Simulation-Induced Bias in Quantities of Interest*

Carlisle Rainey†

Holger L. Kern‡

November 1, 2019

Abstract

Researchers commonly obtain point estimates of quantities of interest by simulating model coefficients, transforming these simulated coefficients into simulated quantities of interest, and then taking the average of the simulated quantities of interest. In contrast, other researchers directly transform coefficient estimates into estimated quantities of interest using the invariance property of maximum likelihood estimators. These two approaches are not equivalent. We show that averaging simulated quantities of interest approximately doubles the transformation-induced bias of Rainey (2017).

Political scientists employ maximum likelihood (ML) to estimate a variety of statistical models. ML estimators have desirable and widely understood properties. But for many research questions, the model coefficient estimates are not the quantities of greatest interest to the researcher. Following King, Tomz, and Wittenberg (2000), researchers often focus on substantively meaningful quantities of interest, such as predicted probabilities, expected counts, marginal effects, and first differences. The literature offers two methods to compute these quantities of interest.

First, the literature suggests that researchers use the invariance property of ML estimators to find the ML estimates of the quantities of interest. First, use ML to estimate the model coefficients. Then transform these point estimates of the coefficients directly into the point estimates of the quantities of interest (King 1998, pp. 75–76; Casella and Berger 2002, pp. 320–321). We refer to this approach as the "plug-in" estimator. Herron (1999) recommends using the invariance property. Software packages such as the `margins` command in Stata (StataCorp 2017), the margins package in R (Leeper 2018), and the `predict()` function in R for the glm (R Core Team 2018) and polr (Venables and Ripley 2002) classes implement this approach.

Second, the literature suggests that researchers use the average of the simulated quantities of interest as the point estimator. We refer to this estimator as the "average-of-simulations" estimator.

*All computer code necessary for replication is available on GitHub. We thank Bill Berry, Christopher Gandrud, Michael Hanmer, John Holbein, Gary King, Justin Kirkland, Thomas Leeper, Matt Pietryka, Arthur Spirling, Michael Tomz, Jason Wittenberg, and Chris Wlezien for helpful comments. We also thank audiences at Florida State University and the 2018 Texas Methods Meeting for productive discussions. All remaining errors are our own.

†Carlisle Rainey is Associate Professor of Political Science, Florida State University, 540 Bellamy, Tallahassee, FL, 32306. (crainey@fsu.edu).

‡Holger L. Kern is Assistant Professor of Political Science, Florida State University, 541 Bellamy, Tallahassee, FL, 32306. (hkern@fsu.edu).

King, Tomz, and Wittenberg (2000) suggest this approach. Software packages such as CLARIFY for Stata (Tomz, Wittenberg, and King 2003) and Zelig for R (Imai, King, and Lau 2008; Choirat et al. 2018) implement this approach.

Rainey (2017) introduces the concept of "transformation-induced bias" and shows that the plug-in approach leads to meaningful bias in the point estimates of the quantities of interest. But Rainey (2017) does not consider the average-of-simulations estimator. This raises the question: "Does the average-of-simulations estimator suffer the same bias?"

We show that not only does the average-of-simulations estimator suffer from the same bias, but it effectively duplicates the cause of the bias and compounds the transformation-induced bias with a similar, additional bias that we refer to as "simulation-induced bias."

Transformation-Induced Bias

The invariance property of ML estimators allows a researcher to find the ML estimate of a function of a parameter by first using ML to estimate the model parameter and then applying the function to that estimate (or "plugging in the estimate") (King 1998, pp. 75–76; Casella and Berger 2002, pp. 320–321). More formally, suppose a researcher uses ML to estimate a statistical model in which $y_i \sim f(\theta_i)$, where $i \in \{1, \ldots, N\}$ and f represents a probability distribution. The parameter θ_i connects to a design matrix X of k explanatory variables and a column of ones by a link function $g(\cdot)$, so that $g(\theta_i) = X_i\beta$, where $\beta \in \mathbb{R}^{k+1}$ represents a vector of parameter with length $k+1$. The researcher uses ML to compute estimates $\hat{\beta}^{\text{mle}}$ for the parameter vector β. We denote the function that transforms model coefficients into quantities of interest as $\tau(\cdot)$. For example, if the researcher uses a logit model and focuses on the predicted probability for a specific observation X_c, then $\tau(\beta) = \text{logit}^{-1}(X_c\beta) = \dfrac{1}{1+e^{-X_c\beta}}$. The researcher can use the invariance property to quickly obtain a ML estimate of the predicted probability: $\hat{\tau}^{\text{mle}} = \tau\left(\hat{\beta}^{\text{mle}}\right) = \text{logit}^{-1}\left(X_c\hat{\beta}^{\text{mle}}\right) = \dfrac{1}{1+e^{-X_c\hat{\beta}^{\text{mle}}}}$.

As Rainey (2017) shows, using the invariance property to transform unbiased model coefficient estimates can introduce bias into estimated quantities of interest. 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. Rainey (2017) defines these as

$$\text{total } \tau\text{-bias} = \underbrace{\mathrm{E}\left[\tau\left(\hat{\beta}^{\text{mle}}\right)\right] - \tau\left[\mathrm{E}\left(\hat{\beta}^{\text{mle}}\right)\right]}_{\text{transformation-induced}} + \overbrace{\tau\left[\mathrm{E}\left(\hat{\beta}^{\text{mle}}\right)\right] - \tau\left(\beta\right)}^{\text{coefficient-induced}}. \tag{1}$$

Transformation-induced τ-bias behaves systematically. The shape of the transformation $\tau(\cdot)$ determines the direction of the bias. In general, any strictly convex (concave) $\tau(\cdot)$ creates upward (downward) transformation-induced τ-bias.

The direction and magnitude of the coefficient-induced τ-bias depend on the choice of $\tau(\cdot)$

and the bias in the coefficient estimates, but an unbiased estimator $\hat{\beta}^{\text{mle}}$ implies the absence of coefficient-induced τ-bias. We consider coefficient-induced τ-bias only briefly in the section "**??**", but set it aside for the bulk of our analysis. Instead, we extend the concept of transformation-induced bias from ML estimates to the average of simulations.

The Definition of Simulation-Induced Bias

Rather than rely on the invariance property of ML estimators to compute a point estimate for a quantity of interest, King, Tomz, and Wittenberg (2000) suggests the following simulation-based approach:

1. *Fit the model.* Use ML to estimate the model coefficients $\hat{\beta}^{\text{mle}}$ and their covariance $\hat{V}\left(\hat{\beta}^{\text{mle}}\right)$.
2. *Simulate the coefficients.* Simulate a large number M of coefficient vectors $\tilde{\beta}^{(i)}$, for $i \in \{1, 2, \ldots, M\}$, using $\tilde{\beta}^{(i)} \sim \text{MVN}\left[\hat{\beta}^{\text{mle}}, \hat{V}\left(\hat{\beta}^{\text{mle}}\right)\right]$, where MVN represents the multivariate normal distribution.
3. *Convert simulated coefficients into simulated quantity of interest.* Compute $\tilde{\tau}^{(i)} = \tau\left(\tilde{\beta}^{(i)}\right)$ for $i \in \{1, 2, \ldots, M\}$. Most quantities of interest depend on the values of the explanatory variables. In this situation, researchers either focus on a specific observation (typically some kind of "average case") or average across all sample observations (Hanmer and Kalkan 2013). In any case, the transformation $\tau(\cdot)$ includes this choice.[1]
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 $\hat{\tau}^{\text{avg}} = \frac{1}{M}\sum_{i=1}^{M} \tilde{\tau}^{(i)}$.[2]

The literature offers two ways to estimate quantities of interest: (1) calculate the ML estimator $\hat{\tau}^{\text{mle}}$ using the invariance property of ML estimators or (2) calculate the average of simulations $\hat{\tau}^{\text{avg}}$ using the algorithm described in King, Tomz, and Wittenberg (2000). How does $\hat{\tau}^{\text{avg}}$ compare to $\hat{\tau}^{\text{mle}}$?

If the transformation of coefficient estimates into an estimate of the quantity of interest is always convex (or always concave), then Jensen's inequality allows the simple statement relating $\hat{\tau}^{\text{avg}}$ and $\hat{\tau}^{\text{mle}}$ given in Lemma 1 .

Lemma 1 *Suppose a nondegenerate ML estimator $\hat{\beta}^{mle}$. Then any strictly convex (concave) $\tau(\cdot)$ guarantees that $\hat{\tau}^{avg}$ is strictly greater (less) than $\hat{\tau}^{mle}$.*

This result is intuitive, but see the Appendix for the proof. Since we simulate using a multivariate normal distribution, $\tilde{\beta}$ has a symmetric distribution. By definition, $\hat{\tau}^{\text{mle}}$ equals the mode of the

[1] As King, Tomz, and Wittenberg (2000) note, this step might require additional simulation, to first introduce and then average over fundamental uncertainty. We ignore this additional step since it does not affect our argument.

[2] In the discussion that follows, we assume no Monte Carlo error exists in $\hat{\tau}^{\text{avg}}$. In other words, we assume that M is sufficiently large so that $\hat{\tau}^{\text{avg}} = \text{E}\left[\tau\left(\tilde{\beta}\right)\right]$, where $\tilde{\beta} \sim \text{MVN}\left[\hat{\beta}^{\text{mle}}, \hat{V}\left(\hat{\beta}^{\text{mle}}\right)\right]$.

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 $\tau(\cdot)$ pulls $\tau(\tilde{\beta})$ in toward $\hat{\tau}^{\text{mle}}$. If $\tilde{\beta}$ happens to fall above the mode $\hat{\beta}^{\text{mle}}$, then $\tau(\cdot)$ 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 $\hat{\tau}^{\text{mle}}$. We refer to the expectation of this difference between $\hat{\tau}^{\text{avg}}$ and $\hat{\tau}^{\text{mle}}$ as "simulation-induced bias," so that

$$\text{simulation-induced } \tau\text{-bias} = \mathrm{E}\left(\hat{\tau}^{\text{avg}}\right) - \mathrm{E}\left(\hat{\tau}^{\text{mle}}\right).$$

Theorem 1 compares the sum of simulation- and transformation-induced τ-bias in $\hat{\tau}^{\text{avg}}$ to transformation-induced τ-bias in $\hat{\tau}^{\text{avg}}$.

Theorem 1 *Suppose a nondegenerate ML estimator $\hat{\beta}^{mle}$. Then for any strictly convex or concave $\tau(\cdot)$, the sum of the simulation-induced and transformation-induced τ-bias in $\hat{\tau}^{avg}$ is strictly greater in magnitude than the transformation-induced τ-bias in $\hat{\tau}^{mle}$.*

Regardless of the direction the simulation-induced and transformation-induced τ-bias, Theorem 1 shows that the magnitude of the combination in $\hat{\tau}^{\text{avg}}$ is *always* larger than the transformation-induced bias alone in $\hat{\tau}^{\text{mle}}$ for strictly convex or concave $\tau(\cdot)$.

The Intuition of Simulation-Induced Bias

Theorem 1 shows that $\hat{\tau}^{\text{avg}}$ compounds both transformation-induced τ-bias with simulation-induced τ-bias. But is this bias substantively important? Monte Carlo experiments allow us to assess this directly, but an analytical approximation provides a helpful guideline.

We approximate the simulation-induced τ-bias in $\hat{\tau}^{\text{avg}}$ as

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

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

This approximation appears similar to the approximation for the transformation-induced τ-bias, which (adjusting notation slightly) Rainey (2017, p. 405, Eq. 1) computes as

$$\text{t.i. } \tau\text{-bias} \approx \frac{1}{2}\sum_{r=1}^{k+1}\sum_{s=1}^{k+1} H_{rs}\left[\mathrm{E}\left(\hat{\beta}^{\text{mle}}\right)\right] V_{rs}\left(\hat{\beta}^{\text{mle}}\right), \qquad (3)$$

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

When we compare Equations 2 and 3, we yet again compare the *expectation of a function* with the *function of the expectation*. Therefore, Equations 2 and 3 are not exactly equal. But, as a rough guideline, we should expect them to be similar. And to the extent that the two are similar, the additional simulation-induced τ-bias in $\hat{\tau}^{\text{avg}}$ is about the same as the transformation-induced τ-bias in $\hat{\tau}^{\text{mle}}$.

Because of the similarity between Equations 2 and 3, the simulation-induced τ-bias becomes large under the conditions identified by Rainey (2017) as leading to large transformation-induced τ-bias: when the non-linearity in the transformation $\tau(\cdot)$ is severe and when the standard errors of $\hat{\beta}^{\text{mle}}$ are large. While the transformation-induced τ-bias vanishes as the number of observations grows large, it can be substantively meaningful for the sample sizes commonly encountered in social science research (Rainey 2017).

The Intuition of Simulation-Induced Bias

To develop the intuition for the theoretical results above, we example a stylized example with simulations, an alterative analytical approach, and an empirical example.

Using a Drastic, Convex Transformation: $\tau(\mu) = \mu^2$

To develop an intuition for the simulation-induced τ-bias in $\hat{\tau}^{\text{avg}}$, consider the simple, unrealistic (but heuristically useful) scenario in which $y_i \sim \text{N}(0, 1)$, for $i \in \{1, 2, \ldots, n = 100\}$, and the researcher wishes to estimate μ^2. Suppose that the researcher knows that the variance equals one but does not know that the mean μ equals zero. The researcher uses the unbiased maximum likelihood estimator $\hat{\mu}^{\text{mle}} = \frac{\sum_{i=1}^{n} y_i}{n}$ of μ, but ultimately cares about the quantity of interest $\tau(\mu) = \mu^2$. The researcher can use the invariance property to compute the ML estimate $\tau(\mu)$ as $\hat{\tau}^{\text{mle}} = \left(\hat{\mu}^{\text{mle}}\right)^2$. Alternatively, the researcher can use the simulation-based approach, estimating $\tau(\mu)$ as $\hat{\tau}^{\text{avg}} = \frac{1}{M}\sum_{i=1}^{M} \tau\left(\tilde{\mu}^{(i)}\right)$, where $\tilde{\mu}^{(i)} \sim \text{N}\left(\hat{\mu}^{\text{mle}}, \frac{1}{\sqrt{n}}\right)$ for $i \in \{1, 2, \ldots, M\}$.

The true value of the quantity of interest is $\tau(0) = 0^2 = 0$. However, the ML estimator $\hat{\tau}^{\text{mle}} = \left(\hat{\mu}^{\text{mle}}\right)^2$ equals zero if and only if $\hat{\mu}^{\text{mle}} = 0$. Otherwise, $\hat{\tau}^{\text{mle}} > 0$. Since $\hat{\mu}^{\text{mle}}$ almost surely differs from zero, $\hat{\tau}^{\text{mle}}$ is biased upward.

Moreover, even if $\hat{\mu}^{\text{mle}} = 0$, $\tilde{\mu}^{(i)}$ almost surely differs from zero. If $\tilde{\mu}^{(i)} \neq 0$, then $\left(\tilde{\mu}^{(i)}\right)^2 > 0$. Thus, $\hat{\mu}^{\text{avg}}$ is almost surely larger than the true value $\tau(\mu) = 0$ even when $\hat{\mu} = 0$.

We illustrate this fact clearly by repeatedly simulating y and computing $\hat{\tau}^{\text{mle}}$ and $\hat{\tau}^{\text{avg}}$. Figure **??** shows the first four of 10,000 total simulations. The figure shows how the unbiased estimate $\hat{\mu}^{\text{mle}}$ is translated into $\hat{\tau}^{\text{mle}}$ and $\hat{\tau}^{\text{avg}}$.

First, to find $\hat{\tau}^{\text{avg}}$, we complete three steps: (1) simulate $\tilde{\mu}^{(i)} \sim \text{N}\left(\hat{\mu}^{\text{mle}}, \frac{1}{10}\right)$ for $i \in \{1, 2, \ldots, M = 1{,}000\}$, (2) calculate $\tilde{\tau}^{(i)} = \tau\left(\tilde{\mu}^{(i)}\right)$, and (3) calculate $\hat{\tau}^{\text{avg}} = \frac{1}{M}\sum_{i=1}^{M} \tilde{\tau}^{(i)}$. The rug plot along the horizontal axis shows the distribution of $\tilde{\mu}$. The hollow points in Figure **??** shows the transformation of each point $\tilde{\mu}^{(i)}$ into $\tilde{\tau}^{(i)}$. The rug plot along the vertical axis shows the distribution of $\tilde{\tau}$. Focus on the top-left panel of Figure **??**. Notice that $\hat{\mu}^{\text{mle}}$ estimates the true value $\mu = 0$ quite well. However, after simulating $\tilde{\mu}$ and transforming $\tilde{\mu}$ into $\tilde{\tau}$, the $\tilde{\tau}$s fall far from the true value $\tau(0) = 0$. The dashed orange line shows the average of $\tilde{\tau}$. Notice that although $\hat{\mu}^{\text{mle}}$ is unusually close to the truth $\mu = 0$ in this sample, $\hat{\tau}^{\text{avg}}$ is substantially biased upward.

Second, to find $\hat{\tau}^{\text{mle}}$, we simply transform $\hat{\mu}^{\text{mle}}$ directly using $\hat{\tau}^{\text{mle}} = \left(\hat{\mu}^{\text{mle}}\right)^2$. The solid green lines show this transformation. The convex transformation $\tau(\cdot)$ has the effect of lengthening the right tail of the distribution of $\tilde{\tau}$, pulling the average well above the mode. This provides the basic intuition for Lemma 1.

The remaining panels of Figure 1 repeat this process with three more random samples. Each

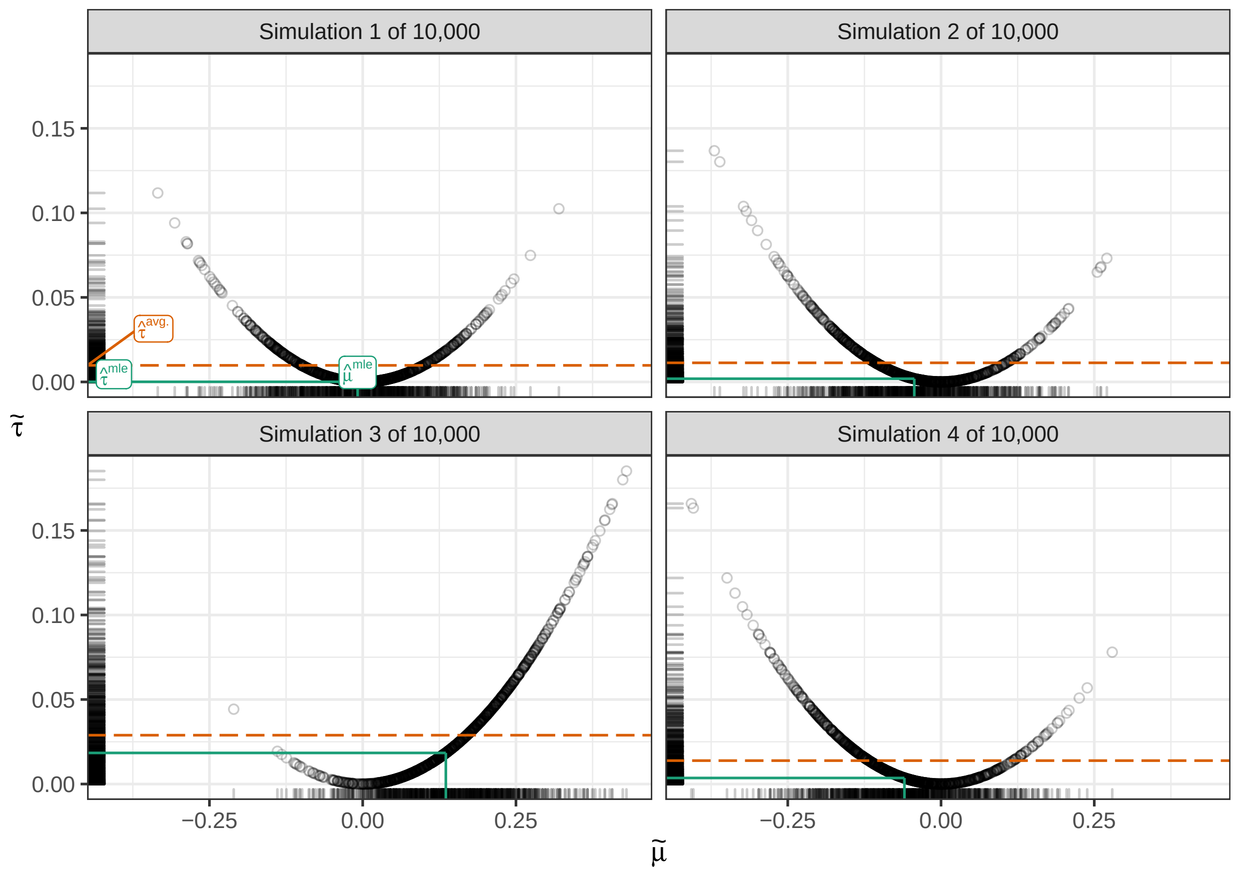

Figure 1: The first four Monte Carlo simulations of $\hat{mu}$. These four panels illustrate the relationship between $\hat{\tau}^{\text{mle}}$ and $\hat{\tau}^{\text{avg}}$ described by Lemma 1 and Theorem 1.

sample presents a similar story — the convex transformation stretches the distribution of $\tilde{\tau}$ to the right, which pulls $\hat{\tau}^{\text{avg}}$ above $\hat{\tau}^{\text{mle}}$.

We repeat this process 10,000 total times to produce 10,000 estimates $\hat{\mu}^{\text{mle}}$, $\hat{\tau}^{\text{mle}}$, and $\hat{\tau}^{\text{avg}}$. Figure 2 shows the density plots for the 10,000 estimates (i.e., the sampling distributions of $\hat{\mu}^{\text{mle}}$, $\hat{\tau}^{\text{mle}}$, and $\hat{\tau}^{\text{avg}}$). As we know analytically, $\hat{\mu}^{\text{mle}}$ is unbiased with a standard error of $\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{100}} = \frac{1}{10}$. Both $\hat{\tau}^{\text{mle}}$ and $\hat{\tau}^{\text{avg}}$ are biased upward, but $\hat{\tau}^{\text{avg}}$ is about twice as biased. Theorem 1 shows why this must be the case.

Using the Law of Iterated Expectations

Alternatively, we can develop the intuition behind our argument analytically via the law of iterated expectations. It helps to alter the notation slightly, making two implicit dependencies explicit. We explain each change below and use the alternate, more expansive notation only in this section.

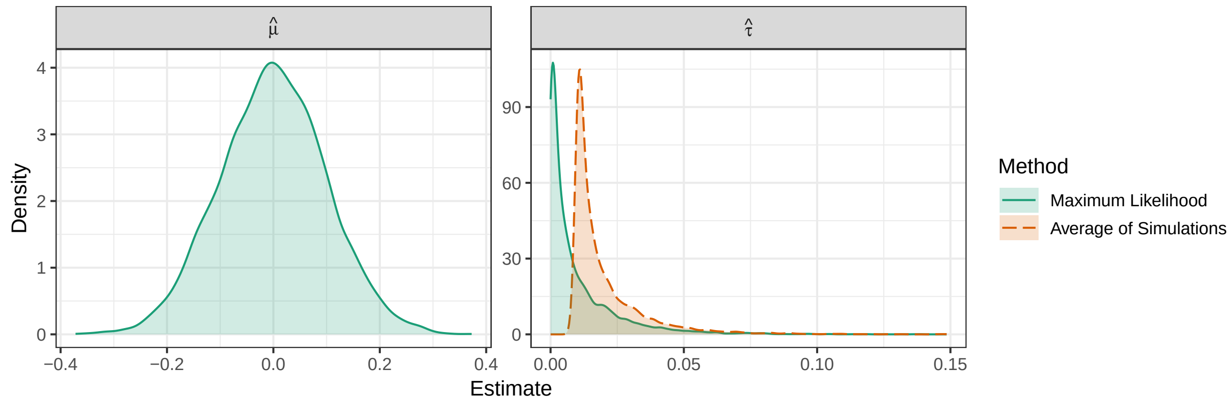

Figure 2: The sampling distributions of $\hat{\beta}^{\text{mle}}$, $\hat{\tau}^{\text{mle}}$, and $\hat{\tau}^{\text{avg}}$.

The law of iterated expectations states that $\mathrm{E}_Y\left(\mathrm{E}_{X|Y}(X \mid Y)\right) = \mathrm{E}_X(X)$, where X and Y represent random variables. The three expectations occur with respect to three different distributions: E_Y denotes the expectation with respect to the marginal distribution of Y, $\mathrm{E}_{X|Y}$ denotes the expectation with respect to the conditional distribution of $X \mid Y$, and E_X denotes the expectation with respect to the marginal distribution of X.

Outside of this section, we understand that the distribution of $\tilde{\beta}$ depends on $\hat{\beta}^{\text{mle}}$ and could be written as $\tilde{\beta} \mid \hat{\beta}^{\text{mle}}$. To remain consistent with previous work, especially King, Tomz, and Wittenberg (2000) and Herron (1999), we use $\tilde{\beta}$ to represent $\tilde{\beta} \mid \hat{\beta}^{\text{mle}}$. The definition of $\tilde{\beta}$ makes this usage clear. In this section only, we use $\tilde{\beta} \mid \hat{\beta}^{\text{mle}}$ to represent the conditional distribution of $\tilde{\beta}$ and use $\tilde{\beta}$ to represent the <u>un</u>conditional distribution of $\tilde{\beta}$. Intuitively, one might imagine (1) generating a data set y, (2) estimating $\hat{\beta}^{\text{mle}}$, and (3) simulating $\tilde{\beta} \mid \hat{\beta}^{\text{mle}}$. If we do steps (1) and (2) just once, but step (3) repeatedly, we have a sample from the conditional distribution $\tilde{\beta} \mid \hat{\beta}^{\text{mle}}$. If we do steps (1), (2), and (3) repeatedly, then we have a sample from the <u>un</u>conditional distribution $\tilde{\beta}$. The unconditional distribution helps us understand the nature of the simulation-induced τ-bias.

Applying the law of iterated expectations, we obtain $\mathrm{E}_{\tilde{\beta}}\left(\tilde{\beta}\right) = \mathrm{E}_{\hat{\beta}^{\text{mle}}}\left(\mathrm{E}_{\tilde{\beta}|\hat{\beta}^{\text{mle}}}(\tilde{\beta} \mid \hat{\beta}^{\text{mle}})\right)$. The three identities below connect the three key quantities from Theorem 1 to three versions of $\mathrm{E}_{\hat{\beta}^{\text{mle}}}\left(\mathrm{E}_{\tilde{\beta}|\hat{\beta}^{\text{mle}}}(\tilde{\beta} \mid \hat{\beta}^{\text{mle}})\right)$, with the transformation $\tau(\cdot)$ applied at different points.

$$\tau\left[\underset{\hat{\beta}^{\text{mle}}}{\mathrm{E}}\left(\underset{\tilde{\beta}|\hat{\beta}^{\text{mle}}}{\mathrm{E}}\left(\tilde{\beta} \mid \hat{\beta}^{\text{mle}}\right)\right)\right] = \tau\left[\underset{\tilde{\beta}}{\mathrm{E}}\left(\tilde{\beta}\right)\right] = \tau\left[\mathrm{E}\left(\hat{\beta}^{\text{mle}}\right)\right], \tag{4}$$

$$\underset{\hat{\beta}^{\text{mle}}}{\mathrm{E}}\left(\tau\left[\underset{\tilde{\beta}|\hat{\beta}^{\text{mle}}}{\mathrm{E}}\left(\tilde{\beta} \mid \hat{\beta}^{\text{mle}}\right)\right]\right) = \underset{\hat{\beta}^{\text{mle}}}{\mathrm{E}}\left(\tau\left[\hat{\beta}^{\text{mle}}\right]\right) = \underset{\hat{\beta}^{\text{mle}}}{\mathrm{E}}\left(\hat{\tau}^{\text{mle}}\right), \text{ and } \quad \longleftarrow \quad \text{Switch } \tau \text{ and an E once.} \tag{5}$$

$$\underset{\hat{\beta}^{\text{mle}}}{\mathrm{E}}\left(\underset{\tilde{\beta}|\hat{\beta}^{\text{mle}}}{\mathrm{E}}\left(\tau\left[\tilde{\beta} \mid \hat{\beta}^{\text{mle}}\right]\right)\right) = \underset{\tilde{\beta}}{\mathrm{E}}\left(\tau\left[\tilde{\beta}\right]\right) = \underset{\tilde{\beta}}{\mathrm{E}}\left(\hat{\tau}^{\text{avg}}\right). \quad \longleftarrow \quad \text{Switch } \tau \text{ and an E again.} \tag{6}$$

If we subtract Equation 5 from Equation 4, we obtain the transformation-induced τ-bias in $\hat{\tau}^{\text{mle}}$ (see Equation 1 for the definition of transformation-induced τ-bias). To move from Equation 4 to Equation 5 we must swap $\tau(\cdot)$ with an expectation once. This implies that, if $\tau(\cdot)$ is convex, Equation 5 must be greater than Equation 4. This, in turn, implies that the bias is positive.

To obtain the τ-bias in $\hat{\tau}^{\text{avg}}$ we must subtract Equation 6 from Equation 4. But to move from Equation 4 to Equation 6 we must swap $\tau(\cdot)$ with an expectation *twice*. Again, if $\tau(\cdot)$ is convex, then Equation 6 must be greater than Equation 4. However, because we expect $\hat{\beta}^{\text{mle}}$ and $\tilde{\beta} \mid \hat{\beta}^{\text{mle}}$ to have similar distributions, we should expect the additional swap to roughly double the bias in $\hat{\tau}^{\text{avg}}$ compared to $\hat{\tau}^{\text{mle}}$. It is this additional swap that leads to simulation-induced τ-bias.

Reanalysis of Holland (2015)

Holland (2015) presents a nuanced theory that describes the conditions under which politicians choose to enforce laws and supports the theoretical argument with a rich variety of evidence. In particular, it elaborates on the *electoral* incentives of politicians to enforce laws. We borrow three Poisson regressions and hypotheses about a single explanatory variable to illustrate how the ML estimate can differ from the simulation average.

Holland writes:

> My first hypothesis is that enforcement operations drop off with the fraction of poor residents in an electoral district. So district poverty should be a negative and significant predictor of enforcement, but only in politically decentralized cities [Lima and Santiago]. Poverty should have no relationship with enforcement in politically centralized cities [Bogota] once one controls for the number of vendors.

To evaluate this claim, we refit Model 1 from Table 2 in Holland (2015) for each city. We use each model to compute the percent increase in the enforcement operations for each district in the city if the percent of the district in the lower class dropped by half. For example, in the Villa Maria El Triunfo district in Lima, 84% of the district is in the lower class. If this dropped to 42%, then the average of simulations suggests that the number of enforcement operations would increase by about 284% (from about 5 to about 19). The ML estimate, on the other hand, suggests an increase of 264% (from about 5 to about 17). The ML estimate, then, is about 7% smaller than the simulation average—a noticeable shrinkage.

Figure 3 shows how the estimates change (usually shrink) for all districts when we switch from the averages of simulations to ML estimates. Table 1 presents the details for the labelled cases in Figure 3. In Bogota, the estimate shrinks by 9% in Sante Fe and 14% in Usme. In Lima, the estimate shrinks by 4% in Chacalacayo and 7% Villa Maria El Triunfo. The shrinkage is much larger in Santiago, where the standard errors for the coefficient estimates are much larger. The estimate shrinks by about 47% in San Ramon and 53% in La Pintana. The median shrinkage is 7% in Bogota,

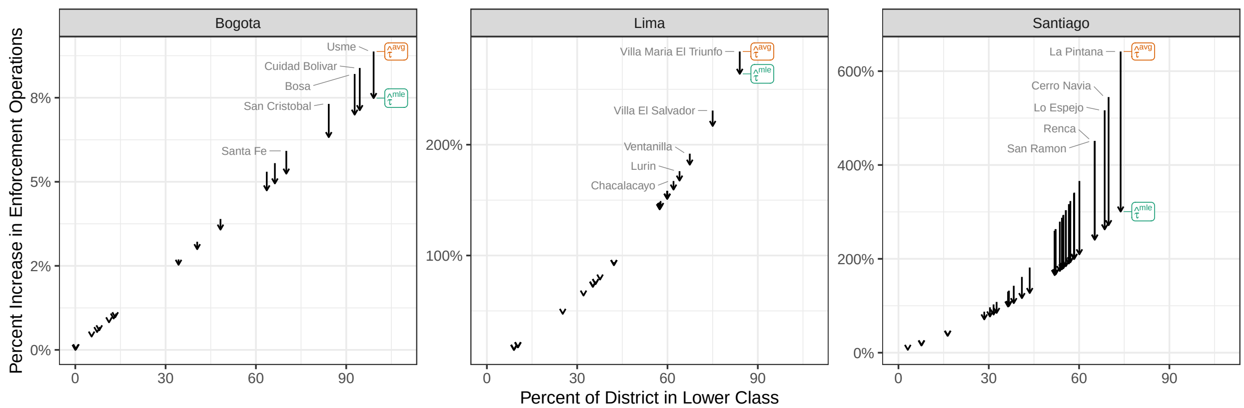

Figure 3: This figure compares the average of simulations with the ML estimate using three Poisson regression models from Holland (2015). The quantity of interest is the percent increase in the enforcement operations when the percent of a district in the lower class drops by half. The arrows show how the estimates change when we switch from the average of simulations to the ML estimate.

2% in Lima, and 36% in Santiago. For many districts in Santiago, the average of simulations is about *twice* the ML estimate. These estimates clearly show that the average of simulations and the ML estimates can meaningfully differ in actual analyses.

Table 1: This table presents the details for the districts labelled in Figure 3.

		Average of Simulations			**ML Estimate**			
City	**District**	**% Change**[a]	**From**[b]	**To**[c]	**% Change**	**From**	**To**	**Shrinkage**[d]
Bogota	**Usme**	9%	5.5	5.8	7%	5.4	5.8	16%
	Cuidad Bolivar	8%	6.1	6.4	7%	5.9	6.3	15%
	Bosa	8%	7.1	7.4	7%	6.9	7.4	15%
	San Cristobal	7%	12.6	13.4	6%	12.5	13.3	14%
	Santa Fe	6%	27.3	29.0	5%	26.6	28.0	11%
Lima	**Villa Maria El Triunfo**	284%	5.3	19.5	264%	4.7	17.1	7%
	Villa El Salvador	231%	7.3	23.5	217%	6.8	21.4	6%
	Ventanilla	192%	8.4	23.4	182%	8.2	23.0	5%
	Lurin	176%	6.9	17.4	168%	6.4	17.1	5%
	Chacalacayo	167%	6.7	16.4	159%	6.2	16.1	5%
Santiago	**La Pintana**	642%	1.4	4.0	301%	0.8	3.4	53%
	Cerro Navia	545%	1.5	4.2	272%	1.0	3.6	50%
	Lo Espejo	517%	1.4	4.3	263%	1.0	3.5	49%
	Renca	451%	1.3	4.1	241%	1.0	3.4	47%
	San Ramon	451%	1.2	4.0	241%	1.0	3.3	47%

[a] Quantity of interest; percent change in enforcement operations when the percent in the lower class drops by half.
[b] Enforcement operations when the percent in the lower class equals its observed value.
[c] Enforcement operations when the percent in the lower class equals half its observed value.
[d] Shrinkage in the quantity of interest due to switching from the average of simulations to the ML estimator.

Conclusion

Many social scientists turn to King, Tomz, and Wittenberg (2000) for advice on how to interpret, summarize, and present empirical results. By highlighting the importance of reporting substantively meaningful quantities of interest, it has significantly improved empirical research in political science and neighboring disciplines. Depending on the statistical software used, researchers estimate quantities of interest either with the average of simulated quantities of interest (e.g., Clarify in Stata, Zelig in R) or using the invariance property of ML estimators (e.g., margins in Stata and R). In practice, researchers' choice between these two estimators seems idiosyncratic rather than principled, depending on their preferred software package rather than any statistical criteria. Further, the methodological literature has not distinguished or compared the two approaches to estimating quantities of interest.

We show that when researchers use the average of simulations to estimate quantities of interest, they roughly double the transformation-induced bias described by Rainey (2017). We refer to this unnecessary additional bias as "simulation-induced bias." The good news is that the fix is easy: we do not have to use the simulation-based approach to obtain a point estimate for a quantity of interest. Instead, we can simply plug estimated coefficients into the transformation to obtain the ML estimate of the quantity of interest. We recommend that statistical software does this by default.

A Proofs

A.1 Proof of Lemma 1

Proof By definition,

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

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

$$\hat{\tau}^{\text{avg}} > \tau\left[\mathrm{E}\left(\tilde{\beta}\right)\right].$$

However, because $\tilde{\beta} \sim \text{MVN}\left[\hat{\beta}^{\text{mle}}, \hat{V}\left(\hat{\beta}^{\text{mle}}\right)\right]$, $\mathrm{E}\left(\tilde{\beta}\right) = \hat{\beta}^{\text{mle}}$, 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. ■

A.2 Proof of Theorem 1

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

For the concave case, it follows similarly that $\underbrace{\mathrm{E}\left(\hat{\tau}^{\text{avg}}\right) - \tau\left[\mathrm{E}\left(\hat{\beta}^{\text{mle}}\right)\right]}_{\text{s.i. and t.i. } \tau\text{-bias in } \hat{\tau}^{\text{avg}}} < \underbrace{\mathrm{E}\left(\hat{\tau}^{\text{mle}}\right) - \tau\left[\mathrm{E}\left(\hat{\beta}^{\text{mle}}\right)\right]}_{\text{t.i. } \tau\text{-bias in } \hat{\tau}^{\text{mle}}} < 0$. ■

References

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

Choirat, Christine, James Honaker, Kosuke Imai, Gary King, and Olivia Lau. 2018. *Zelig: Everyone's Statistical Software*.

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.

Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8(1):83–98.

Holland, Alisha C. 2015. "The Distributive Politics of Enforcement." *American Journal of Political Science* 59(2):357–371.

Imai, Kosuke, Gary King, and Olivia Lau. 2008. "Toward a common framework for statistical analysis and development." *Journal of Computational and Graphical Statistics* 17(4):892–913.

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.

Leeper, Thomas J. 2018. *margins: Marginal Effects for Model Objects*.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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

StataCorp. 2017. *Stata 15 Base Reference Manual*. College Station, TX: Stata Press.

Tomz, Michael, Jason Wittenberg, and Gary King. 2003. "Clarify: Software for Interpreting and Presenting Statistical Results." *Journal of Statistical Software* 8(1).

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth ed. New York: Springer.