Areas." *American Political Science Review* 114(4):1230–1246.
- Hainmueller, Jens, Dominik Hangartner and Giuseppe Pietrantuono. 2017. "Catalyst or crown: Does naturalization promote the long-term social integration of immigrants?" *American Political Science Review* 111(2):256–276.
- 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.
- Hollibaugh, Gary E. and Lawrence S. Rothenberg. 2018. "The Who, When, and Where of Executive Nominations: Integrating Agency Independence and Appointee Ideology." *American Journal of Political Science* 62(2):296–311.
- Hyytinen, Ari, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen and Janne Tukiainen.

2018. “Public Employees as Politicians: Evidence from Close Elections.” *American Political Science Review* 112(1):68–81.
- Jensen, Nathan M., Michael G. Findley and Daniel L. Nielson. 2020. “Electoral institutions and electoral cycles in investment incentives: A field experiment on over 3,000 US municipalities.” *American Journal of Political Science* 64(4):807–822.
- Jiang, Junyan. 2018. “Making Bureaucracy Work: Patronage Networks, Performance Incentives, and Economic Development in China.” *American Journal of Political Science* 62(4):982–999.
- Kim, Sung Eun. 2018. “Media Bias against Foreign Firms as a Veiled Trade Barrier: Evidence from Chinese Newspapers.” *American Political Science Review* 112(4):918–938.
- Kriner, Douglas L. and Andrew Reeves. 2015. “Presidential Particularism and Divide-the-Dollar Politics.” *American Political Science Review* 109(1):155–171.
- Laitin, David D. and Rajesh Ramachandran. 2016. “Language policy and human development.” *American Political Science Review* 110(3):457–480.
- Li, Zhao. 2018. “How Internal Constraints Shape Interest Group Activities: Evidence from Access-Seeking PACs.” *American Political Science Review* 112(4):1067–1082.
- Lipsy, Phillip Y. 2015. “Explaining Institutional Change: Policy Areas, Outside Options, and the Bretton Woods Institutions.” *American Journal of Political Science* 59(2):341–356.
- Mohr, Zachary, Jo Ellen V. Pope, Martha E. Kropf and Mary Jo Shepherd. 2019. “Strategic Spending: Does Politics Influence Election Administration Expenditure?” *American Journal of Political Science* 63(2):427–438.
- Mummolo, Jonathan and Erik Peterson. 2018. “Improving the Interpretation of Fixed Effects Regression Results.” *Political Science Research and Methods* 6(4):829–835.
- Nellis, Gareth and Niloufer Siddiqui. 2018. “Secular Party Rule and Religious Violence in Pakistan.” *American Political Science Review* 112(1):49–67.
- Paglayan, Agustina S. 2019. “Public-Sector Unions and the Size of Government.” *American Journal of Political Science* 63(1):21–36.
- Pond, Amy and Christina Zafeiridou. 2020. “The Political Importance of Financial Performance.” *American Journal of Political Science* 64(1):152–168.

- Rainey, Carlisle. 2017. "Transformation-Induced bias: Unbiased Coefficients Do Not Imply Unbiased Quantities of Interest." *Political Analysis* 25(3):402–409.
- Rogowski, Jon C. 2016. "Presidential Influence in an Era of Congressional Dominance." *American Political Science Review* 110(2):325–341.
- Schultz, Kenneth A. and Justin S. Mankin. 2019. "Is Temperature Exogenous? The Impact of Civil Conflict on the Instrumental Climate Record in Sub-Saharan Africa." *American Journal of Political Science* 63(4):723–739.
- Shepherd, Michael E. and Hye Young You. 2020. "Exit Strategy: Career Concerns and Revolving Doors in Congress." *American Political Science Review* 114(1):270–284.
- Szakonyi, David. 2018. "Businesspeople in elected office: Identifying private benefits from firm-level returns." *American Political Science Review* 112(2):322–338.
- Volden, Craig and Alan E Wiseman. 2014. *Legislative effectiveness in the United States congress: The lawmakers*. Cambridge University Press.
- Volden, Craig and Alan E. Wiseman. 2018. "Legislative effectiveness in the United States senate." *Journal of Politics* 80(2):731–735.
- Xu, Yiqing and Yang Yao. 2015. "Informal institutions, collective action, and public investment in rural China." *American Political Science Review* 109(2):371–391.
- Zhu, Boliang. 2017. "MNCs, Rents, and Corruption: Evidence from China." *American Journal of Political Science* 61(1):84–99.