Unnecessary Bias

Do Not Use the Average of Simulations to Estimate the Quantity of Interest[*]

Carlisle Rainey[†]

Abstract

Following King, Tomz, and Wittenberg (2000), researchers commonly convert coefficient estimates into an estimate of the quantity of interest using the average of simulations. However, other researchers follow the advice of King (1998) and use the invariance property of maximum likelihood estimates to quickly convert the coefficients into the quantity of interest. These approaches are not equivalent, yet researchers seem to give little thought to the choice. I show that the average of simulations can introduce substantial bias compared to the maximum likelihood estimate. In general, when reporting point estimates of the quantity of interest, researchers should report the maximum likelihood estimate, not the average of the simulations.

Suppose the researcher uses maximum likelihood to estimate a statistical model in which $y_i \sim f(\theta_i)$, where $i \in \{1, ..., N\}$ and f represents a probability distribution. The parameter θ_i is connected to a design matrix X of k explanatory variables and a column of ones by a link function g, so that $g(\theta_i) = X_i \beta$, where $\beta \in \mathbb{R}^N$ represents a vector of coefficients with length $k + 1$. The researcher can use maximum likelihood to compute estimates $\hat{\beta}^{\text{mle}}$ for the parameter vector β.[1] Note that I adopt a frequentist perspective, so $\hat{\beta}^{\text{mle}}$ represents a *random variable* that varies across samples.

But researchers usually care about a function τ of the model coefficients β rather than the model coefficients themselves. Following King, Tomz, and Wittenberg (2000), I refer to this function τ as the "quantity of interest." The researcher can use the invariance property $\hat{\tau}^{\text{mle}} = \tau\left(\hat{\beta}^{\text{mle}}\right)$ to quickly obtain a maximum likelihood estimate of the quantity of interest (King 1998, pp. 75-76, and Casella and Berger 2002, pp. 320-321).

The Cost of Transformation

The invariance property does not come without cost. The transformation of unbiased model coefficient estimates introduces bias into the estimate of the quantity of interest. If the coefficient estimates are biased, the transformation-induced bias can, but generally does not, offset the bias in the coefficient estimates. To separate the sources of bias, Rainey (2017, p. 404) decomposes the bias in the estimate of the quantity of interest, which he refers to as total τ-bias, into two components: transformation-induced τ-bias and coefficient-induced τ-bias. These are defined as

$$\text{total } \tau\text{-bias} = \underbrace{\text{E}[\tau(\hat{\beta}^{\text{mle}})] - \tau[\text{E}(\hat{\beta}^{\text{mle}})]}_{\text{transformation-induced}} + \overbrace{\tau[\text{E}(\hat{\beta}^{\text{mle}})] - \tau(\beta)}^{\text{coefficient-induced}}$$

The direction and magnitude of the coefficient-induced τ-bias depends on the choice of τ and the bias in the coefficient estimates, but an unbiased estimator $\hat{\beta}$ implies no coefficient-induced τ-bias. I do not consider

[1]I do not consider other estimators of β. I superscript the estimate with "mle" to emphasize that $\hat{\beta}^{\text{mle}}$ represents the *maximum likelihood* estimate.

coefficient-induced τ-bias any further.

Transformation-induced τ-bias, though, can be understood based on the shape of the transformation. In general, any strictly convex (concave) τ creates upward (downward) transformation-induced τ-bias.

The Average of Simulations

Some commonly used software, such as Stata's `margins` command (and Leeper's `margins` port into R), use the invariance property to estimate the quantity of interest.

But other software, such as `CLARIFY` for Stata and `Zelig` for R, adopts an alternative approach recommended by King, Tomz, and Wittenberg (2000). King, Tomz, and Wittenberg (2000) suggest the following algorithm:

1. *Fit the model.* Use maximum likelihood to estimate the model coefficients $\hat{\beta}^{\mathrm{mle}}$ and their covariance $\hat{V}\left(\hat{\beta}^{\mathrm{mle}}\right)$.
2. *Simulate the coefficients.* Simulate a large number M of coefficient vectors $\tilde{\beta}^{(i)}$ for $i \in \{1, 2, ..., M\}$ using $\tilde{\beta}^{(i)} \sim N\left[\hat{\beta}^{\mathrm{mle}}, \hat{V}\left(\hat{\beta}^{\mathrm{mle}}\right)\right]$, where N is the multivariate normal distribution.
3. *Convert simulated coefficients to simulated quantity of interest.* Compute $\tilde{\tau}^{(i)} = \tau\left(\tilde{\beta}^{(i)}\right)$ for $i \in \{1, 2, ..., M\}$. Most quantities of interest depend on the values of the explanatory variables. In this case, the researcher must choose to focus on a particular scenario, or perhaps average across several scenarios (Hanmer and Kalkan 2013). In any case, the transformation τ includes this choice.[2]
4. *Average the simulations of the quantity of interest.* Estimate the quantity of interest using the average of the simulations of the quantity of interest, so that $\hat{\tau}^{\mathrm{avg.}} = \frac{1}{M} \sum_{i=1}^{M} \tilde{\tau}^{(i)}$.

In the discussion that follows, I assume no Monte Carlo error exists in $\hat{\tau}^{\mathrm{avg.}}$. That is, make M large enough to assume that $\hat{\tau}^{\mathrm{avg.}} = \mathrm{E}\left[\tau\left(\tilde{\beta}\right)\right]$, where $\tilde{\beta} \sim N\left[\hat{\beta}^{\mathrm{mle}}, \hat{V}\left(\hat{\beta}^{\mathrm{mle}}\right)\right]$.

One might use alternative approach to generate $\tilde{\beta}$, such as parametric or non-parametric bootstrap. As long as $\tilde{\beta}$ approximately follows a normal distribution, then the intuitions and best practices I offer below still apply.

The Average of Simulations Versus the Maximum Likelihood Estimate

Applied researchers seem to use $\hat{\tau}^{\mathrm{avg.}}$ and $\hat{\tau}^{\mathrm{mle}}$ interchangeably. But the preceding discussion raises questions. How does $\hat{\tau}^{\mathrm{avg.}}$ compare to $\hat{\tau}^{\mathrm{mle}}$? Are they the same? How are they different? Is one more biased than the other?

If the transformation is always convex (or always concave), then Jensen's inequality allows a simple statement about the relationship between the average of simulations and the maximum likelihood estimate given in Lemma 1.

Lemma 1 *Suppose a maximum likelihood estimator $\hat{\beta}^{mle}$. Then any strictly convex (concave) τ guarantees that $\hat{\tau}^{avg.}$ is strictly greater [less] than $\hat{\tau}^{mle}$.*

Proof By definition,
$$\hat{\tau}^{\mathrm{avg.}} = \mathrm{E}\left[\tau\left(\tilde{\beta}\right)\right].$$

Using Jensen's inequality (Casella and Berger 2002, p. 190, Thm. 4.7.7), we know that $\mathrm{E}\left[\tau\left(\tilde{\beta}\right)\right] > \tau\left[\mathrm{E}\left(\tilde{\beta}\right)\right]$, so that
$$\hat{\tau}^{\mathrm{avg.}} > \tau\left[\mathrm{E}\left(\tilde{\beta}\right)\right].$$

---

[2]As King et al. note, this step might require additional simulation, but most times not. My arguments do not depend on this simplification.

However, because $\tilde{\beta} \sim N\left[\hat{\beta}^{\text{mle}}, \hat{V}\left(\hat{\beta}^{\text{mle}}\right)\right]$, $\text{E}\left(\tilde{\beta}\right) = \hat{\beta}^{\text{mle}}$, so that

$$\hat{\tau}^{\text{avg.}} > \tau\left(\hat{\beta}^{\text{mle}}\right).$$

Of course, $\hat{\tau}^{\text{mle}} = \tau\left(\hat{\beta}^{\text{mle}}\right)$ by definition, so that

$$\hat{\tau}^{\text{avg.}} > \hat{\tau}^{\text{mle}}.$$

The proof for concave τ follows similarly. ∎

This result is intuitive. By assumption, $\tilde{\beta}$ has a symmetric distribution. By definition, $\hat{\tau}^{\text{mle}}$ simply equals the mode of the distribution of $\tau(\tilde{\beta})$. But the distribution of $\tau(\tilde{\beta})$ is *not* symmetric. If $\tilde{\beta}$ happens to fall below the mode $\hat{\beta}^{\text{mle}}$, then τ pulls $\tau(\tilde{\beta})$ in toward $\hat{\tau}^{\text{mle}}$. If $\tilde{\beta}$ happens to fall above the mode $\hat{\beta}^{\text{mle}}$, then τ pushes $\tau(\tilde{\beta})$ away from $\hat{\tau}^{\text{mle}}$. This creates a right-skewed distribution for $\tau(\tilde{\beta})$, which pushes the average $\hat{\tau}^{\text{avg.}}$ above $\hat{\tau}^{\text{mle}}$.

For a convex transformation, Lemma 1 shows that $\hat{\tau}^{\text{avg.}}$ is always larger than $\hat{\tau}^{\text{mle}}$. But does this imply that $\hat{\tau}^{\text{avg.}}$ is *more biased* than $\hat{\tau}^{\text{mle}}$? Theorem 1 shows this is the case.

Theorem 1 *Suppose a maximum likelihood estimator $\hat{\beta}^{mle}$. Then for any strictly convex or concave τ, the transformation-induced τ-bias for $\hat{\tau}^{avg.}$ is strictly greater in magnitude than the transformation-induced τ-bias for $\hat{\tau}^{mle}$.*

Proof According to Theorem 1 of Rainey (2017, p. 405), $\text{E}\left(\hat{\tau}^{\text{mle}}\right) - \tau\left[\text{E}\left(\hat{\beta}^{\text{mle}}\right)\right] > 0$. Lemma 1 shows that for any convex τ, $\hat{\tau}^{\text{avg.}} > \hat{\tau}^{\text{mle}}$. It follows that $\underbrace{\text{E}(\hat{\tau}^{\text{avg.}}) - \tau\left[\text{E}\left(\hat{\beta}^{\text{mle}}\right)\right]}_{\text{t.i. } \tau\text{-bias in } \hat{\tau}^{\text{avg.}}} > \underbrace{\text{E}\left(\hat{\tau}^{\text{mle}}\right) - \tau\left[\text{E}\left(\hat{\beta}^{\text{mle}}\right)\right]}_{\text{t.i. } \tau\text{-bias in } \hat{\tau}^{\text{mle}}} > 0$. For the concave case, it follows similarly that $\underbrace{\text{E}(\hat{\tau}^{\text{avg.}}) - \tau\left[\text{E}\left(\hat{\beta}^{\text{mle}}\right)\right]}_{\text{t.i. } \tau\text{-bias in } \hat{\tau}^{\text{avg.}}} < \underbrace{\text{E}\left(\hat{\tau}^{\text{mle}}\right) - \tau\left[\text{E}\left(\hat{\beta}^{\text{mle}}\right)\right]}_{\text{t.i. } \tau\text{-bias in } \hat{\tau}^{\text{mle}}} < 0$ ∎

Regardless of whether the transformation-induced τ-bias is positive or negative, Theorem 1 shows that the magnitude of the bias is *always* larger for $\hat{\tau}^{\text{avg.}}$ than $\hat{\tau}^{\text{mle}}$.

An Approximation for the Additional Bias in $\hat{\tau}^{\text{avg.}}$

Theorem 1 guarantees that $\hat{\tau}^{\text{avg.}}$ is more biased than $\hat{\tau}^{\text{mle}}$. This raises yet more questions. By how much? Is the bias trivial? Or is it substantial? Monte Carlo experiments allow one to assess this directly, but an analytical approximation provides a helpful rule of thumb.

I approximate the *additional* transformation-induced τ-bias in $\hat{\tau}^{\text{avg.}}$ compared to $\hat{\tau}^{\text{mle}}$ as

$$
\begin{aligned}
\text{additional t.i. } \tau\text{-bias in } \hat{\tau}^{\text{avg.}} &= \underbrace{\left(\mathrm{E}(\hat{\tau}^{\text{avg.}}) - \tau\left[\mathrm{E}(\hat{\beta}^{\text{mle}})\right]\right)}_{\text{t.i. } \tau\text{-bias in } \hat{\tau}^{\text{avg.}}} - \underbrace{\left(\mathrm{E}(\hat{\tau}^{\text{mle}}) - \tau\left[\mathrm{E}(\hat{\beta}^{\text{mle}})\right]\right)}_{\text{t.i. } \tau\text{-bias in } \hat{\tau}^{\text{mle}}} \\
&= \mathrm{E}(\hat{\tau}^{\text{avg.}}) - \mathrm{E}(\hat{\tau}^{\text{mle}}) \\
&= \mathrm{E}\left(\hat{\tau}^{\text{avg.}} - \hat{\tau}^{\text{mle}}\right) \\
&= \mathrm{E}\left(\mathrm{E}\left[\tau(\tilde{\beta})\right] - \tau(\hat{\beta}^{\text{mle}})\right) \\
&= \mathrm{E}\left(\underbrace{\mathrm{E}\left[\tau(\tilde{\beta})\right] - \tau\left[\mathrm{E}(\tilde{\beta})\right]}_{\substack{\text{approximated in Eq. 1,} \\ \text{p. 405, of Rainey (2017)}}} \right) \\
&\approx \mathrm{E}\left[\frac{1}{2}\sum_{r=1}^{k+1}\sum_{s=1}^{k+1} H_{rs}(\hat{\beta}^{\text{mle}})\,\hat{V}_{rs}(\hat{\beta}^{\text{mle}}) \right]
\end{aligned}
\tag{1}
$$

where the remaining expectation occurs with respect to $\hat{\beta}^{\text{mle}}$, $H(\hat{\beta}^{\text{mle}})$ represents the Hessian matrix of second derivatives of τ at the point $\hat{\beta}^{\text{mle}}$ and, conveniently, $\hat{V}(\hat{\beta}^{\text{mle}})$ represents the estimated covariance matrix $\hat{\beta}^{\text{mle}}$.

This is approximation is similar to the approximation for the transformation-induced τ-bias for $\hat{\beta}^{\text{mle}}$, which adjusting notation slightly, Rainey (2017, p. 405, Eq. 1) computes as

$$
\text{t.i. } \tau\text{-bias for } \hat{\beta}^{\text{mle}} = \frac{1}{2}\sum_{r=1}^{k+1}\sum_{s=1}^{k+1} H_{rs}\left[\mathrm{E}(\hat{\beta}^{\text{mle}})\right] V_{rs}(\hat{\beta}^{\text{mle}}),
\tag{2}
$$

where $H\left[\mathrm{E}(\hat{\beta}^{\text{mle}})\right]$ represents the Hessian matrix of second derivatives of τ at the point $\mathrm{E}(\hat{\beta}^{\text{mle}})$ and $V(\hat{\beta}^{\text{mle}})$ represents the covariance matrix of the sampling distribution of $\hat{\beta}^{\text{mle}}$.

When we compare Equations 1 and 2, we are yet again comparing the *average of a function* with the *function of that average*. Therefore, Equations 1 and 2 are not exactly equal. But, as a rule of thumb, we should expect them to be similar. And to the extent that this rule holds, the *additional* transformation-induced τ-bias in $\hat{\tau}^{\text{avg.}}$ is about the same as the transformation-induced τ-bias in $\hat{\tau}^{\text{mle}}$. Therefore, as a rule of thumb, I suggest that the transformation-induced τ-bias in $\hat{\tau}^{\text{avg.}}$ will be about *double* the transformation-induced τ-bias in $\hat{\tau}^{\text{mle}}$.[3]

Because of the similarity in Equations 1 and 2, the difference between $\hat{\tau}^{\text{avg.}}$ and $\hat{\tau}^{\text{mle}}$ becomes large under the same conditions that Rainey (2017) notes transformation-induced τ-bias becomes large: when the non-linearity in the transformation in severe and when the standard errors of $\hat{\beta}^{\text{mle}}$ are large.

Illustrative Simulation

As an illustration, consider the Poisson regression model $y_i \sim \text{Poisson}(\lambda_i)$, where $\lambda_i = e^{(-2+x_i)}$ for $i \in \{1, 2, ..., 100\}$. To create the x_is, I simply took 100 draws from a standard normal distribution. Assume that the researcher wants to estimate the instantaneous marginal effect of x on $\mathrm{E}(y)$ so that $\tau(\beta) = \frac{d\,\mathrm{E}(y)}{dx} = e^{(\beta_{cons}+\beta_x x)}$ for x ranging from

---

[3]When writing Rainey (2017), I considered using the difference between $\hat{\tau}^{\text{avg.}}$ and $\hat{\tau}^{\text{mle}}$ to estimate (and correct for) the transformation-induced τ-bias in $\hat{\tau}^{\text{mle}}$. This crude form of parametric bootstrap does a reasonable job of reducing the bias, but tends to increase the MSE of the estimated quantity of interest.

about -3 to about +3.

Following the procedures discussed above, I generate 1,000 data sets and use each data set to estimate $\hat{\tau}^{\mathrm{mle}}$ and $\hat{\tau}^{\mathrm{avg.}}$. Note that the transformation using is convex, so Theorem 1 guarantees that the transformation-induced τ-bias in both $\hat{\tau}^{\mathrm{mle}}$ and $\hat{\tau}^{\mathrm{avg.}}$ will be positive. The rule of thumb suggests about twice as much bias in $\hat{\tau}^{\mathrm{avg.}}$ as in $\hat{\tau}^{\mathrm{mle}}$.

Figure 1 shows the bias in $\hat{\tau}^{\mathrm{avg.}}$ and $\hat{\tau}^{\mathrm{mle}}$ compared to the true value $\tau(\beta)$. Especially notice three features of this plot. First, the bias is substantial. For $\hat{\tau}^{\mathrm{mle}}$, the bias is about $\frac{1}{8}$ the size of the true effect. For $\hat{\tau}^{\mathrm{avg.}}$, the bias is about $\frac{1}{4}$ the size of the true effect. Second, notice that the bias occurs in the expected direction. Because the transformation $\tau(\beta) = \frac{d\,\mathrm{E}(y)}{dx} = e^{(\beta_{cons}+\beta_x x)}$ is convex, the bias is positive. Upon careful inspection of Figure 1, you might notice that the bias is slightly *negative* in the interval from 0 to $\frac{1}{4}$. In this region, the negative coefficient-induced τ-bias (Poisson regression coefficients are biased toward zero) happens to be larger than the positive transformation-induced τ-bias. Theorem 1 only addresses transformation-induced τ-bias. Third, notice that the bias in $\hat{\tau}^{\mathrm{avg.}}$ is about twice as large as the bias in $\hat{\tau}^{\mathrm{mle}}$, as the rule of thumb suggests.
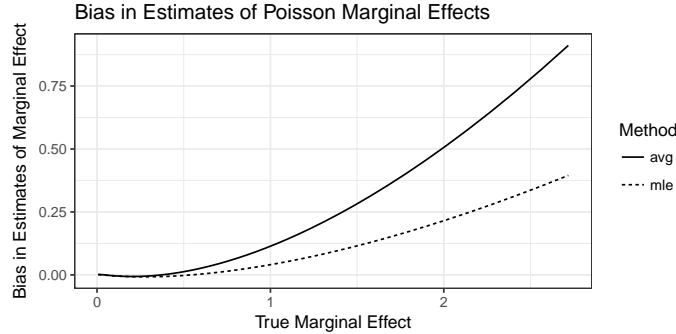

Figure 1: This figure shows the bias in the estimates of the marginal effects in a Poisson regression model. Notice that the convex transformation $\tau(\beta) = \frac{d\,\mathrm{E}(y)}{dx} = e^{(\beta_{cons}+\beta_x x)}$ creates a positive bias (see Theorem 1) and that the bias in $\hat{\tau}^{\mathrm{avg.}}$ is about twice as large as the bias in $\hat{\tau}^{\mathrm{mle}}$ (compare Equations 1 and 2).

Example: Supreme Court Decisions

To unify explanation of U.S. Supreme Court decisions, George and Epstein (1992) fit a single probit model that combines the legal and extralegal models of Court decision-making to a data set of 64 decisions. The authors model the probability of a conservative decision as a function of whether the Solicitor General filed an Amicus brief (SG = 1) or not (SG = 0) and 10 other explanatory variables. See George and Epstein (1992) for the more details of the model.

I use this model illustrate the potential impact of using the simulation average rather than the maximum likelihood estimate of the quantity of interest. I focus on two potential quantities of interest: the probability of a conservative decision and the effect of the Solicitor General filing a brief. Table 1 summarizes these quantities of interest.

For each quantity of interest, I compute an estimate using the average of simulation and maximum likelihood. First, I use both the average of simulations and maximum likelihood to estimate the probability of a conservative decision for each combination of explanatory variables included in the data set. Second, I use both approaches to

Table 1: This table provides the details of the quantities of interest from George and Epstein's (1992) model of U.S. Supreme Court decisions.

Description	Notation	Change in Key Explanatory Variable	Values for Other Explanatory Variables
probability of a conservative decision	$\tau(\beta) = \Phi(X_c\beta)$	none	every observed combination
effect of a Solicitor General brief on the probability of a conservative decision	$\tau(\beta) = \Phi(X_{\text{high}}\beta) - \Phi(X_{\text{low}}\beta)$	for X_{high}, SG = 1, and for X_{low}, SG = 0	every observed combination

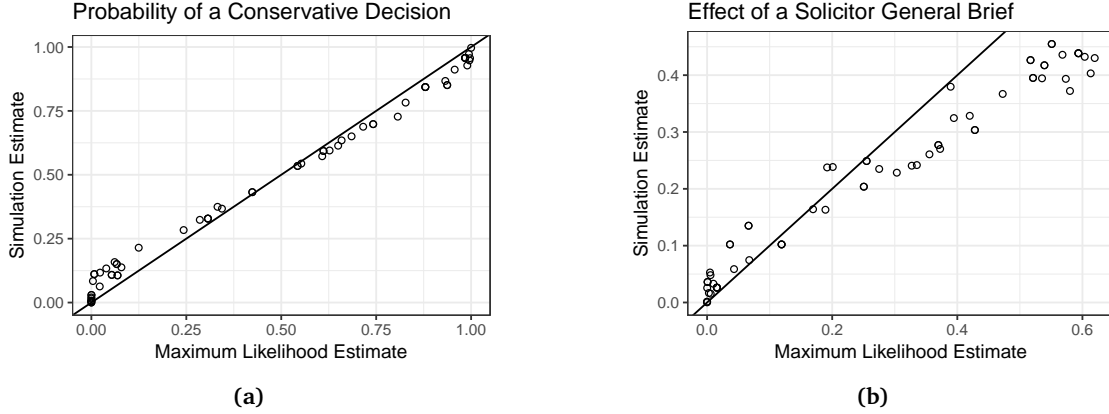**(a)**   **(b)**

Figure 2: The figure shows the relationship between the simulation average and the maximum likelihood estimate two the quantities of interest. The left panel (a) shows the probability of a conservative decision. Notice that the simulation average tends falls about the maximum likelihood estimate when the probability is low–where the transformation is convex–and below the maximum likelihood estimate when the probability is high–where the transformation is concave. The right panel (b) shows the effect of a brief by the Solicitor General on the probability of a conservative decision.

estimate the effect of a Solicitor General brief on the probability of a conservative decision. I define this effect as the *difference* in the probability of a conservative decision for each observation in the data set, if that observation changed from one in which the Solicitor General *did not* file a brief (SG = 0) to one in which the Solicitor General *did* file a brief (SG = 1).

Figure 2 compares the estimates. First, consider the estimates of the probability of a conservative decision in Figure 2a. The pattern is clear: when the chance of a conservative decision is less than 50%, the average of the simulations is too large. In this region, the transformation (the normal cdf) is convex. When the chance of a conservative decision is greater than 50%, the average of the simulations is too small. In this region, the transformation is concave. When the chance of a conservative decision is closer to 50%, the differences between the average of the simulations and the maximum likelihood estimate are smaller, because the transformation is more linear in this area. The same is true for chances close to 0% and 100%.

Further, some of the differences are quite large. For example, when maximum likelihood suggests a chance of about 5%, the average of the simulation suggests a chance of about 10%. This difference may seem small at first (i.e., only 5 percentage points), but the average of simulations is about *double* the maximum likelihood estimate.

Now consider the estimates of the effect of the Solicitor General filing an Amicus brief in Figure 2b. The largest differences appear in the upper-right corner of the plot. For this group of observation, the average of simulations suggests than a brief from the Solicitor General increases the chance of a conservative decision by about 40 percentage points. On the other hand, the maximum likelihood estimate suggests an increase of about 60 percentage points. This difference is certainly meaningful, the maximum likelihood estimate is 50% larger than

the average of the simulations.

A Note on Hanmer and Kalkan (2013)

Hanmer and Kalkan (2013) discuss two approaches to computing quantities of interest: the typical "average-case" approach and their recommended "observed-value" approach. With either approach, the researcher estimates the quantity of interest–the change in the expected value of the outcome variable (e.g., the probability of a conservative decision) as a key explanatory variable changes from a low value to a high value (e.g., SG changes from 0 to 1). But the researcher must also deal with the other explanatory variables in the model, because these variables alter the effect of interest. The average-case approach sets the other explanatory variables at a typical value, such as the median. Hanmer and Kalkan (2013) suggest estimating the quantity of interest for all the observed combinations of the other explanatory variables and then averaging the quantity of interest across the combinations. Hanmer and Kalkan (2013) In my argument, this choice is built into the transformation τ, so their (compelling) argument does not undermine or enhance my own.[4] Because researchers have not drawn a sharp conceptual distinction between the simulation average and maximum likelihood estimates of the quantity of interest, Hanmer and Kalkan (2013) do not adopt a clear position. I read their paper to suggest averaging the maximum likelihood estimates ($\hat{\tau}^{\text{mle}}$) for each observed case (though they do not explicit write this.) However, in their Stata code, they average the simulation averages ($\hat{\tau}^{\text{avg.}}$) for each observed case.

When I average the maximum likelihood estimates of the effect of a Solicitor General brief in the left panel (b) of Figure 2 (as Hanmer and Kalkan (2013) seem to suggest in their paper), I obtain an estimated effect of 0.28. When I average the simulation averages (as Hanmer and Kalkan do in their Stata code), I obtain an estimated effect of 0.23. The maximum likelihood estimate is about 22% larger than the simulation average–the choice matters.

The important point is this: Hanmer and Kalkan (2013) draw a distinction between the average-case and observed-value approaches to choosing the quantity of interest. This paper draws a distinction between estimating quantities of interest (whether average-case or observed-value) using the simulation average and the maximum likelihood estimate. Regardless of whether the researcher uses the average-case approach or the observed-value approach, the simulation average is more biased than the maximum likelihood estimate.

Conclusion

Substantive researchers in political science tend to estimate their quantity of interest using the average of simulation (e.g., CLARIFY, Zelig) or using the invariance property of maximum likelihood estimates (margins in Stata and R). The choice between the two seems idiosyncratic rather than principled. However, the choice is clear. Rainey (2017) introduces the idea of transformation-induced bias. This paper shows that the average of the simulation roughly *doubles* this bias. In many cases, the researcher has small standard errors. In this case, the additional bias is minimal. But in other cases, such as when the researcher has large standard errors or focuses on a highly non-linear transformation of the model coefficients, the bias is substantial. And the fix is easily. Simply plugging the coefficient estimates into the transformation. Our software should do this by default.

---

[4] I generally agree with their arguments in favor of the observed-value approach, but I recommend researchers plot the distribution of effects rather than summarizing them into a single average.

References

Casella, George, and Roger L. Berger. 2002. *Statistical Inference*. 2nd ed. Pacific Grove, CA: Duxbury.

George, Tracey E., and Lee Epstein. 1992. "On the Nature of Supreme Court Decision Making." *American Political Science Review* 86(2):323–337.

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.

King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: Michigan University Press.

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):341–355.

Rainey, Carlisle. 2017. "Transformation-Induced Bias: Unbiased Coefficients Do Not Imply Unbiased Quantities of Interest." *Political Analysis* 25:402–409.