Simulation-Induced Bias in Quantities of Interest[*]

Carlisle Rainey[†]

Holger L. Kern[‡]

October 30, 2018

Abstract

Following King, Tomz, and Wittenberg (2000) 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 approaches are not equivalent. We show both analytically and empirically that computing quantities of interest using the average of simulations can introduce substantial bias. To avoid this bias, researchers can use the invariance property to calculate maximum likelihood estimates of quantities of interest.

Political scientists employ maximum likelihood (ML) to estimate a variety of statistical models. ML estimators have desirable and widely understood properties. Researchers fit several common models with ML, including logit and probit models (along with their ordered and multinomial variants), Poisson and negative binomial models, and many varieties of duration models. Even least-squares corresponds to the ML estimates for the normal-linear model. Political scientists (e.g., Katz and King 1999, Mebane 2000, and Sartori 2003) commonly rely ML estimators to fit theoretically motivated statistical models. But for many research questions, the model coefficient estimates are not the quantities of greatest interest to the researcher.

King, Tomz, and Wittenberg (2000) urges researchers to focus on substantively meaningful quantities of interest, such as predicted probabilities, expected counts, marginal effects, and first differences, and provides a method to compute these quantities and the associated uncertainty. Before the publication of King, Tomz, and Wittenberg (2000) and the development of CLARIFY for Stata (Tomz, Wittenberg, and King 2003) and Zelig for R (Imai, King, and Lau 2008; Choirat

et al. 2018), political scientists tended to report lengthy tables of coefficient estimates and pay little attention to the substantive meaning of the estimates beyond their sign and statistical significance.

This foundational work spawned a swath of extensions of the method and software implementations. For example, Licht (2011) extends King, Tomz, and Wittenberg's simulation-based approach to Cox proportional hazard models, and Gandrud (2015) develops the software to implement the method. Williams and Whitten (2012) describes how researchers can use simulation to more fully understand dynamic models of time-series cross-sectional data, and Williams and Whitten (2011) implements their suggestions.

King, Tomz, and Wittenberg (2000) had a tremendous impact on substantive research as well. Citations demonstrate its broad usage. As of September 2018, Google Scholar lists 3,746 citations for King, Tomz, and Wittenberg (2000) 1,487 citations for Tomz, Wittenberg, and King (2003), making it one of the most cited political methodology articles. Further Google Scholar lists 1,487 citations for Tomz, Wittenberg, and King (2003), 349 citations for Imai, King, and Lau (2008), and 334 citations for Choirat et al. (2018). In short, the valuable advice offered by King, Tomz, and Wittenberg (2000) has influenced many scholars and improved research in political science and the social sciences more generally.

In addition to focusing researchers on more relevant quantities of interest, King, Tomz, and Wittenberg (2000) popularized the simulation-based approach to compute these quantities of interest. With the simulation-based approach, researchers simulate model coefficients, transform these simulated coefficients into simulated quantities of interest, and finally summarize the distribution of the simulated quantities of interest.

Our paper addresses one aspect of the advice in King, Tomz, and Wittenberg (2000): to use the average of the simulated quantities of interest as the point estimator. We refer to this estimator as the "average of simulations." We show both analytically and empirically that averaging simulations compounds the transformation-induced bias that Rainey (2017) describes. Researchers can easily avoid compounding the transformation-induced bias by using ML estimates of quantities of interest rather than the average of simulations. Researchers can easily obtain estimates of the quantities of interest from coefficient estimates using the invariance property of ML estimators.

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 (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

2

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: $\hat{\tau}^{\text{mle}} = \tau\left(\hat{\beta}^{\text{mle}}\right) = \text{logit}^{-1}\left(X_c\hat{\beta}^{\text{mle}}\right) = \dfrac{1}{1 + e^{-X_c\hat{\beta}^{\text{mle}}}}$.

Software implementations differ in whether they rely on the invariance property or King, Tomz, and Wittenberg (2000)'s simulation-based approach. Several 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, uses the invariance property. Other 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), report the average of simulations. In the Stata code to create marginal effects plots for a probit model, Brambor, Clark, and Golder (2006) use the average of simulations.

The methodological literature is similarly divided. Herron (1999) uses the invariance property to derive an ML estimator for predicted probabilities in limited dependent variable models (and uses simulation to compute confidence intervals). Even though Herron (1999) cites an earlier version of King, Tomz, and Wittenberg (2000), it does not mention the fact that its approach differs from King, Tomz, and Wittenberg (2000). Carsey and Harden (2013) follows King, Tomz, and Wittenberg (2000) and uses the average of simulations. The literature seems to incorrectly treat both estimators as essentially equivalent. However, as we show, the two are not equivalent. Indeed, the ML estimator has a distinct advantage over the average of simulations.

Transformation-Induced τ-Bias

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

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

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 "A Note of Caution", 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 Average of Simulations

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 $\hat{\beta}^{\mathrm{mle}}$ 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 $i \in \{1, 2, \ldots, M\}$, using $\tilde{\beta}^{(i)} \sim \mathrm{MVN}\left[\hat{\beta}^{\mathrm{mle}}, \hat{V}\left(\hat{\beta}^{\mathrm{mle}}\right)\right]$, 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 $\hat{\tau}^{\mathrm{avg}} = \frac{1}{M} \sum_{i=1}^{M} \tilde{\tau}^{(i)}$.[2]

A Definition of Simulation-Induced Bias

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

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 $\hat{\tau}^{\mathrm{avg}}$ and $\hat{\tau}^{\mathrm{mle}}$ 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 $\hat{\tau}^{mle}$.*

Proof By definition,

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

---

[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 $\hat{\tau}^{\mathrm{avg}}$. In other words, we assume that M is sufficiently large so that $\hat{\tau}^{\mathrm{avg}} = \mathrm{E}\left[\tau\left(\tilde{\beta}\right)\right]$, where $\tilde{\beta} \sim \mathrm{MVN}\left[\hat{\beta}^{\mathrm{mle}}, \hat{V}\left(\hat{\beta}^{\mathrm{mle}}\right)\right]$.

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

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

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

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

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

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

The proof for concave τ follows similarly. ∎

This result is intuitive. Since we simulate using a multivariate normal distribution, $\tilde{\beta}$ has a symmetric distribution. By definition, $\hat{\tau}^{\mathrm{mle}}$ 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 $\hat{\beta}^{\mathrm{mle}}$, then $\tau(\cdot)$ pulls $\tau(\tilde{\beta})$ in toward $\hat{\tau}^{\mathrm{mle}}$. If $\tilde{\beta}$ happens to fall above the mode $\hat{\beta}^{\mathrm{mle}}$, then $\tau(\cdot)$ pushes $\tau(\tilde{\beta})$ away from $\hat{\tau}^{\mathrm{mle}}$. This creates a right-skewed distribution for $\tau(\tilde{\beta})$, which pushes the average $\hat{\tau}^{\mathrm{avg}}$ above $\hat{\tau}^{\mathrm{mle}}$.

For a convex transformation, Lemma 1 shows that $\hat{\tau}^{\mathrm{avg}}$ is always larger than $\hat{\tau}^{\mathrm{mle}}$. We refer to the expectation of this difference between $\hat{\tau}^{\mathrm{avg}}$ and $\hat{\tau}^{\mathrm{mle}}$ as "simulation-induced bias," so that

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

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

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 $\hat{\tau}^{mle}$.*

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

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

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 $\tau(\cdot)$.

An Approximation 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

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

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

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

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

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 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.
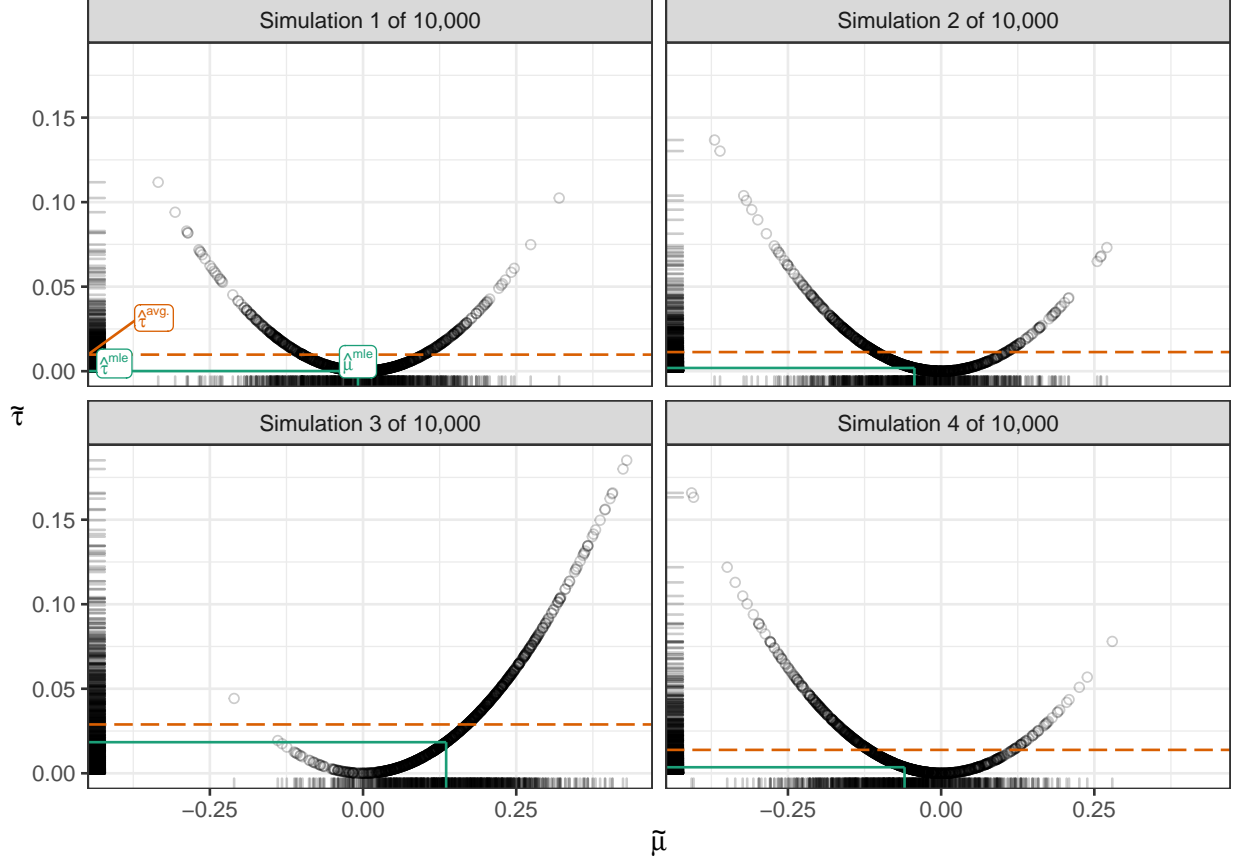

Figure 1: The first four Monte Carlo simulations of $\hat{mu}$. These four panels illustrate the relationship between $\hat{\tau}^{\text{mle}}$ and $\hat{\tau}^{\text{avg}}$ described by Lemma 1 and Theorem 1.

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 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
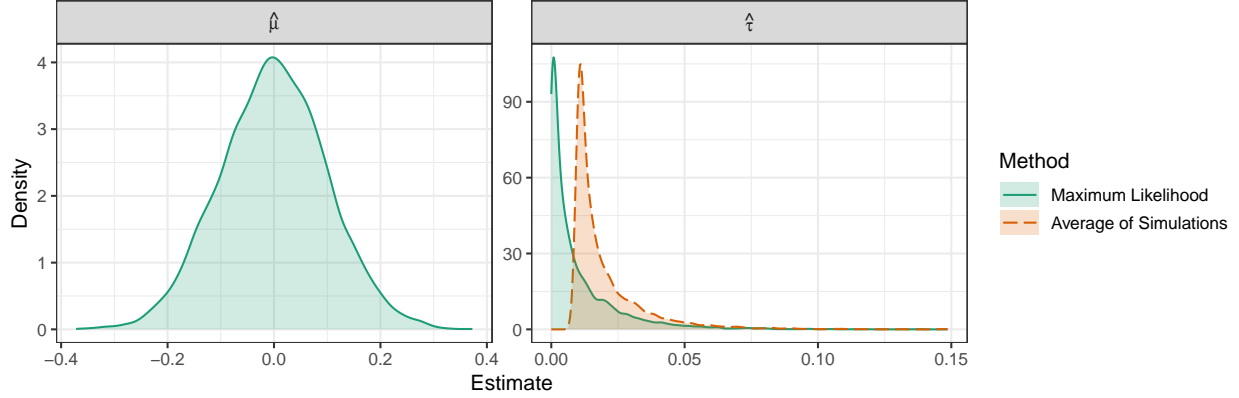
this must be the case.

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

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.

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

9

$\mathrm{E}_{\hat{\beta}^{\mathrm{mle}}}\left(\mathrm{E}_{\tilde{\beta}|\hat{\beta}^{\mathrm{mle}}}(\tilde{\beta} \mid \hat{\beta}^{\mathrm{mle}})\right)$, with the transformation $\tau(\cdot)$ applied at different points.

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

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

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

If we subtract Equation 5 from Equation 4, we obtain the transformation-induced τ-bias in $\hat{\tau}^{\mathrm{mle}}$ (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 $\hat{\tau}^{\mathrm{avg}}$ 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 $\hat{\beta}^{\mathrm{mle}}$ and $\tilde{\beta} \mid \hat{\beta}^{\mathrm{mle}}$ to have similar distributions, we should expect the additional swap to roughly double the bias in $\hat{\tau}^{\mathrm{avg}}$ compared to $\hat{\tau}^{\mathrm{mle}}$. It is this additional swap that leads to simulation-induced τ-bias.

Illustrations of Simulation-Induced Bias

To establish that simulation-induced τ-bias is more than a theoretical curiosity, we use three illustrations to show that simulation-induced τ-bias can matter substantively. First, we reproduce a published analysis from Holland (2015) to show that the average of simulations can meaningfully differ from the ML estimate. Second, we perform a Monte Carlo analysis estimates of marginal effects from a simple Poisson regression to show that the magnitude of the bias can exceed the truth. Third, we compute the coefficient-, transformation-, and simulation-induced bias for estimates of first-differences from a realistic logistic regression to show that simulation-induced bias (1) behaves similarly to transformation-induced bias and (2) can remain even for samples large enough to essentially eliminate coefficient-induced 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.
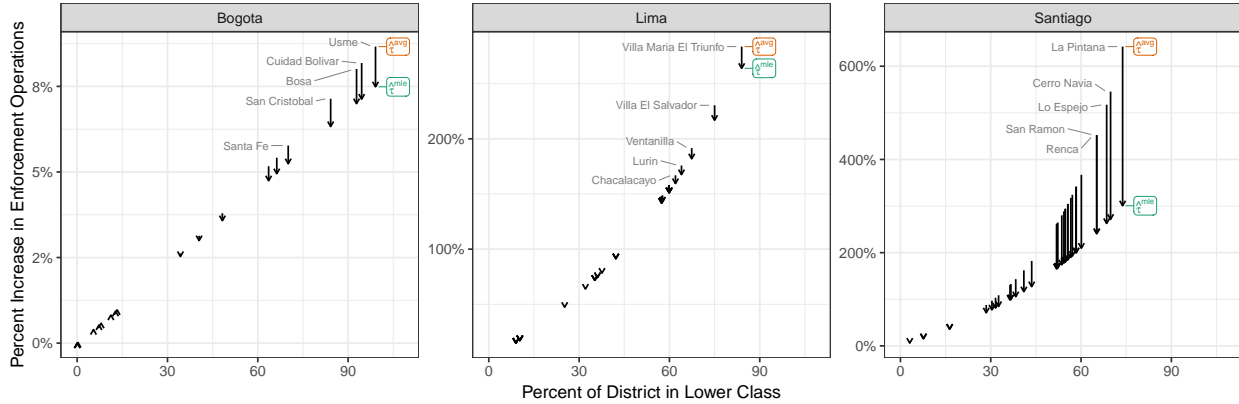

Figure 3: This figure compares the average of simulations with the ML estimate using three Poisson regression models from Holland (2015). The quantity of interest is the percent increase in the enforcement operations when the percent of a district in the lower class drops by half. The arrows show how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

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, 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.

City	District	Average of Simulations			ML Estimate			
		% Change[a]	From[b]	To[c]	% Change	From	To	Shrinkage[d]
Bogota	Usme	9%	5.5	5.8	7%	5.4	5.8	14%
	Cuidad Bolivar	8%	6.1	6.4	7%	5.9	6.3	13%
	Bosa	8%	7.1	7.4	7%	6.9	7.4	13%
	San Cristobal	7%	12.7	13.4	6%	12.5	13.3	11%
	Santa Fe	6%	27.3	28.9	5%	26.6	28.0	9%
Lima	Villa Maria El Triunfo	284%	5.2	19.4	264%	4.7	17.1	7%
	Villa El Salvador	230%	7.2	23.3	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	4%
Santiago	La Pintana	642%	1.4	4.0	301%	0.8	3.4	53%
	Cerro Navia	546%	1.5	4.3	272%	1.0	3.6	50%
	Lo Espejo	517%	1.4	4.4	263%	1.0	3.5	49%
	Renca	452%	1.3	4.1	241%	1.0	3.4	47%
	San Ramon	452%	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.

Stylized Example from Rainey (2017)

To illustrate transformation-induced bias, Rainey (2017) considers the simple, fictional model

$$\log(\text{income}_i) = \beta_{cons} + \beta_{edu}\text{education}_i + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$, education is measured in years, and income is measured in thousands of dollars. Rainey uses the invariance property to compute ML estimates of the median income among those with 20 years of education $\text{med}(\text{income} \mid \text{education} = 20) = e^{\beta_{cons}+20\beta_{edu}}$. Following Rainey, we suppose that $N = 10$, $\beta_{cons} = 2.5$, $\beta_{edu} = 0.1$, $\sigma^2 = 1$, and education takes on integers roughly uniform from 10 to 20. In this case, $\tau(\beta_{cons}, \beta_{edu}) = e^{\beta_{cons}+20\beta_{edu}} \approx \$90k$.

We use Monte Carlo simulation to obtain the sampling distributions of $\hat{\tau}^{\text{mle}}$ and $\hat{\tau}^{\text{avg}}$. As in Rainey (2017), we find that $\hat{\tau}^{\text{mle}}$ is biased upward by about \$15k. But we find that that average of simulations is biased upward by about \$33k. We know from statistical theory that there is no coefficient-induced τ-bias in this example. Because this transformation is strictly convex, we know from Rainey (2017) that the transformation-induced τ-bias is positive. Similarly, our Lemma 1 reveals that every average of simulations must be greater than the associated ML estimate. Therefore, our Theorem 1 reveals that the *expectation* of the average of simulation must be greater than the expectation of the (upwardly biased) ML estimator. Further, notice that, consistent with

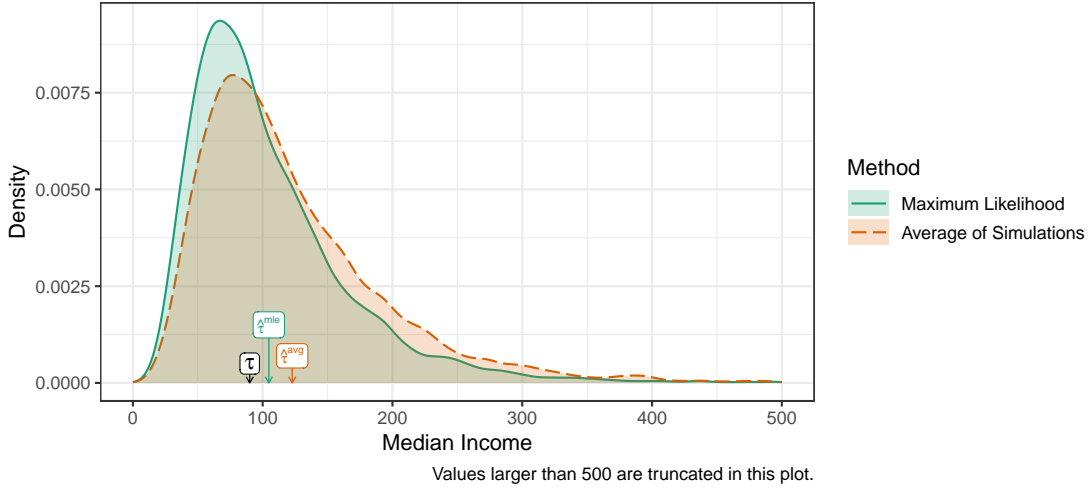our approximation, the bias in the average of simulation is about twice the bias in the ML estimator.

Figure 4: This figure shows the sampling distributions of the average of simulations and the ML estimator of med(income | education = 20). The labels indicate the true value τ of about \$90k (in black) and the expectations of the the ML estimate $\hat{\tau}^{\text{mle}}$ of about \$105k (in green) and the average of simulations $\hat{\tau}^{\text{avg}}$ of about \$122k (in orange).

First Differences in a Logistic Regression Model

Next, we illustrate our concepts for a more complex, realistic logistic regression model. To obtain a plausible data-generating process and realistic set of explanatory variables, we base our Monte Carlo study on estimates and data from Berry, DeMeritt, and Esarey (2010). This paper serves as an ideal example because it is widely cited, uses a large data set, and builds on a familiar series of substantive and methodological papers, including Wolfinger and Rosenstone (1980), Nagler (1991, 1994), and Altman and McDonald (2003). For our data-generating process, we use the model specification and coefficients that Berry, DeMeritt, and Esarey (2010) reports in their Table 1, Column 2. To generate suitably-small samples of explanatory variables, we randomly draw a sample of 100 observations and subsequently add in 100 more , then 200 more, and then 400 more observations to create samples of 100, 200, 400, and 800 observations from the original data set from Berry, DeMeritt, and Esarey (2010).

In our Monte Carlo simulations, we keep the explanatory variables fixed and use the coefficient estimates from Berry, DeMeritt, and Esarey (2010) to simulate 10,000 outcome variables. For each simulated outcome variable, we compute $\hat{\beta}$, $\hat{\tau}^{\text{mle}}$, and $\hat{\tau}^{\text{avg}}$. Finally, we use these estimates to decompose the τ-bias into its coefficient-induced, transformation-induced, and simulation-induced components.

As our quantity of interest, we calculate the change in the probability of turning out to vote if the election registration deadline occurs 10 days sooner (a one standard deviation change). However, the effect of this 10-day shift (and the coefficient-, transformation-, and simulation-induced τ-bias) depends on the original closing date as well as the values of the other explanatory variables. To capture and present this heterogeneity, we estimate the coefficient-, transformation-, and simulation-induced biases for 250 randomly-chosen observations from the full data set.

Figure 5 shows the resulting coefficient-, transformation-, and simulation-induced τ-bias (columns) for the four different sample sizes (rows). Each point shows the bias for one of the 250 observations. Points that fall outside the shaded region have a bias greater than 10%.

First, note that the simulation-induced τ-bias (right-hand column) has a magnitude comparable to the well-known small sample bias in logistic regression coefficients (left-hand column). The magnitude of the simulation-induced τ-bias is similar to the magnitude of the transformation-induced τ-bias (middle column), in line with our approximation. Second, at least in some scenarios, the simulation-induced bias is sufficiently large to meaningfully affect results. For at least some observations, the simulation-induced bias remains larger than 10% until the sample size exceeds 400. Third, while each type of bias disappears asymptotically, the biases disappear at different rates. Rainey (2017) shows that transformation-induced bias can disappear more slowly than coefficient-induced bias. We see the same pattern for simulation-induced bias. Especially for certain observations, the simulation-induced bias remains large even when the coefficient-induced bias has nearly disappeared.

A Note on Hanmer and Kalkan (2013)

Hanmer and Kalkan (2013) recommend that researchers avoid the *average-case* approach for computing quantities of interest and instead use the *observed-value* approach. With either approach, researchers estimate the quantity of interest as a key explanatory variable changes its value. However, with non-linear models, researchers must also consider the other explanatory variables in the model since they alter the quantity of interest. The average-case approach sets the other explanatory variables to central values such as the mean, median, or mode. Hanmer and Kalkan (2013), in contrast, suggests estimating the quantity of interest for all sample observations, leaving the explanatory variables except for the key variable of interest at their observed values, and then averaging the estimates across the sample. Here, our notation captures this choice as part of the transformation $\tau(\cdot)$, so Hanmer and Kalkan's (compelling) argument does not undermine or enhance our own.[3]

Because researchers have not drawn a sharp conceptual distinction between the simulation

---

[3]We generally agree with the arguments in favor of the observed-value approach but recommend that researchers plot the distribution of quantities of interest in addition to providing a summary measure such as their average. See Ai and Norton (2003) for discussion and examples.
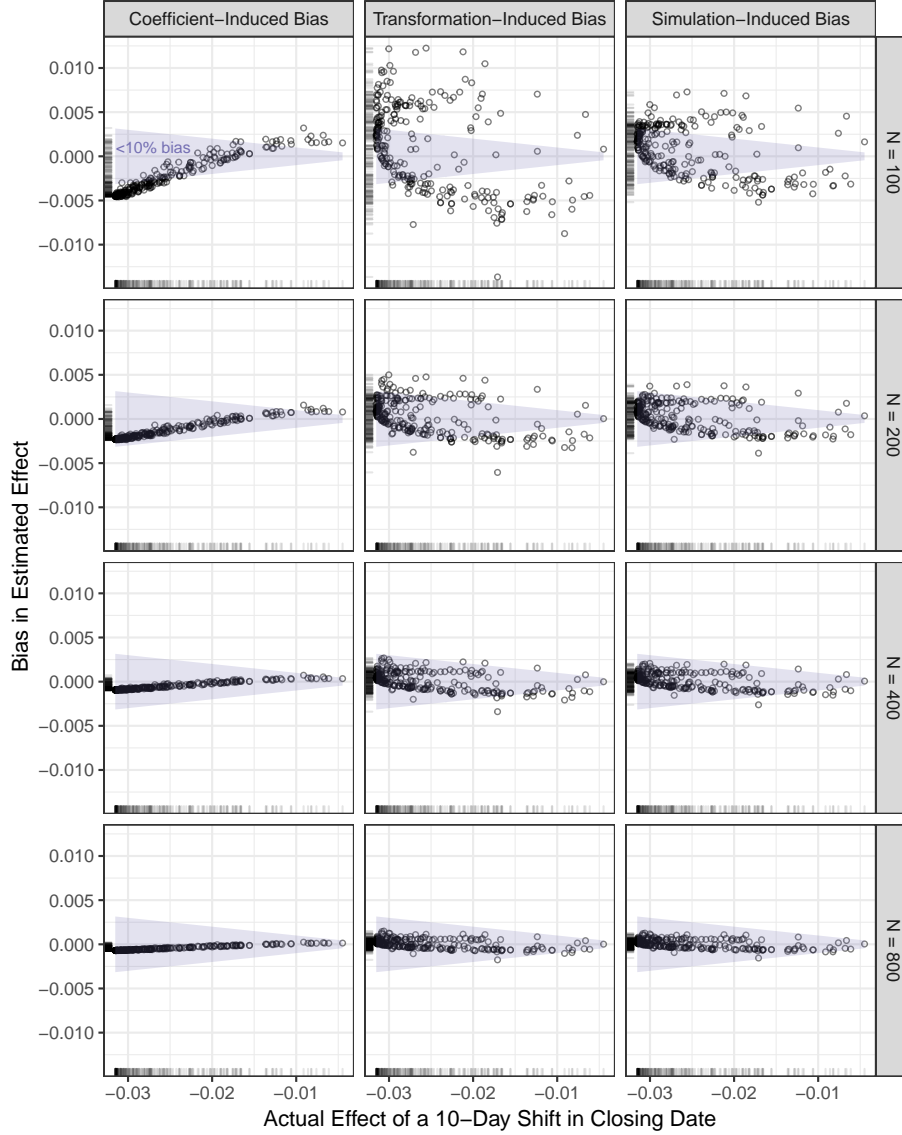
Figure 5: The figure shows the coefficient-induced, transformation-induced, and simulation-induced τ-biases for a logistic regression model based on Berry, DeMeritt, and Esarey (2010). The points that fall outside the shaded region have bias greater than 10%.

average estimator and the ML estimator, Hanmer and Kalkan (2013) does not discuss this choice. Since it explicitly builds on King, Tomz, and Wittenberg (2000), we interpret Hanmer and Kalkan (2013) as relying on the simulation average estimator when computing quantities of interest. The replication archive for the article confirms that this is indeed the case.

The important point is this: Hanmer and Kalkan (2013) draws a distinction between the average-case and observed-value approaches to computing quantities of interest. Our paper draws a

distinction between computing estimates of quantities of interest (whether average-case or observed-value based) using the invariance property of ML estimators or using King, Tomz, and Wittenberg's (2000) simulation-based approach. Regardless of whether researchers use the average-case approach or the observed-value approach, the simulation-based approach leads to estimates that include simulation-induced bias that researchers can easily avoid by relying on the invariance property of ML estimators instead.

A Note of Caution

Our analysis focuses on the bias in estimators of quantities of interest. In particular, we focus additional bias attributable averaging simulations–what we call simulation-induced bias. We urge caution along three dimensions.

First, we remain agnostic about the best strategy to compute standard errors and confidence intervals. Commonly employed approaches include the delta method, simulation, and the bootstrap (e.g., Efron and Tibshirani (1993). For example, the margins package (Leeper 2018) in R reports ML estimates of the quantities of interest, but allows the user to choose among the delta method (default), simulation, and the bootstrap. Krinsky and Robb (1991) presents some limited Monte Carlo evidence that these approaches lead to similar inferences, but the literature has not offered a detailed comparison of these approaches.

Second, the presence of simulation-induced τ-bias (and the transformation-induced τ-bias of Rainey [2017]) does not always imply greater total τ-bias. Indeed, the three sources of τ-bias (simulation-induced, transformation-induced, and coefficient-induced) can cancel, so that just the right amount of simulation- and transformation-induced τ-bias can cancel the coefficient-induced τ-bias. Indeed, simulation- and transformation-induced τ-bias can work much like Firth's (1993) modified score function to correct the coefficient-induced τ-bias.

Third, applied researchers care about more than bias. In particular, they care about both bias and *variance*. What can we say about the effects of simulation and transformation on the variance? In particular, what about the effect on the mean squared error (MSE)?

At first, it might seem like an estimator that is strictly greater than an upwardly biased estimator cannot have a smaller MSE. But this intuition is wrong.

Suppose a convex transformation and define $\delta = \hat{\tau}^{\text{avg}} - \hat{\tau}^{\text{mle}}$. In this case, δ is deterministic for a particular sample, but a random variable across samples. However, (for a convex transformation) δ is strictly positive. By defintion $\text{MSE}\left(\hat{\tau}^{\text{mle}}\right) = \text{E}\left[(\hat{\tau}^{\text{mle}} - \tau)^2\right]$. Similarly,

$\text{MSE}\left(\hat{\tau}^{\text{avg}}\right) = \text{E}\left[(\hat{\tau}^{\text{avg}} - \tau)^2\right]$. We can manipulate $\text{MSE}\left(\hat{\tau}^{\text{avg}}\right)$ as follows:

$$
\begin{aligned}
\text{MSE}\left(\hat{\tau}^{\text{avg}}\right) &= \text{E}\left[(\hat{\tau}^{\text{mle}} + \delta - \tau)^2\right] \quad \text{(def. of } \delta) \\
&= \text{E}\left[((\hat{\tau}^{\text{mle}} - \tau) + \delta)^2\right] \quad \text{(rearrange)} \\
&= \text{E}\left[(\hat{\tau}^{\text{mle}} - \tau)^2 + 2\delta(\hat{\tau}^{\text{mle}} - \tau) + \delta^2\right] \quad \text{(expand)} \\
&= \text{E}\left[(\hat{\tau}^{\text{mle}} - \tau)^2\right] + 2\text{E}\left[\delta(\hat{\tau}^{\text{mle}} - \tau)\right] + \text{E}\left[\delta^2\right] \quad \text{(distribute expectation)} \\
&= \text{MSE}\left(\hat{\tau}^{\text{mle}}\right) + 2\text{E}\left[\delta(\hat{\tau}^{\text{mle}} - \tau)\right] + \text{E}\left[\delta^2\right] \quad \text{(def. of MSE)} \\
\text{MSE}\left(\hat{\tau}^{\text{avg}}\right) - \text{MSE}\left(\hat{\tau}^{\text{mle}}\right) &= 2\text{E}\left[\delta(\hat{\tau}^{\text{mle}} - \tau)\right] + \text{E}\left[\delta^2\right] \quad \text{(rearrange)} \\
&= 2\text{E}\left[\delta(\hat{\tau}^{\text{mle}} - \tau)\right] + \text{E}\left[\delta^2\right] \quad \text{(rearrange)} \\
&= 2\left[\text{E}(\delta)E(\hat{\tau}^{\text{mle}} - \tau) + \text{Cov}(\delta, \hat{\tau}^{\text{mle}} - \tau)\right] + \text{E}\left[\delta^2\right] \quad \text{(def. of covariance)} \\
&= 2\underbrace{\text{Cov}(\delta, \hat{\tau}^{\text{mle}} - \tau)}_{\text{sign unknown}} + 2\underbrace{\text{E}(\delta)}_{\text{positive}}\underbrace{E(\hat{\tau}^{\text{mle}} - \tau)}_{\text{positive}} + \underbrace{\text{E}\left[\delta^2\right]}_{\text{positive}} \quad \text{(rearrange)} \quad (7)
\end{aligned}
$$

At our level of generality, $\text{Cov}(\delta, \hat{\tau}^{\text{mle}} - \tau)$ can be positive or negative. If it is negative and larger than the second and third terms of Equation 7, then the MSE for the simulation average is smaller than the ML estimator. However, this condition does not hold in general.

As an example, Figure 6 shows the bias, standard deviation (SD), and root mean squared error (RMSE) for the sampling distributions from Rainey (2017) shown in Figure 4. In this case, the average of simulations has a higher bias, SD, and RMSE. The ML estimator dominates the average of simulations.
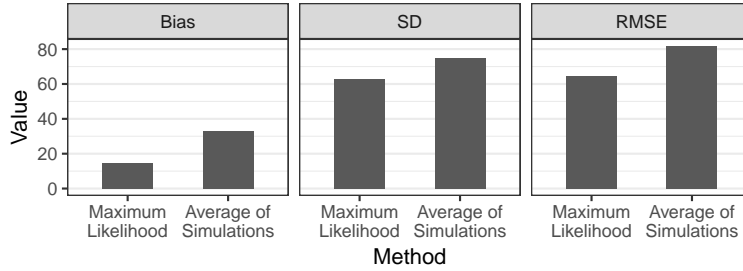

Figure 6: This figure compares the bias, SD, and RMSE of the sampling distributions of the average of simulations and the ML estimator of med(income | education = 20).

But not all cases are as simple as the stylized example from Rainey (2017). That simple scenario has an increasing, convex transformation and no coefficient-induced bias. To illustrate the complexity of the issue, we use Monte Carlo simulation to compute the bias, SD, and RMSE of estimates of marginal effects from a simple logit model. Let $\pi_i = \text{logit}^{-1}(\beta_{\text{cons}} + \beta_{\text{x}} x_i)$ for $i = \{1, 2, ..., 100\}$, $\beta_{\text{cons}} = 0$, and $\beta_{\text{x}} = 1$. We create x_i by drawing from a standard normal distribution. In this

scenario, the transformation is neither increasing nor convex and the coefficients estimates have bias. Figure 7 shows the bias, variance, MSE of the estimates of the marginal effect as x ranges from -3 to $+3$.
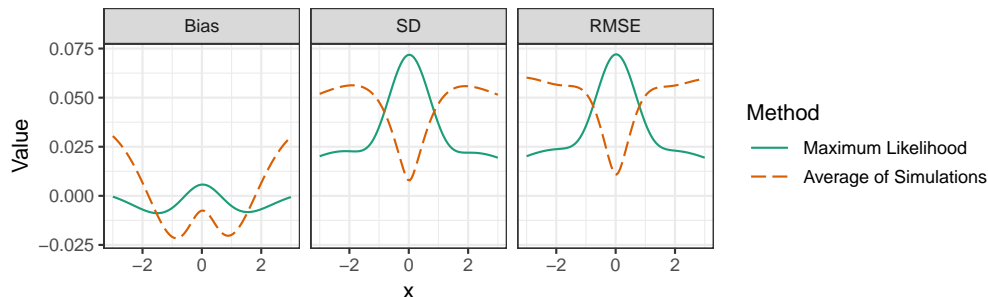

Figure 7: This figure compares the bias, SD, and RMSE of the sampling distributions of the simulation average and the ML estimators of the marginal effect of x in a simple logit model.

In this simple simulation, the ML estimate has a smaller absolute bias for about 80% of the range of x, a smaller variance in about 72% of the range of x, and a smaller MSE for about 75% of the range of x. Thus, in some scenarios, the average of simulations have a smaller total τ-bias, RMSE, or both. Note, though, that these patterns do not necessarily generalize to other scenarios, including other statistical models, other true coefficients, and other quantities of interest. These simulations do illustrate, though, the complexity of the tradeoff.

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 that Rainey (2017) describes. 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.

References

Ai, Chunrong, and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80(1):123–129.

Altman, Micah, and Michael P McDonald. 2003. "Replication with Attention to Numerical Accuracy." *Political Analysis* 11(3):302–307.

Berry, William D., Jacqueline H. R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential." *American Journal of Political Science* 54(1):105–119.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding interaction models: Improving empirical analyses." *Political analysis* 14(1):63–82.

Carsey, Thomas M, and Jeffrey J Harden. 2013. *Monte Carlo Simulation and Resampling Methods for Social Science.* Sage Publications.

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.*

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1):27–38.

Gandrud, Christopher. 2015. "simPH: An R Package for Showing Estimates from Cox Proportional Hazard Models including for Interactive and Nonlinear Effects." *Journal of Statistical Software* 65(3):1–20.

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.

Katz, Jonathan, and Gary King. 1999. "A Statistical Model for Multiparty Electoral Data." *American Political Science Review* 93:15–32.

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.

Krinsky, Itzhack, and A. Leslie Robb. 1991. "Three Methods for Calculating the Statistical Properties of Elasticities: A Comparison." *Empirical Economics* 16(2):199–209.

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

Licht, Amanda A. 2011. "Change Comes with Time: Substantive Interpretation of Nonproportional Hazards in Event History Analysis." *Political Analysis* 19(2):227–243.

Mebane, Walter R. 2000. "Coordination, Moderation, and Institutional Balancing in American Presidential and House Elections." *American Political Science Review* 94(01):37–57.

Nagler, Jonathan. 1991. "The Effect of Registration Laws and Education on US Voter Turnout." *The American Political Science Review* 85(4):1393–1405.

Nagler, Jonathan. 1994. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38(1):230–255.

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.

Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions." *Political Analysis* 11(2):111–138.

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.

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

Williams, Laron K., and Guy D. Whitten. 2012. "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models." *The Journal of Politics* 74(3):685–693.

Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* Yale University Press.