THE IMPACT OF ESTIMATION UNCERTAINTY ON COVARIATE EFFECTS IN NONLINEAR MODELS

IVAN JELIAZKOV AND ANGELA VOSSMEYER

Abstract. Covariate effects are a key consideration in model evaluation, forecasting, and policy analysis, yet their dependence on estimation uncertainty has been largely overlooked in previous work. We discuss several approaches to covariate effect evaluation in nonlinear models, examine computational and reporting issues, and illustrate the practical implications of ignoring estimation uncertainty in a simulation study and applications to educational attainment and crime. The evidence reveals that failing to consider estimation variability and relying solely on parameter point estimates may lead to nontrivial biases in covaraite effects that can be exacerbated in certain settings, underscoring the pivotal role that estimation uncertainty can play in this context.

Keywords: Covariate effect; Discrete data; Marginal effect; Nonlinear model; Partial effect.
JEL Codes: C10, C18, C50.

1. Introduction

Unlike the case of linear models, parameter interpretation in nonlinear settings is more difficult and context-dependent because the effect of any covariate typically depends on other covariates and parameters in the model. In addition, a key implication of nonlinearity for covariate effect estimation is that it is essential to deal with both data variability and parameter uncertainty. While some attention has been paid to tackling the former of these considerations, studies rarely address the latter despite its importance. We show that covariate effects constructed without accommodating parameter variability can be misleading, and discuss conditions when differences can be particularly large.

To motivate the discussion, consider the popular probit model for binary data, where $y_i \in \{0, 1\}$, and given covariate and parameter vectors x_i and β, respectively, the model is given by

$$\Pr(y_i = 1|x_i, \beta) = \Phi(x_i'\beta), \qquad i = 1, \ldots, n, \tag{1}$$

where $\Phi(\cdot)$ is the standard normal cdf. The impact of changing the jth covariate x_{ij} on the outcome probability in (1) can be evaluated in a number of ways, leading to differences in terminology and reporting conventions across fields and disciplines. In this paper we employ the term *covariate effect* to refer to that impact more broadly, subsuming terms such as marginal, partial, or average effects, when the

Date: April 2016.

Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697. E-mail: ivan@uci.edu. Robert Day School of Economics and Finance, Claremont McKenna College, 500 E. Ninth St. Claremont, CA 91711. Email: angela.vossmeyer@cmc.edu.

covariates are continuous, discrete, ordinal, etc. The derivative of (1) (or marginal effect) with respect to x_{ij} is given by

$$\frac{\partial \Pr(y_i = 1|x_i, \beta)}{\partial x_{ij}} = \phi(x_i'\beta)\beta_j, \tag{2}$$

where $\phi(\cdot)$ is the normal pdf. This reveals that the impact of changing x_{ij} depends on the entire vector x_i and all parameters β, not just β_j, and it is nonlinear due to the presence of $\phi(\cdot)$ in (2). The same points are true in the case of discrete changes, where for two vectors $x_i^\dagger$ and $x_i^\ddagger$ that differ only in the value of x_{ij},

$$\Pr\left(y_i = 1|x_i^\dagger, \beta\right) - \Pr\left(y_i = 1|x_i^\ddagger, \beta\right) = \Phi\left(x_i^{\dagger\prime}\beta\right) - \Phi\left(x_i^{\ddagger\prime}\beta\right), \tag{3}$$

which makes it clear that nonlinearity, due to $\Phi(\cdot)$, and dependence on all other covariates and parameters, are still present for the effect of a given x_{ij}.

In the past, marginal effects have often been computed at the average value of the covariates $\bar{x}$ and the parameter point estimate $\hat{\beta}$ (see, e.g., [5] or [10]):

$$\delta_{j1} = \phi\left(\bar{x}'\hat{\beta}\right)\hat{\beta}_j. \tag{4}$$

Unfortunately, $\bar{x}$ can be meaningless if the set of covariates involves categorical, indicator, or any other kind of discrete variables. The effect can be unrepresentative even with continuous covariates if $\bar{x}$ falls in a low density region (e.g., if that distribution is multimodal), in which case policy recommendations based on $\bar{x}$ may not reflect the effects in the population. Despite these deficiencies, computations based on (4) have frequently been employed in the literature.

To avoid some of the aforementioned issues, covariate effects can be evaluated by averaging (2) over the sample (sometimes called average partial effects), i.e.,

$$\delta_{j2} = \frac{1}{n}\sum_{i=1}^{n} \phi\left(x_i'\hat{\beta}\right)\hat{\beta}_j = \overline{\phi\left(x_i'\hat{\beta}\right)}\hat{\beta}_j. \tag{5}$$

The computation in (5), by virtue of averaging the effects rather than the covariates as in equation (4), is more representative of the actual effects in the sample. A comparison of the estimators in (4) and (5) is offered in [9], who derives several important results on their relationship and provides conditions under which their difference can be signed.

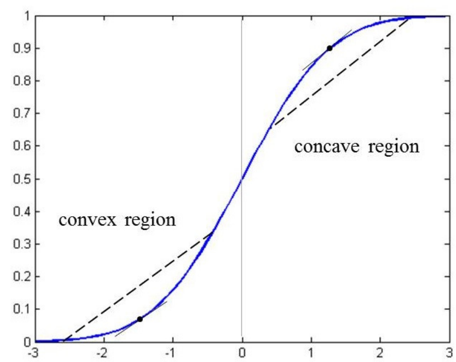
Unfortunately, neither δ_{j1} in (4) nor δ_{j2} in (5) account for the uncertainty due to estimating the parameters as both employ only the point estimate $\hat{\beta}$, rather than its full distribution. This can be clearly inadequate in nonlinear settings because Jensen's inequality applies – for a convex function $h(\cdot)$, the inequality states that

$$E[h(z)] \geq h[E(z)],$$

whereas the opposite is true when $h(\cdot)$ is concave. In the probit context, the link function $\Phi(\cdot)$ has both concave and convex regions, as illustrated in Figure 1, so that the presence of estimation variability would imply that estimates that ignore it (such as those in equations (4) and (5)) could exhibit biases of unknown sign and magnitude. Because nonlinearity affects the mean of the estimated covariate effect, not just its variance, confidence interval adjustments alone are insufficient to resolve this problem. It is also worth noting that the estimation variability in *all* components of β, not only β_j, will impact the estimated covariate effect of

x_j. These considerations emphasize the important role that parameter uncertainty plays in the computation, reporting, and interpretation of covariate effects.

FIGURE 1. Jensen's inequality can impact covariate effect estimates in an unknown direction depending on the shape of the link function.



In practical terms, this calls for another layer of integration that relates to uncertainty in the parameters. In Bayesian analysis, marginalization over the parameters is done with respect to the posterior distribution $\pi(\beta|y) \propto f(y|\beta)\,\pi(\beta)$ where $f(y|\beta)$ is the likelihood and $\pi(\beta)$ is the prior (see, e.g., [4], [7]). Specifically, the Bayesian computation involves

$$\begin{aligned}
\delta_{j3} &= \int \frac{\partial \Pr(y_i = 1|x, \beta)}{\partial x_j} f(x)\, \pi(\beta|y)\, dx d\beta \\
&\approx \frac{1}{nM} \sum_{i=1}^{n} \sum_{m=1}^{M} \frac{\partial \Pr\left(y_i = 1|x_i, \beta^{(m)}\right)}{\partial x_j} \\
&= \frac{1}{nM} \sum_{i=1}^{n} \sum_{m=1}^{M} \phi\left(x_i' \beta^{(m)}\right) \beta_j^{(m)} \\
&= \overline{\phi(x_i'\beta)\,\beta_j}
\end{aligned}, \tag{6}$$

where marginalization over x employs the empirical distribution of the covariates $f(x)$ as in as in the case of (5), and integration over the parameters is done using a sample from the posterior $\beta^{(m)} \sim \pi(\beta|y)$, $m = 1, \ldots, M$. When dealing with discrete covariates, the computation would, similarly to (3), need to be modified to

$$\delta_{j3} = \int [\Pr\left(y_i = 1|x^{\dagger}, \beta\right) - \Pr\left(y_i = 1|x^{\ddagger}, \beta\right)] f(x)\, \pi(\beta|y)\, dx d\beta.$$

The classical (or frequentist) analogue to (6) would involve integration of the estimate of β with respect to its sampling distribution $q(\hat{\beta})$, rather than the posterior $\pi(\beta|y)$. One possibility is to employ the parametric bootstrap to sample $\hat{\beta}$ from its asymptotic (e.g., Gaussian) distribution. Another option, if computational costs are minor or moderate, is to employ the paired bootstrap to obtain draws $\hat{\beta}$ without relying on the parametric approximation. In either case, the draws from the asymptotic distribution can be used to construct the classical analogue to (6),

$$\delta_{j3} = \frac{1}{nM} \sum_{i=1}^{n} \sum_{m=1}^{M} \phi\left(x_i' \hat{\beta}^{(m)}\right) \hat{\beta}_j^{(m)}. \tag{7}$$

As noted in [1], standard econometric packages can be used for bootstrapping, reducing the practical costs associated with implementing (7).

A few remarks are in order regarding the estimators discussed so far. First, δ_{j3} obtained by equation (6) will approach the estimate from equation (7) as the Bayesian posterior distribution approaches the classical asymptotic distribution. Second, the parametric and paired bootstrap estimates of δ_{j3} would be close when the paired bootstrap distribution is approximately Gaussian. Third, as $n \to \infty$, δ_{j3} from (6) or (7) will converge to δ_{j2} from (5), noting that for ill-conditioned data matrices, data sets with unbalanced outcomes, or weakly identified models, this convergence may be slow. Fourth, the quantities in (6) and (7) provide estimates that integrate out both variability in the sample and estimation uncertainty, and therefore basic summaries such as confidence or credibility bands that are commonly reported for other estimates would not be sensible here as there is no residual uncertainty that such bands could represent. It is important to note, however, that variability in the covariate effect can still be captured by summarizing the distribution of the terms $\phi(x_i'\beta^{(m)})\beta_j^{(m)}$ or $\phi(x_i'\hat{\beta}^{(m)})\hat{\beta}_j^{(m)}$ that enter the averages in (6) or (7), e.g. by reporting quantiles or presenting a histogram. Note also that other interesting aspects of the covariate effect can be represented analogously, by conditioning on particular variables or by integrating either over the parameter distribution or the data distribution, but not both. In the former case, we are left with a distribution of the covariate effect over the units in the sample (individuals, firms, etc.); in the latter, we have a distribution of the average (over the sample units) effect as a function of the parameter uncertainty. Both of these marginal distributions and their summaries may be interesting and important in particular contexts and could be easily obtained as a by-product of the more general average δ_{j3}. Regardless of which methods are considered suitable and how covariate effects are ultimately summarized in a particular setting, a key point is that parameter uncertainty directly affects the point estimates for those effects.

2. Practical Illustrations

To illustrate the points made earlier, we examine two applications and augment the findings with simulation evidence to make points that are difficult to address with real data. Our first application involves educational attainment using data from the National Longitudinal Survey of Youth (NLSY79). Educational attainment has been the subject of active discourse in academic and policy circles, but it is useful for our purposes because the attainment of education can naturally be categorized by various thresholds such as the completion of high school or college and allows the study of how variations in the dependent variable (for a given set of covariates) will affect the behavior of the estimated covariate effects. The second application considers a Poisson model for crime data, which offers a useful distinction from the probit example because under the Poisson link (which is strictly convex) we can determine the sign of the bias caused by ignoring variability.

2.1. **Probit Example.** In the education application we use data from [7] and examine three probit models under alternative specifications for the binary outcomes – in the first model, $y_i = 1\{\text{high school degree}\}$; in the second, $y_i = 1\{\text{some college}\}$; and in the third model, $y_i = 1\{\text{college or graduate degree}\}$. Under the first discretization, the fraction of ones is 77%, in the second case it is 42%, and in the last case, it is 19%. In addition to exploiting alternative outcome variables, we also focus on different subgroups in the sample in order to evaluate the differences in the estimated covariate effects in less well-behaved contexts.

For this application, we use data on cohorts 14-17 of age in 1979 for whom a family income variable (average family income, in tens of thousands of 1980 dollars, at age 16 and 17) is available. The sample consists of 3923 individuals and includes variables on an individual's family at the age of 14 including the highest grade completed by the father and mother, whether the mother worked, family income (stabilized by a square root transformation), and whether the youth lived in an urban area or the South at the age of 14. We also include the individual's gender and race, and to control for age cohort affects, we add indicator variables for an individual's age in 1979. Additional discussion of the data can be found in [7].

The results of the analysis using the three definitions of y_i are presented in Table 1. Estimation was performed by Gibbs sampling, and the posterior means and standard deviations were very close to the frequentist MLE and standard errors which were also obtained as a robustness check. The coefficients presented in the table accord with findings in the literature. Based on these different regressions, our goal in the next section is to compare the behavior of δ_1, δ_2, and δ_3, computed for the family income variable, in several different settings.

Table 1. Posterior means and standard deviations.

	High School		Some College		College	
Covariate	Mean	SD	Mean	SD	Mean	SD
Constant	-1.218	0.117	-2.412	0.120	-3.465	0.160
Mother Working	0.073	0.050	0.048	0.044	-0.034	0.050
Urban	-0.214	0.062	0.055	0.053	0.011	0.064
South	0.028	0.053	0.060	0.048	0.074	0.057
Father Education	0.058	0.008	0.074	0.008	0.086	0.009
Mother Education	0.039	0.010	0.047	0.010	0.076	0.012
Family Income	1.76	0.15	1.16	0.12	1.25	0.14
Female	0.293	0.049	0.134	0.042	0.059	0.050
Black	0.298	0.061	0.157	0.056	-0.038	0.067
Age 15	-0.044	0.069	0.023	0.064	-0.088	0.077
Age 16	0.007	0.075	0.063	0.067	-0.043	0.077
Age 17	0.232	0.077	0.235	0.068	0.202	0.078

2.1.1. *Some Basic Lessons.* We begin our discussion by referring the reader to Table 2 and noting the effect of the sample size on the behavior of δ_1 and δ_2 relative to δ_3, which we measure by gauging the approximate percentage differences $\lambda_1 = 100(\ln(\delta_1) - \ln(\delta_3))$ and $\lambda_2 = 100(\ln(\delta_2) - \ln(\delta_3))$. Here δ_3 serves as a benchmark that accounts for all sources of variability, which is compared to δ_1 and δ_2, which do not do so. We consider random subsamples of size 100, 500, and 1000, in addition to the full data sample and several specific subsamples of interest in policy analysis. As the sample size n increases, the first 4 rows of Table 2 indicate that δ_2 moves closer to δ_3 as seen by the decrease in the absolute values of λ_2. From the table we also see that the behavior of δ_1 across the three specifications of y_i is more erratic and convergence of δ_1 to δ_3 is not guaranteed to occur at all, or to occur quickly, as the sample size n increases. As expected, in larger samples the parameters are estimated more precisely and the draws of β used to compute δ_3 in equation (6) are close to $\hat{\beta}$ that appears in the computation of δ_2 in equation (5) leading to smaller discrepancies; however, even in well-behaved cases (e.g., when $y_i = 1\{\text{some college}\}$) and large samples, δ_1 need not converge to δ_3 because $\bar{x}$ in equation (4) may not converge to anything useful or sensible.

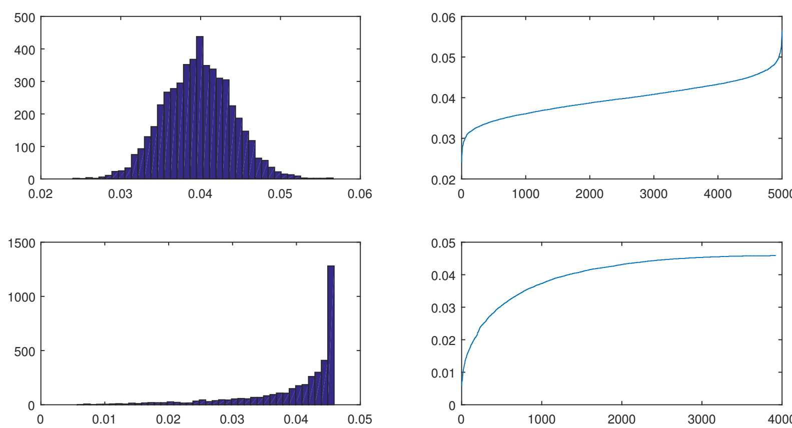
In irregular settings, e.g., in the presence of ill-conditioned covariate matrices, data sets with unbalanced outcomes, or weakly identified models, δ_1 could be affected severely, and even the effect on δ_2 may be quite large. This can be seen from rows 5-9 in Table 2, where the sample is restricted to study such scenarios. Column 3 of the table presents the condition number (C) of the data matrix. In general, a high condition number reflects high dependencies in the data, implying a poorly-conditioned data matrix. Confining the sample to individuals who have educated parents leads to large differences between δ_1 and δ_3 and δ_2 and δ_3. This subsample is characterized not only by correlation in the covariates, but also contains a high degree of classification, especially when $y_i = 1\{\text{high school degree}\}$ because educated parents almost always have children who have at least a high school degree. Therefore, it is not surprising that this is where we see the largest differences even when the sample size is relatively large. Rows 5 and 6 of the table illustrate how the covariate effect differences vary as we restrict the sample to higher levels of parental education. Rows 7, 8, and 9 show similar developments as the sample is restricted along gender, racial, and parental education lines that have been of interest in earlier work. In these cases, ignored parameter uncertainty can be seen to produce significant biases – even though the samples with educated mothers are often large, Table 2 displays that the bias of δ_2 can be upwards of 22%. This particular segment of the population – young individuals who have educated mothers – is a widely studied area in labor economics, so that employing a significantly biased result such as this, can have major policy implications and lead to ill-informed decisions. Some policy considerations are examined in section 2.1.2.

TABLE 2. Percentage differences $\lambda_1 = 100(\ln(\delta_1) - \ln(\delta_3))$ and $\lambda_2 = 100(\ln(\delta_2) - \ln(\delta_3))$ for the covariate effect of family income for several subsamples.

			High School		Some College		College	
Subsample	n	C	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
Random	100	121	3.1	3.3	26.6	7.0	−198.7	−5.3
Random	500	101	2.9	0.9	15.3	1.8	2.7	0.8
Random	1000	101	4.4	0.5	12.7	0.8	2.1	0.4
Full	3923	101	5.6	0.2	11.7	0.2	1.0	0.1
Father College	500	309	−56.5	−7.5	3.5	0.8	14.4	1.7
Father College, Mother High School	468	318	−63.7	−14.0	−0.5	0.6	11.4	1.9
Female, Mother Some College	284	236	−448.4	−22.1	−0.2	1.3	22.3	2.7
Black, Mother Some College	124	242	−262.7	−8.5	16.9	4.1	21.4	5.9
Black, Parents < High School	336	98	4.6	2.3	4.5	1.9	−13.7	−4.8

As discussed at the end of Section 1, confidence bands for δ_3 are not necessary as there is no residual uncertainty that can be represented. However, two marginal distributions can be obtained as a by-product of the draws of δ_3, which may be of interest in certain contexts. For the attainment of some college, Figure 2 represents the distribution of the covariate effect of income (i) as a function of parameter uncertainty and (ii) over the units of the full NLSY sample. Separately integrating over the data and parameter distributions elicits aspects of the data and estimation uncertainty that would otherwise not be available with δ_1 or δ_2.

FIGURE 2. Distribution of the average effect as a funciton of (i) parameter uncertainty (top two panels) and (ii) the units in the sample (bottom two panels); in each case, a histogram and a plot of the ordered values are presented.



2.1.2. *Policy Analysis.* In this section, we examine the differences that the three covariate effect estimates—δ_1, δ_2, and δ_3—would have on policy implications. Table 3 reports the three covariate effect estimates for the income variable and shows that, in all instances, δ_3 implies a higher impact of the income variable on the the probability of achieving a high school degree than δ_1 and δ_2. A comparison between δ_2 and δ_3 indicates that, in this sample, at least one percentage point difference will emerge due to failure to account for parameter uncertainty. The sign of the difference also indicates that, on average, these subsamples of individuals are on the convex part of the link function and that the presence of important increasing returns will be missed by covariate effect estimates that ignore parameter variability. Finally, comparing δ_1, δ_2 and δ_3 in the third row of Table 3, leads to dramatic differences in policy recommendations: at one end of the spectrum, the value of δ_1 suggests that conditionally upon advanced maternal eduction, income does not affect the probability of finishing high school for minorities, while δ_2 and δ_3 clearly reveal that income still plays an important role in completing high school.

2.1.3. *Additional Considerations.* The results from our probit application indicate that differences in covariate effect estimates can occur not only in small samples,

Table 3. The covariate effect of income on obtaining at least a high school degree for several subsamples.

	Results		
Subsample	δ_1	δ_2	δ_3
Father College	0.094	0.153	0.165
Father College, Mother High School	0.068	0.112	0.129
Black, Mother Some College	0.009	0.117	0.127

but also in larger samples with ill-conditioned covariate matrices, high degree of classification, unbalanced outcomes, and so on. We now focus on the effect that the variability of the covariates x can have on δ_1 and δ_2, and do this by constructing a simulation study as it is not straightforward to change the variance of covariates while holding their coefficients fixed in a real data application. For the simulation study, the covariate vector x_i, $i = 1, \ldots, n$, includes an intercept and 2 continuous variables which are generated from the distribution $N\left(0, \sigma_x^2\right)$, where σ_x^2 is set to 0.5, 1, and 5 to explore the effect of changes in the variability in the regressors. We let $\beta = (0, 0.2, 0.3)'$ and $y_i = 1\left\{x_i'\beta + \varepsilon_i > 0\right\}$, where $\varepsilon_i \sim N(0,1)$. The effect of the sample size on the point estimates is again studied by setting n to 100, 500, and 1000. The results of the simulation study, presented in Table 4, are quite interesting and somewhat unexpected, but can be easily rationalized. In particular, δ_1 improves as σ_x^2 decreases and δ_2 improves as σ_x^2 increases. The former is due to the fact that with lower variability of the covariates, replacing them by their mean (as done in computing δ_1) is not as detrimental as it would be otherwise; the latter is due to the fact that as σ_x^2 increases, large values of $x_i'\beta$ lead to small values of $\phi(x_i'\beta)$, which in turn implies that the impact of β_j is small, regardless of whether it varies (as in δ_3) or not (as in δ_2). It is interesting to note that even though most of the differences diminish with larger samples, some remain relatively stable across sample sizes.

Table 4. Percentage differences $\lambda_1 = 100(\ln(\delta_1) - \ln(\delta_3))$ and $\lambda_2 = 100(\ln(\delta_2) - \ln(\delta_3))$ for the covariate effect of x_2 in the simulated data over 25 Monte Carlo replications.

	$\sigma_x^2 = 0.5$		$\sigma_x^2 = 1$		$\sigma_x^2 = 5$	
n	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
100	19.4	14.4	13.4	3.5	26.6	2.7
500	4.1	0.8	7.0	0.7	27.3	0.5
1000	4.4	0.4	7.1	0.4	33.9	0.3

In closing, we note that even though in the probit example the link function is symmetric and its shape is such that extreme values of $x_i'\beta$ lead to low values of the covariate effects, in other cases this need not be true. One important case is the Poisson model, which is considered next.

2.2. **Poisson Example.** The Poisson link $\exp(x_i'\beta)$ is asymmetric and exhibits an increasing derivative, whereby the effects of covariate and estimation uncertainty may play a strong role. For this model, the marginal effects for the jth component of x_i are

$$\begin{aligned}\delta_{j1} &= \exp(\bar{x}'\hat{\beta})\hat{\beta}_j,\\ \delta_{j2} &= \overline{\exp(x_i'\hat{\beta})}\hat{\beta}_j,\\ \delta_{j3} &= \overline{\exp(x_i'\beta)\beta_j},\end{aligned}$$

which can be contrasted with the the probit effects in equations (4), (5), and (6). In this example, we re-evaluate earlier findings in a new context in which, because of the global convexity of the Poisson link, δ_1 and δ_2 should exhibit negative bias.

The data set for our example comes from [6], which was made available by [11]. It was constructed from arrest records maintained by the California Department of Justice and earnings records from the California Employment Development Program for a random sample of men born in 1960-1962. The dependent variable y_i is the number of arrests, whereas covariates include prior convictions, average sentence length, time in prison, number of quarters employed, income (measured in tens of thousands of dollars), and race indicators for black and Hispanic. Parameter estimation was performed by Accept-Reject Metropolis-Hastings simulation (see [8], [2], [3]), and estimates were confirmed by maximum likelihood (which is an input in the simulation algorithm). In this example, we focus on the effect of income on number of arrests.

The results are presented in Table 5 for the full data sample and several different subsamples (in the table, C is used again to denote the condition number of the data matrix). The results again point to a number of instances of practically relevant biases, especially in small samples and the specific subsamples in rows 5-9 of Table 5. Note that, unlike the results in Table 2, the sign of the differences in Table 5 is always negative, as theoretically predicted due to the convexity of the link function. While knowing the sign of the bias is helpful, it does not mitigate the dangers of ignoring parameter uncertainty, as its magnitude will generally be context-specific.

TABLE 5. Percentage differences $\lambda_1 = 100(\ln(\delta_1) - \ln(\delta_3))$ and $\lambda_2 = 100(\ln(\delta_2) - \ln(\delta_3))$ for the covariate effect of income for several subsamples.

			Results	
Subsample	n	C	λ_1	λ_2
Random	200	286	−33.7	−6.6
Random	400	304	−25.0	−3.0
Random	1000	286	−20.9	−1.2
Full	2725	288	−21.2	−0.4
Previously imprisoned	134	248	−47.3	−6.1
Black	439	192	−15.0	−1.0
Hispanic and previously imprisoned	32	271	−111.0	−13.4
White and previously imprisoned	54	323	−1113.6	−16.4
Previously imprisoned for more than 1 year	77	229	−80.5	−10.4

2.2.1. *Policy Analysis.* We now focus on the implications of δ_1, δ_2, and δ_3 for setting policy. In particular, while Table 5 presents the biases that result from the different covariate effect approaches, this section considers their implications in practice. Table 6 presents the actual values of δ_1, δ_2, and δ_3 for the effect of our income on the number of arrests. For white individuals who have been previously imprisoned, the increase in income is predicted to reduce the number of arrests by 2.38. The interpretation of this result is vastly different than that of δ_1, which shows nearly no effect and could mislead policy makers into believing that income for white, previously imprisoned individuals is not pertinent. However, once parameter and sampling variability are accommodated (as in δ_3), income has a very large effect – actually the largest among the race groups considered. This result suggests that policies aimed at creating employment opportunities for inmates can be an important tool for reducing the number of future arrests and crimes. The table also shows that $\delta_2$is much closer to δ_3 than δ_1, the discrepanies are still meaningful from a policy standpoint.

Table 6. The covariate effect of income ($10,000) on number of arrests for several subsamples.

	Results		
Subsample	δ_1	δ_2	δ_3
Previously imprisoned	-0.54	-0.81	-0.87
Hispanic and previously imprisoned	-0.39	-1.05	-1.20
White and previously imprisoned	0.00	-2.02	-2.38
Previously imprisoned for more than 1 year	-0.78	-1.57	-1.75

3. Conclusion

This paper has considered the problem of assessing covariate effects in nonlinear models. Two traditional approaches employing point estimates of the parameters—one evaluating the effect at the average of the covariates and the other averaging the effect over the observations in the sample—are contrasted with an estimator that also accounts for estimation uncertainty. This uncertainty should be accounted for by integrating over the distribution of the model parameters. Our study shows that even though estimation uncertainty tends to diminish with larger samples, there are instances where failure to incorporate it in covariate effect estimation can lead to significant biases, even in large samples. For this reason, we advocate that both Bayesian and frequentist researchers report covariate effects that account for estimation uncertainty (in addition to covariate variability) when reporting their findings or make policy recommendations.

References

[1] Brownstone, David (2001), "Discrete Choice Modeling for Transportation," *Travel Behaviour Research: The Leading Edge*, 97-124.

[2] Chib, S. and E. Greenberg (1995), "Understanding the MetropolisHastings algorithm,"*American Statistician*, 49, 327-335.

[3] Chib, S. and I. Jeliazkov (2005), "Accept-reject Metropolis-Hastings sampling and marginal likelihood estimation,"*Statistica Neerlandica*, 59, 30-44.

[4] Chib, S. and I. Jeliazkov (2006), "Inference in Semiparametric Dynamic Models for Binary Longitudinal Data," *Journal of the American Statistical Association*, 101, 685-700.
[5] Greene, William, *Econometric Analysis*, 6th Edition. New Jersey: Prentice Hall, 2008.
[6] Grogger, Jeffrey(1991), "Certainty vs. Severity of Punishment," *Economic Inquiry*, 29, 297-309.
[7] Jeliazkov, I., J. Graves, and M. Kutzbach (2008), "Fitting and Comparison of Models for Multivariate Ordinal Outcomes," *Advances in Econometrics: Bayesian Econometrics*, 23, 115-156.
[8] Tierney, L. (1994), "Markov chains for exploring posterior distributions"(with discussion), *Annals of Statistics*, 22, 1701-1762.
[9] Verlinda, J. A. (2006), "A Comparison of Two Common Approaches for Estimating Marginal Effects in Binary Choice Models," *Applied Economics Letters*, 13, 77-80.
[10] Wooldridge, Jeffrey, *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press, 2002.
[11] Wooldridge, Jeffrey, *Introductory Econometrics: A Modern Approach*, 4th edition. Mason, OH: South-Western, 2009.