Supplemental Material (not copyedited or formatted) for: Gregory A. Caldeira, Daniel Lempert. 2020.
"Selection of Cases for Discussion: The U.S. Supreme Court, OT 1939, 1968, and 1982."
Journal of Law and Courts 8(2). DOI: https://doi.org/10.1086/709915.

Appendix

An Alternative Approach to Calculating Second Differences

After fitting a binary choice model, it is now standard practice to present (changes in) predicted probabilities to describe the marginal effect of a given covariate. It is also now well-understood that the magnitude of these marginal effects is a function of the values that all other independent variables take on, since binary choice models are nonlinear. We now discuss some implications of this fact that are relevant for calculating second differences. In our exposition and application, we discuss marginal effects after fitting a logit model, and we consider binary covariates and so are concerned with the effects of discrete changes (or "first differences"), but our basic argument is applicable more generally.

Fitting a logit yields the following function for calculating predicted probabilities: letting z be the covariate of interest, $x_1, \ldots, x_n$ the other independent variables, α the constant, and β_i the estimated coefficients,

$$\Pr(y = 1 \mid x_1, \ldots, x_n, z) = \frac{\exp(\alpha + \beta_1 x_1 + \ldots + \beta_n x_n + \beta_z z)}{1 + \exp(\alpha + \beta_1 x_1 + \ldots + \beta_n x_n + \beta_z z)}. \tag{1}$$

This function's rate of increase *increases* the closer $\Pr(y = 1)$ is to 0.5; it takes on an "S-shape." Consider now the calculation of the discrete change in $\Pr(y = 1 \mid x_1, \ldots, x_n)$ as z goes from 0 to 1. Assuming with minimal loss of generality that $\beta_z > 0$, the slope of the function means that this discrete change will be the smallest when $x_1, \ldots, x_n$ is chosen such that $\Pr(y = 1 \mid x_1, \ldots, x_n, z = 0)$ is as close to 0 as possible, or when $\Pr(y = 1 \mid x_1, \ldots, x_n, z = 1)$ is as close to 1 as possible; and the discrete change will be largest when when $x_1, \ldots, x_n$ is chosen such that $\Pr(y = 1 \mid x_1, \ldots, x_n, z = 1)$ and $\Pr(y = 1 \mid x_1, \ldots, x_n, z = 0)$ are equidistant from 0.5. The critical point is that this discrete change is a function of the choice of $x_1, \ldots, x_n$.

Now suppose one is interested in whether the effect of a covariate on $\Pr(y = 1)$ differs across two groups—in our case, two terms of the Court. A standard approach is to fit a logit model with an interactive term indicating group membership (e.g., Rainey 2016). Letting

g be the binary variable indicating group membership, we have the following function for calculating predicted probabilities:

$$\begin{aligned}&\Pr(y = 1 \mid x_1, \ldots, x_n, g, z) \\ &\quad = \frac{\exp(\alpha + \beta_g g + \beta_1 x_1 + \ldots + \beta_n x_n + \beta_z z + \beta_{g1} g x_1 + \ldots + \beta_{gn} g x_n + \beta_{gz} g z)}{1 + \exp(\alpha + \beta_g g + \beta_1 x_1 + \ldots + \beta_n x_n + \beta_z z + \beta_{g1} g x_1 + \ldots + \beta_{gn} g x_n + \beta_{gz} g z)}. \qquad (2)\end{aligned}$$

For any constituent term not interacted with the group variable, $\beta_{gi} = 0$ by assumption in (2).

An appropriate way to determine the difference in effects across groups is to calculate a second difference:

$$\begin{aligned}\Delta\Delta[\Pr(Y)] \equiv\ &[\Pr(y = 1 \mid x_1, \ldots, x_n, g = 0, z = 1) - \Pr(y = 1 \mid x_1, \ldots, x_n, g = 0, z = 0)] \\ &- [\Pr(y = 1 \mid x_1, \ldots, x_n, g = 1, z = 1) - \Pr(y = 1 | x_1, \ldots, x_n, g = 1, z = 0)]. \qquad (3)\end{aligned}$$

This is the effect in "Group 1" subtracted from the effect in "Group 0"—the difference in effects between groups (i.e., the difference between the first differences). For brevity, we will refer to the second term on the right hand side of (3) as the "baseline probability" for Group 0 and fourth term as the "baseline probability" for Group 1.

Simplifying from (2), the baseline probability for Group 0 is:

$$\Pr(y = 1 \mid x_1, \ldots, x_n, g = 0, z = 0) = \frac{\exp(\alpha + \beta_1 x_1 + \ldots + \beta_n x_n)}{1 + \exp(\alpha + \beta_1 x_1 + \ldots + \beta_n x_n)}, \qquad (4)$$

and the baseline probability for Group 1 is:

$$\Pr(y = 1 \mid x_1, \ldots, x_n, g = 1, z = 0) = \frac{\exp(\alpha + \beta_g + \beta_1 x_1 + \ldots + \beta_n x_n + \beta_{g1} x_1 + \ldots + \beta_{gn} x_n)}{1 + \exp(\alpha + \beta_g + \beta_1 x_1 + \ldots + \beta_n x_n + \beta_{g1} x_1 + \ldots + \beta_{gn} x_n)}. \quad (5)$$

We present (4) and (5) to make clear that unless β_g (the group intercept) and $\beta_{g1}, \ldots, \beta_{gn}$ (the coefficients attending the interaction terms) are each equal to 0—or $\beta_g + \beta_{g1} x_1 + \ldots + \beta_{gn} x_n$ otherwise sum to 0—the baseline probability for Group 0 will not equal the baseline probability for Group 1. Thus, because of the nonlinearity of (2), the effect for Group 0 will not equal the effect of Group 1, even if $\beta_{gz} = 0$. More generally, the second difference depends not only on β_{gz}, but also on $\beta_{g1}, \ldots, \beta_{gn}$—and thus, we emphasize, on the choice of $x_1, \ldots, x_n$—as well as on β_g. This has some perhaps under-appreciated consequences for the choice of $x_1, \ldots x_n$ when calculating second differences.

Indeed, the choice of $x_1, \ldots x_n$ when calculating second differences gets very little attention in the literature. The typical approach seems to be set x_i in (4), the baseline probability for Group 0, equal to x_i in (5), the baseline probability for Group 1. For example, x_i may be mean or modal value of the covariate in the estimation sample. We call this the "representative case" approach. The representative case approach addresses the question, "how does the effect of some covariate of interest vary across two units that have identical covariate values, but are in different groups?" Note, though, that these two units may have very different baseline probabilities, with the implications that we have discussed above.

An alternative to the representative case approach is the "observed-value" approach, which has recently been defended (Hanmer and Kalkan 2013) as often the most appropriate way to calculate first differences (i.e., for assessing non-interactive effects). The observed-value approach involves calculating effect sizes for every observation in the sample by varying only the covariate of interest and holding the other variables at their in-sample value, and averaging those effects to give a summary measure. It is reasonably straightforward to extend the observed-value approach to calculating second differences, but we simply note here that the (average) baseline probability may be very different across groups for this approach too. For a more thorough discussion of the representative case and observed-value and approaches, and a summary of current methodological advice to social scientists, see Long and Mustillo (2019).

That baseline probabilities will vary across groups for both the representative case and the observed-value approach, and the attendant implications for calculation of second differences, does not make these these calculations *wrong*. Under the assumptions of the model, they are correct. Still, in some contexts, we suggest that the relevant question might *not* be "how does some effect vary across two units with the same covariate values, but in different groups?" but, instead, "how does some effect vary across two units with the same baseline probability (of a positive outcome), but in different groups?" This implies calculating second differences by setting $x_1, \ldots x_n$ so that $\beta_g + \beta_{g1}x_1 + \ldots + \beta_{gn}x_n \approx 0$.

Software: sdcasepick

We briefly describe the Stata program `sdcasepick`, which allows for convenient calculation of effect size comparisons that follow the recommendations here.[12] After the user estimates a logit (or other binary choice) regression that includes a (binary) group indicator, the software allows the user to specify any covariate of interest included in the previously estimated regression and a range of baseline probabilities; within that range, the software finds the pair of observations (one in each group) in the estimation sample that are closest in terms of baseline probability. The software then calculates first and second differences (and associated standard errors) by setting the other covariates at the values in this closest pair of observations. Thus, effect size comparisons are based on cases with comparable baseline probabilities, as we recommend. An additional advantage is that, as implemented, the cases forming the bases for comparison are guaranteed to be present in the sample, which ameliorates concerns about out-of-sample

[12] `sdcasepick` is included with replication materials and relies on Stata's `margins` command and Long and Freese's (2014) `m*` commands.

Supplemental Material (not copyedited or formatted) for: Gregory A. Caldeira, Daniel Lempert. 2020.
"Selection of Cases for Discussion: The U.S. Supreme Court, OT 1939, 1968, and 1982."
Journal of Law and Courts 8(2). DOI: https://doi.org/10.1086/709915.

comparisons and extreme counterfactuals.