

# Behind the Curve and Beyond: Calculating Representative Predicted Probability Changes and Treatment Effects for Non-Linear Models

Bastian Becker<sup>1</sup>

## Abstract

Parameter coefficients from non-linear models are inherently difficult to interpret, and scholars frequently opt for computing and comparing predicted probabilities for variables of interest. In an influential article, Hanmer and Ozan Kalkan (2013) discuss the two most common approaches, the *average case* respectively *observed values* approach, and make a strong case for the latter. In this paper, I propose a further refinement of the observed values approach for the purpose of computing predicted probability changes. This refinement concerns the use of counterfactual values for the independent variable of interest. I demonstrate that accounting for non-linearities with regards to the variable of interest is important to avoid estimation biases. I also discuss the implications of this insight for estimating average treatment effects from observational data.

## 1 Introduction

Parameter coefficients from non-linear models, as they are commonly used for categorical dependent variables, are inherently difficult to interpret. Therefore, it is preferable to opt for computing and comparing predicted probabilities for different values of independent variables (Brambor et al., 2006; Gelman and Pardoe, 2007; Berry et al., 2010).<sup>2</sup> In an influential article, Hanmer and Ozan Kalkan (2013) discuss the two most common approaches to computing such probabilities, the *average case* approach (ACA) respectively *observed values* approach (OVA), and make a strong case for the latter.<sup>3</sup> In essence, predicted probabilities for average cases are not representative for the population at large as they ignore the distribution of covariates, which leads to biased predictions based on models with non-linear functional forms (e.g. logit, probit, poisson). They also present evidence in favor of their claim that the ACA exaggerates probability changes attributed to any independent variable.

---

<sup>1</sup>Bard College Berlin, Berlin, Germany; b.becker@berlin.bard.edu

<sup>2</sup>It can be argued that odds ratios constitute a preferable way to interpret probability models (see, for example, Liao (1994)). However, I focus on predicted probabilities as their use is more common in the current literature.

<sup>3</sup>According to Google Scholar the article has been cited 307 times, 32 of which are publications in the AJPS, APSR, BJPS, CPS, JOP, and PSRM (as of March 13, 2018).

The main uses of predicted probabilities for limited dependent variables are three. First, interest in the predicted probabilities themselves, either at a specific value of an independent variable or for a range of values (e.g. effect plots). Second, an interest in predicted probability changes, which compares predicted probabilities at two different values of an independent variable, to give an indication of effect sizes or treatment effects. Or third, combining the first two uses by computing predicted probabilities *changes* for one independent variable conditional on values of another independent variable (e.g. marginal effect plots). In this paper, I focus on the computation of predicted probability changes—which are central to the two latter uses just mentioned—for logit models with binary dependent variables, but the results generalize to other non-linear models.

Specifically, I propose a refinement of the observed values approach as formalized by Hanmer and Ozan Kalkan (2013). This refinement remains true to the authors' original motivation, which is to compute representative probability changes by accounting for the distribution of independent variables. However, in addition to considering the distribution of covariates, the refinement suggested here also accounts for the distribution of the independent variable for which the probability changes are computed. Furthermore, I show how the same refinement can be applied to estimates of (average) treatment effects based on observational data. I also present results from a simulation analysis to show how the refinement I suggest avoids inducing biases in the estimation of predicted probability changes.

## 2 Predicted Probabilities for Non-Linear Models

### 2.1 The Observed Values Approach (OVA)

The logic underlying models with binary dependent variables is that the distribution of observed outcomes  $Y$  is a chance event, and that each observation  $i$  has an unobserved probability of either outcome, i.e.  $\Pr(1) = p_i$  and  $\Pr(0) = 1 - p_i$ . The full vector of probabilities  $p$  is linked to a matrix of independent variables  $X$  through a function  $G$  that specifies intercept  $\alpha$  and a vector of weights  $\beta$ ,  $G(p) = \alpha + X\beta$ . The most common functional forms are the *logit* and *probit*, and  $\alpha$  and  $\beta$  are estimated such that the resulting probabilities  $p$  maximize the joint likelihood of the model having generated the observed outcomes  $Y$  (Long, 1997). In the following, I focus on the logit model only. The focus on logit models is purely for exemplary purposes and due to their common usage and relative ease of use.<sup>4</sup> Most importantly, the issues raised in the following apply to *all* commonly used models with non-linear link-functions.

The interest here lies in predicted probability changes given a discrete change in an independent variable  $x_k$  from  $c$  to  $d$ . Under a logit model, predicted probabilities can be produced for any observation (or counterfactual case) by applying the inverse logit function  $F$ ,  $p = F(\alpha, \beta, x) = \frac{1}{1 + e^{-(\alpha + x\beta)}}$ . It is possible to separate the contribution of the independent variable  $x_k$  from all other independent variables  $X_{-k}$ , such that the inverse logit is re-expressed as  $\frac{1}{1 + e^{-(\alpha + X_{-k}\beta_{-k} + x_k\beta_k)}}$ . The predicted probability change can then be computed as follows,

<sup>4</sup>It is often noted that logit and probit models do not yield substantively different results.

$$\Delta p_i = \frac{1}{1 + e^{-(\alpha + X_{i-k}\beta_{-k} + d\beta_k)}} - \frac{1}{1 + e^{-(\alpha + X_{i-k}\beta_{-k} + c\beta_k)}}. \quad (2.1)$$

On the right-hand side, all values but covariates  $X_{-k}$  have already been set. Here the difference between the ACA and the OVA becomes most clear. Under the ACA,  $X_{-k}$  would be filled in by  $\bar{X}_{-k}$ , i.e. modal, median, and mean values for each respective variable. By contrast, for the OVA, we compute  $\Delta p_i$  for every observation  $i$  by filling in the respective  $X_{-k}$  values, and the average over the result. As such, predicted probability changes under OVA are better expressed as,

$$\Delta p_{OVA} = \frac{1}{n} \sum_{i=1}^n \Delta p_i, \quad (2.2)$$

where  $n$  indicates the number of observations, and  $X_{-k}$  constitutes a matrix with one row for each observation, and one column for each independent variables (excluding the  $k$ th).

Both approaches, due to the non-linearity induced by the (inverse) logit function, lead virtually never to the same result. As elaborated by Hanmer and Ozan Kalkan (2013) the reason for this is that the effect of  $\beta_k$  is not constant but depends on an observation's values for the independent variables. This can be seen when taking the first derivative of the inverse link function  $F$  with respect to  $x_k$ , which indicates the marginal effect. For the logit model it is (following the expanded formulation of  $F$ ),

$$\frac{\partial p}{\partial x_k} = \beta_k \frac{e^{\alpha + X_{-k}\beta_{-k} + x_k\beta_k}}{(1 + e^{\alpha + X_{-k}\beta_{-k} + x_k\beta_k})^2}. \quad (2.3)$$

As explained above, ACA and OVA differ in their choice of  $X_{-k}$  for which the predicted probabilities are computed. As the first derivative of the average case usually does not have the same value as the average derivative of all cases, they lead to different results. As mentioned before, this concern is not limited to logit models but similarly applies to other non-linear models, where generally,  $\frac{\partial p}{\partial x_k} = f(x\beta)\beta_k$ .

In sum, Hanmer and Ozan Kalkan (2013) make a convincing case for the OVA. In particular, they show that accounting for the observed variability in the independent variables is important to arrive at representative results. However, the authors also warn, "it is crucial to note that the observed-value approach is not foolproof. Though our focus is on setting the variables not being manipulated, setting the values of the variable being manipulated is at least as important" (Hanmer and Ozan Kalkan, 2013, p. 269). Put differently, the authors do not elaborate the choice of counterfactuals for the independent variable of interest,  $x_k$ . This is the point of the departure for the present paper.

## 2.2 Problem: Choosing Counterfactuals

In explicating the OVA, Hanmer and Ozan Kalkan (2013) do not focus on the choice of counterfactual values for the independent variable of interest,  $x_k$ , i.e. the variable whose values are adjusted to compute predicted probability changes. However, this choice is important since, like the choice of  $X_{-k}$ , it affects the value of the first derivative of  $F$  and thus the quantities we are interested in, i.e. predicted probability changes. A common approach in the literature is to use fixed values as counterfactual values (e.g. King and

Zeng, 2006; Aronow and Samii, 2016). I show below that this approach undesirably distorts computations of predicted probability changes, and that counterfactuals for the OVA are better chosen with respect to each observation's value of  $x_k$ . As such, what is offered here is a refinement of the original OVA for the purpose of computing predicted probability changes.

**Table 1:** Common counterfactuals

Counterfactual	$c$	$d$
Interquartile difference	1st quartile of $x_k$	3rd quartile of $x_k$
First differences	a (e.g. 0)	a+1 (e.g. 1)
Standard deviation	0	0 + SD( $x_k$ )
Full range	min( $x_k$ )	max( $x_k$ )

Under the OVA, predicted probability changes are computed based on equation 2.2. Some quantities are given.  $\alpha$  and  $\beta$  have been determined by a previously estimated logit model.  $X_{-k}$  contains the observed values for the corresponding variables. What is left are the counterfactual values of  $c$  and  $d$  which replace the observed values of  $x_k$ . Looking at the empirical literature, counterfactuals that are frequently chosen correspond to interquartile differences, first differences, standard deviations, or the full range of the independent variable of interest. Table 1 shows the exact values for each choice of counterfactual. Having chosen a counterfactual, and following the standard OVA, the respective values are plugged into equation 2.2, and the result indicates the predicted probability change in the dependent variable.

As discussed above, the main advantage of OVA over ACA is that it does not replace  $X_{-k}$  with values of one representative case only,  $\bar{X}_{-k}$ , but takes into account the distribution of all  $X_{-k}$ . As such, it accounts for the distribution of covariates and how they condition the marginal effect of  $x_k$ ,  $\frac{\partial p}{\partial x_k}$ . However, the same concern also applies to  $x_k$  itself (Gelman and Pardoe, 2007). The counterfactual value,  $x_c$ , by which  $x_k$  is replaced affects the distribution of marginal effects. This is not a concern if our interest lies in the marginal effect at a specific counterfactual value. However, if the interest is in predicted probability changes that are representative of the population, it is. Replacing all  $x_k$  values with a single  $x_c$  value ignores variation in  $x_k$  as well as covariation of  $x_k$  with other independent variables ( $X_{-k}$ ), and the resulting variation in  $\frac{\partial p}{\partial x_k}$ . The consequences of this for predicted probability changes depend on whether the replacement shifts observations—on average—towards higher or lower marginal effects.

Importantly not all observation are affected by these “shifts” in the same way. In particular, as shifts  $s$  are the difference between the observed value and the counterfactual value,  $x_c - x_k$ , they depend on the observed value itself, and their distribution,  $S$ , is unknown. Thus, when single counterfactual values are used, predicted probability changes are effectively determined according to the following equation (as  $s = x_c - x_k$ ,  $x_c = s + x_k$ ),

$$\Delta p_{OVA} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{1 + e^{-(\alpha + X_{-k} \beta_{-k} + (x_k + s_d) \beta_k)}} - \frac{1}{1 + e^{-(\alpha + X_{-k} \beta_{-k} + (x_k + s_c) \beta_k)}} \right), \quad (2.4)$$

where  $s_c$  and  $s_d$  are the shifts induced by the respective counterfactual values,  $c$  and  $d$ . As the distribution of the shifts, and thus its consequences for computations of predicted probability changes, are unknown, this appears to be an undesirable set-up. Fortunately, there is a rather straightforward way to avoid such shifts.

### 2.3 Solution: Location-Sensitive OVA

As outlined in the previous section, counterfactuals are most commonly computed based on fixed values. This practice is likely a consequence of counterfactual computations with coefficients from linear models, where the choice of fixed values and the shifts it induces does not affect results (due to the constance of marginal effects). For non-linear models, such shifts should be avoided and this can be done by setting counterfactual relative to observed values. Say a predicted probability change is to be computed for a change in  $x_k$  from  $c$  to  $d$ . Instead of replacing all  $x_k$  by the fixed values (as above), counterfactuals are computed for each observation with respect to its observed  $x_k$ . This can be done by computing the difference between  $c$  and  $d$ , and setting each observation's counterfactuals,  $c_i$  and  $d_i$ , to  $x_k - (d-c)/2$  and  $x_k + (d-c)/2$  respectively. Employing these counterfactuals that are sensitive to each observation's  $x_k$  value leads to the following reformulation of the equation for predicted probability changes,

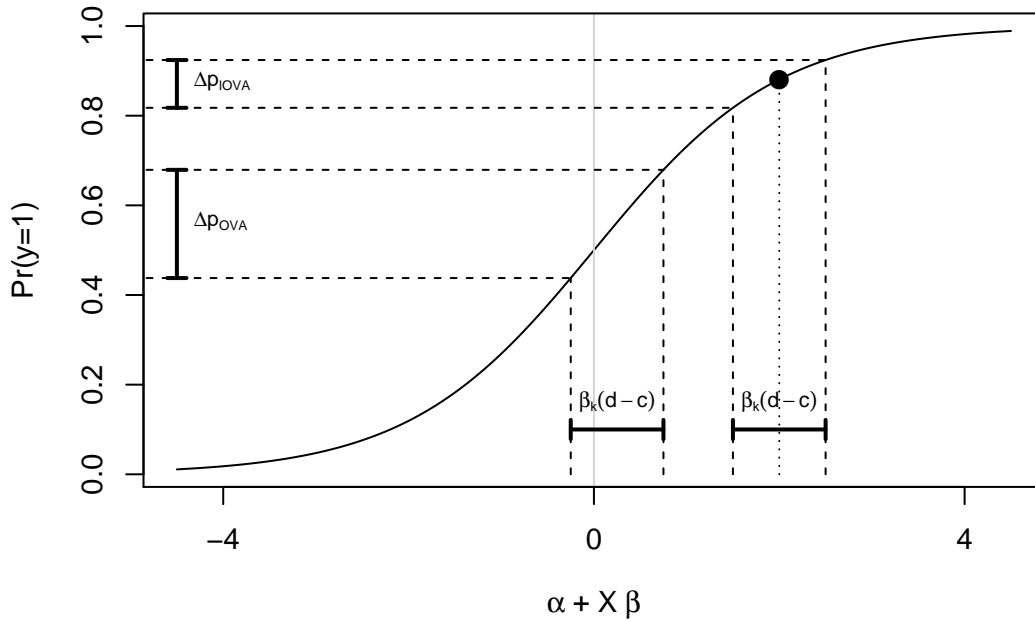
$$\Delta p_{IOVA} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{1 + e^{-(\alpha + X_{-k}\beta_{-k} + (x_k + (d-c)/2)\beta_k)}} - \frac{1}{1 + e^{-(\alpha + X_{-k}\beta_{-k} + (x_k - (d-c)/2)\beta_k)}} \right). \quad (2.5)$$

The location-sensitive OVA (IOVA) for predicted probability changes thus considers the same counterfactual change in  $x_k$  counterfactual range for all observations,  $d - c = d_i - c_i \forall i$ . Furthermore, it does so relative to the originally observed value and thus avoids unnecessary distortion due to shifts induced by the usage of fixed counterfactuals. Figure 1 illustrates the difference between both approaches. In the next section, I show that such a correction can have substantial impact on assessment of predicted probability changes.

## 3 Simulation Analysis

In the following simulation analysis I explore differences between OVA and IOVA for computing predicted probability changes using different counterfactuals. In particular, I discuss differences for counterfactuals based on interquartile differences, first differences, and standard deviations. Counterfactuals based on the full range of the independent variable are excluded here as such predictions are highly sensitive to outliers, and often highly dependent on model specification (King and Zeng, 2006); they should therefore be avoided (unless researchers are interested in the outliers themselves).<sup>5</sup>

<sup>5</sup>The analysis is implemented with the R software environment. The *mvtnorm* package is used to generate population data. Replication files are provided through the journal website.



**Figure 1:** Predicted Probability Changes: OVA vs IOVA (illustration). The black dot corresponds to an hypothetical observed value. Here the shift induced by OVA leads to a higher predicted probability change (for a hypothetical counterfactual change in  $x_k$  from  $c$  to  $d$ ) than under IOVA.

### 3.1 Set-Up

To assess the implications of using OVA with fixed or relative counterfactuals for assessments of predicted probability changes, I simulate population data under different realistic parameter specifications (scenarios), estimate logit models based on a random sample of the simulated population data, and then compare results based on the standard OVA and the IOVA.<sup>6</sup> For each scenario, I simulate 250 populations with 1 000 000 observations. From each population, I draw a random sample of 1000 observations. Simulating both populations and corresponding samples has the advantage that population parameters can be computed directly and used to evaluate sample estimates. The data-generating process (DGP) of the population data is as follows,

$$y \sim \text{B}(1, \pi),$$

$$\pi = \text{F}(\alpha + \beta_1 x_1 + .5x_2 + 2x_3).$$

The DGP is based on the logit model, where the binary dependent variable,  $y$ , is linked to the independent variables through the Binomial function,  $\text{B}$ , and the inverse logit function,  $\text{F}$ .  $\pi$  corresponds to each observation's probability of  $y$  equaling 1 (the probability of  $y$  equaling 0 is  $1 - \pi$ ). There are three independent variables,  $x_1$ ,  $x_2$ , and  $x_3$ , are drawn from a multivariate normal distribution, where the standard deviations of  $x_2$  and  $x_3$  are set to 1, and the correlation coefficients  $\rho(x_1, x_3)$  and  $\rho(x_2, x_3)$  are set to 0. The remaining parameters are altered under the different scenarios, i.e. the parameter

<sup>6</sup>This set-up is similar to the one employed by Hanmer and Ozan Kalkan (2013).

coefficient  $\beta_1$  takes values .5, 1, and 2, the standard deviation  $\sigma_1$  of  $x_1$  takes values .5, 1, and 2, the correlation coefficient  $\rho(x_1, x_2)$  takes either 0 or .5, and the intercept  $\alpha$  is either 0 or 1. Simulations are run for all possible combinations of these parameter values. Table 2 gives an overview of all 36 simulation scenarios. Predicted probability changes are later evaluated for  $x_1$ .

**Table 2:** Simulation Scenarios (Specification of DGP). The event rate,  $1/(1 + e^{-\alpha})$ , of scenarios 1–18 is 50.0%, and 73.1% for scenarios 19–36.

Scenario	$\beta_1$	$\sigma_1$	$\rho(x_1, x_2)$	$\alpha$	Scenario	$\beta_1$	$\sigma_1^2$	$\rho(x_1, x_2)$	$\alpha$
1	0.5	0.5	0.0	0.0	10	0.5	0.5	0.5	0.0
2	1.0	0.5	0.0	0.0	11	1.0	0.5	0.5	0.0
3	2.0	0.5	0.0	0.0	12	2.0	0.5	0.5	0.0
4	0.5	1.0	0.0	0.0	13	0.5	1.0	0.5	0.0
5	1.0	1.0	0.0	0.0	14	1.0	1.0	0.5	0.0
6	2.0	1.0	0.0	0.0	15	2.0	1.0	0.5	0.0
7	0.5	2.0	0.0	0.0	16	0.5	2.0	0.5	0.0
8	1.0	2.0	0.0	0.0	17	1.0	2.0	0.5	0.0
9	2.0	2.0	0.0	0.0	18	2.0	2.0	0.5	0.0
19	0.5	0.5	0.0	1.0	28	0.5	0.5	0.5	1.0
20	1.0	0.5	0.0	1.0	29	1.0	0.5	0.5	1.0
21	2.0	0.5	0.0	1.0	30	2.0	0.5	0.5	1.0
22	0.5	1.0	0.0	1.0	31	0.5	1.0	0.5	1.0
23	1.0	1.0	0.0	1.0	32	1.0	1.0	0.5	1.0
24	2.0	1.0	0.0	1.0	33	2.0	1.0	0.5	1.0
25	0.5	2.0	0.0	1.0	34	0.5	2.0	0.5	1.0
26	1.0	2.0	0.0	1.0	35	1.0	2.0	0.5	1.0
27	2.0	2.0	0.0	1.0	36	2.0	2.0	0.5	1.0

Based on the simulations for each scenario, I first compute the population parameters of interest. That is, I compute average probability changes,  $\Delta p_{IOVA}$ , as specified in equation 2.5, based on population data and the true model parameters of the DGP. To determine whether the OVA specification by Hanmer and Ozan Kalkan (2013) constitutes an unbiased estimator of the theoretically more consistent IOVA specification, I then compute predicted probability changes  $\widehat{\Delta p}_{OVA}$  based on sample data and the estimated model parameters.<sup>7</sup> The OVA approach was specified in equation 2.2. I compare OVA results directly to the predicted probability changes computed based on the IOVA approach,  $\widehat{\Delta p}_{IOVA}$  (equation 2.5).

### 3.2 Results

The simulation results for interquartile changes in the independent variable of interest  $x_1$  are presented in figure 2. Each plot corresponds to one scenario and presents the esti-

<sup>7</sup>The estimated model parameters,  $\hat{\alpha}$ ,  $\hat{\beta}_1$ ,  $\hat{\beta}_2$ , and  $\hat{\beta}_3$ , are determined based on a correctly specified model, with  $y \sim B(1, p)$ , and  $p = F(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$ .

mation error of both measures due to sampling. The specifications of each scenario are indicated on the top and left-hand side of the figure. The average population parameter is indicated in the top-right corner of each plot; the boxes (interquartile range) and whiskers (full range) summarize the distribution of the IOVA (white-shaded box) and OVA estimation errors (grey-shaded box) over the 250 simulations. Negative values indicate that for a given scenarios the estimate was below the population parameter, and positive values that the estimate was above. Note that the spread of the distribution indicates how spread out estimation errors are, it does not indicate the precision (or efficiency) of the estimate. Detailed numeric information on both bias and precision can be found in the Appendix (Table 3). As there was no significant difference in how the two measures perform in terms of precision, the presentation here focuses on estimation bias.

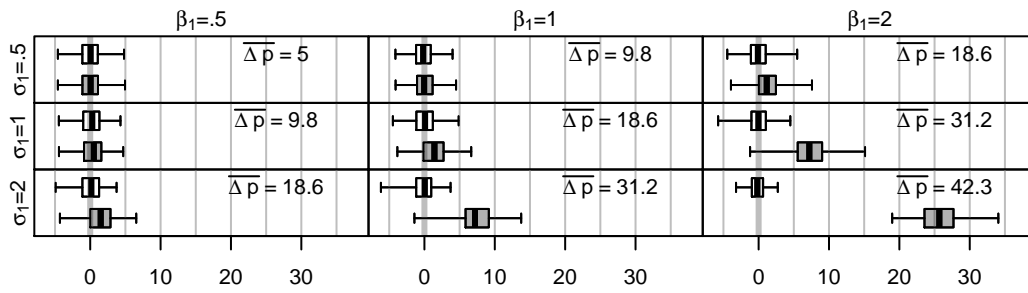
Sampling makes estimation error unavoidable. However, for an unbiased measure, such error will cancel out as more samples are drawn. In other words, in expectation the estimation error is zero. Figure 2 shows that, for interquartile change in the independent variable, estimates based on the IOVA are unbiased across all scenarios. While there are simulations which lead to estimation error, these error cancel out on average. While the OVA estimate is unbiased in some scenarios, it is heavily biased in others. In all scenarios where a bias is present, the OVA approach overestimates the true value. In the most severe scenarios, OVA overestimates the true value by a factor greater than 1.5 (e.g. Panel B, Scenario 18).

Bias in the OVA measure is largely absent when  $x_1$ , the independent variable for which probability changes are computed, plays a negligible role in determining an observation's value on the dependent variable, i.e.  $\beta_1$  is small. Similarly, bias is largely absent if  $x_1$  varies little, i.e.  $\sigma_1$  is small. As either factor,  $\beta_1$  or  $\sigma_1$ , increases, estimation bias increases too. The bias is particularly pronounced if both factors assume higher values. These biases are somewhat more pronounced when the independent variable of interest is correlated with another independent variable (i.e.  $\rho(x_1, x_2) = .5$  in scenarios 10-18 and 28-36). However, as can be seen from comparing scenarios 1 and 10, respectively 19 and 28, the correlation itself appears not to induce any bias. Similarly, intercept shifts also seem to moderate the biases induced by  $\beta_1$  and  $\sigma_1$ . In particular, in scenarios where the intercept  $\alpha$  is not 0 but 1 (scenarios 19-36) the biases are, to a small degree, attenuated.

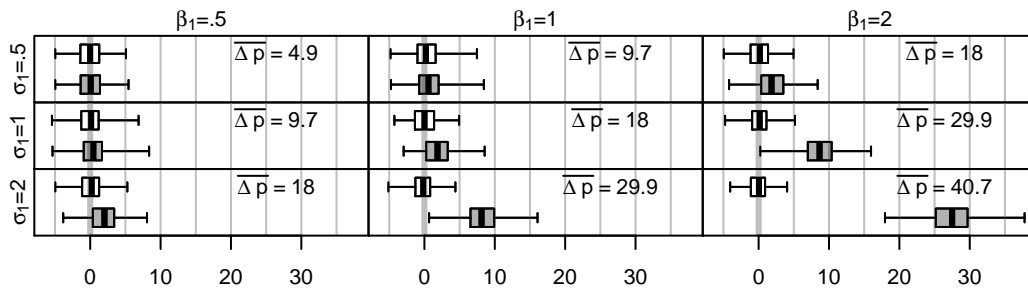
Figure 3 presents the simulation results for counterfactuals based on first differences, i.e.  $c = 0$  and  $d = 1$ . The set up of the figure is the same as above. While there are again no indications of bias in the IOVA measure, the biases of the OVA measure have different patterns. Most importantly, not all biases are positive. In particular, low variability in  $x_1$  combined with high predictive power and a non-zero intercept (i.e. scenarios 21 and 30) leads the OVA estimate to underestimate the true value. Unlike before, correlation with another independent variable and non-zero intercepts appear not to moderate these biases. Instead, in most scenario where  $x_1$  and  $x_2$  are correlated, there is a slight upward shift in the distribution of estimation error. The opposite holds for non-zero intercepts. As the intercept changes from 0 to 1, the distribution of estimation error is shifted downward in most scenarios.

The simulation results for standard deviations changes in the independent variable can be found in the Appendix (Table 4). The patterns mirror those of the interquartile changes, possibly because both counterfactuals take variability in the independent variable into account. That being said, overall estimation error is less pronounced, especially in scenarios

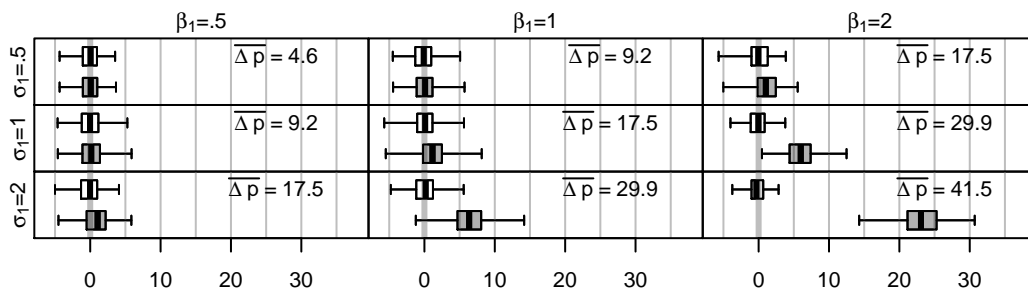




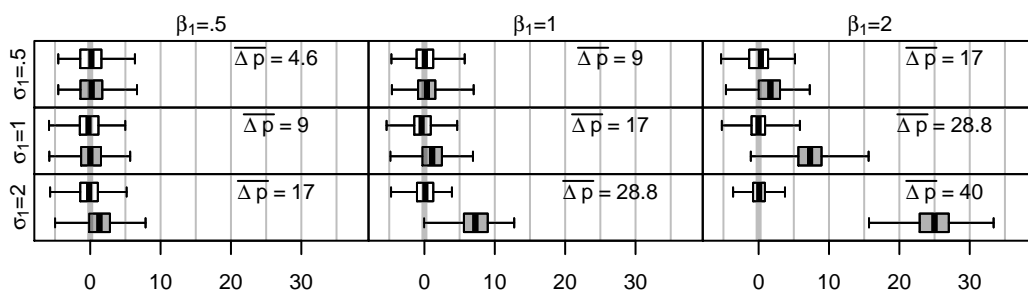
(a) Scenarios 1–9,  $\rho(x_1, x_2) = 0, \alpha = 0$



(b) Scenarios 10–18,  $\rho(x_1, x_2) = .5, \alpha = 0$



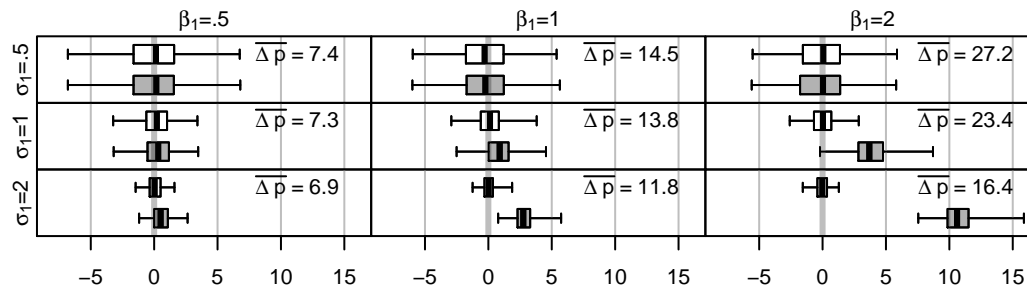
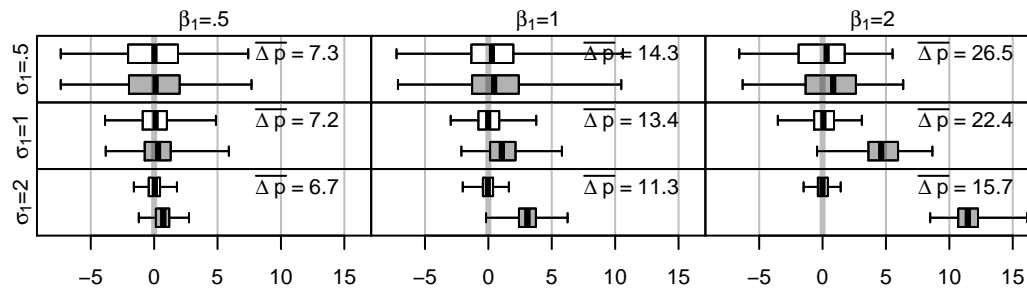
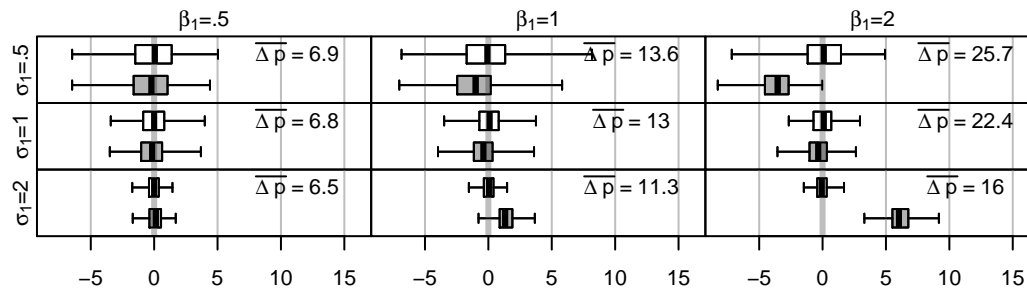
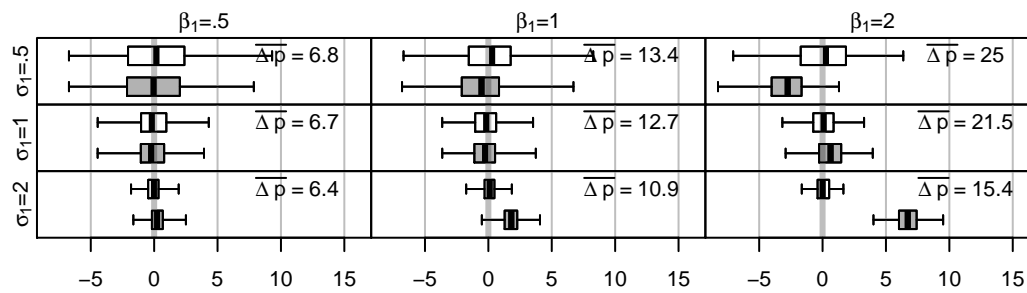
(c) Scenarios 19–27,  $\rho(x_1, x_2) = 0, \alpha = 1$



(d) Scenarios 28–36,  $\rho(x_1, x_2) = .5, \alpha = 1$

**Figure 2:** Estimation error in Predicted Probability Changes based on Interquartile Changes. Each plot corresponds to one scenario and 250 simulations. The average population parameter is indicated in the top-right corner. The boxes (interquartile range) and whiskers (full range) summarize the distribution of estimation errors for the IOVA (white-shaded box) and OVA estimates (grey-shaded box).

where the intercept is not zero.

(a) Scenarios 1–9,  $\rho(x_1, x_2) = 0$ ,  $\alpha = 0$ (b) Scenarios 10–18,  $\rho(x_1, x_2) = .5$ ,  $\alpha = 0$ (c) Scenarios 19–27,  $\rho(x_1, x_2) = 0$ ,  $\alpha = 1$ (d) Scenarios 28–36,  $\rho(x_1, x_2) = .5$ ,  $\alpha = 1$ 

**Figure 3:** Estimation error in Predicted Probability Changes based on First Differences. Each plot corresponds to one scenario and 250 simulations. The average population parameter is indicated in the top-right corner. The boxes (interquartile range) and whiskers (full range) summarize the distribution of estimation errors for the IOVA (white-shaded box) and OVA estimates (grey-shaded box).

## 4 Extensions

### 4.1 Assessing Average Treatment Effects

A common application of predicted probability changes is the computation of average treatment effects from observational data. In particular, scholars are interested in the average predicted change in an outcome variable  $y$  for a given treatment. For observational data, the average treatment effect is then defined as,

$$\text{ATE} = \frac{1}{N} \sum_1^N \frac{\text{E}(y|d_t, X_{-k}, \Theta) - \text{E}(y|d_c, X_{-k}, \Theta)}{d_t - d_c}, \quad (4.1)$$

where  $d_t$  and  $d_c$  are the counterfactual values for the treated and control group respectively, and the expected values of  $y$  are computed given these counterfactual values, covariates  $X_{-k}$ , and model parameters  $\Theta$ .  $d_t$  and  $d_c$  take values 1 and 0 in the case of a dichotomous treatment, but can take other values for continuous variables. In general, it is left to the discretion of the researcher to define those values, and it is most common to chose fixed counterfactual values (Aronow and Samii, 2016).

If the model on the basis of which parameters are estimated is non-linear, the use of fixed counterfactuals is equally problematic as in the standard OVA. Instead of using fixed values it is better to use counterfactuals that are defined relative to each observation's original value for the  $x_k$  variable that is manipulated by the treatment. This implies that  $d_t$  and  $d_c$  in the above equation have to be replaced with  $x_k + (d_t - d_c)/2$  and  $x_k - (d_t - d_c)/2$  respectively. Again, this approach has the advantage to compute treatment effects without inducing any untheorized distortions through the usage of fixed counterfactuals.

### 4.2 Interval Estimates

As for parameter coefficients, it is important to get an understanding of the uncertainty associated with measures of predicted probability change. To keep the presentation as simple as possible I have ignored such uncertainty above and focused on point estimates only. The delta method is the most prominent analytical approach to derive interval estimates from statistical models. However, the delta method has at least two major drawbacks here. First, it only approximates non-linear link functions and thus can lead to estimation errors. Second, the application of the delta method is complicated by the fact that IOVA, just like OVA, relies on comparisons of counterfactuals, and not on a single counterfactual. These problems can be avoided by the use of simulation-based approaches (King, Tomz, and Wittenberg, 2000).<sup>8</sup>

Bootstrapping is the most popular simulation-based approach to statistical estimation. Herron (2000) describes how confidence intervals can be bootstrapped through iterated random draws from the model parameters' estimated multivariate distribution. For each draw, the predicted probabilities of interest are computed. The resulting distribution describes the uncertainty in the respective measure and can be conveniently summarized by the use of confidence intervals, e.g. the central 95% of the distribution. The computation for IOVA measures proceeds analogously. After estimating the model, the researcher

<sup>8</sup>Hanmer and Ozan Kalkan (2013) also use a simulation-based approach to derive confidence intervals.

draws  $S$  (e.g. 5000) parameter samples from the respective multivariate distribution, and for each draw  $s$  computes equation 2.5 by plugging in the drawn  $\alpha$  and  $\beta$  parameter values. Following convention, the resulting distribution of predicted probability changes,  $\overline{\Delta p}_{IOVA}^s$ , is then best summarized by its average point estimate and 95%-confidence interval. If a single measure of uncertainty is preferred, one can also compute the standard error from the simulated distribution. Note that making inferences based on the standard error requires the researcher to assume that the average predicted probability change is normally distributed. Following this approach, the small program accompanying this paper simulates both confidence intervals and standard errors, and compares IOVA and OVA results.

## 5 Conclusion

In this paper, I have shown that combining the OVA with fixed counterfactuals for the independent variable of interest to compute predicted probability changes, and average treatment effects, induces distortions not theoretically motivated or understood. Therefore, I suggested a location-sensitive OVA (IOVA) that evaluates changes in variables of interest relative to observed values, which is theoretically sound and consistent with the motivation of the original OVA. With the help of a simulation analysis, I showed that under certain conditions IOVA leads to widely different results than when OVA is combined with fixed counterfactuals. More importantly, the distortions induced by the use of fixed counterfactuals can lead to both, positive and negative biases, depending on what the underlying data-generating process is and what counterfactuals are used. At the same time, the estimates based on the IOVA specification provided unbiased estimates in all scenarios and for all counterfactuals. In sum, both the theoretical discussion as well as the simulation analysis strongly recommend the use of relative rather than fixed counterfactuals when it comes to the estimation of average predicted probability changes based on non-linear models.

This paper is accompanied by a small program—for use with the statistical software *R*—that allows researchers to derive predicted probability changes for binary logit models using their own data. In addition to point estimates, researchers are usually interested in the uncertainty that accompanies them. Such uncertainty is most commonly summarized by standard errors or interval estimates. As one of the IOVA extensions, I discuss how bootstrapping, a simulation-based approach, can be used for the computation of standard error and interval estimates. The program automatically returns point estimates as well as both measures of uncertainty for all three counterfactuals discussed in this paper, i.e. first differences, standard deviation differences, and interquartile differences. The program also returns these quantities for the OVA, such that they can be directly compared to the IOVA results.

While the elaboration here was limited to continuous independent variables in non-linear models, similar concerns apply to nominal variables.<sup>9</sup> For ordered variables, the approach employed here suggests that instead of comparing two fixed categories, one should compute probability changes relative to an observation's observed category (e.g.

<sup>9</sup>With the exception of dichotomous variables, where the choice of counterfactuals is trivial.

from the observed category to the next). However, further research is needed to extend the approach developed here to non-continuous independent variables in non-linear models.

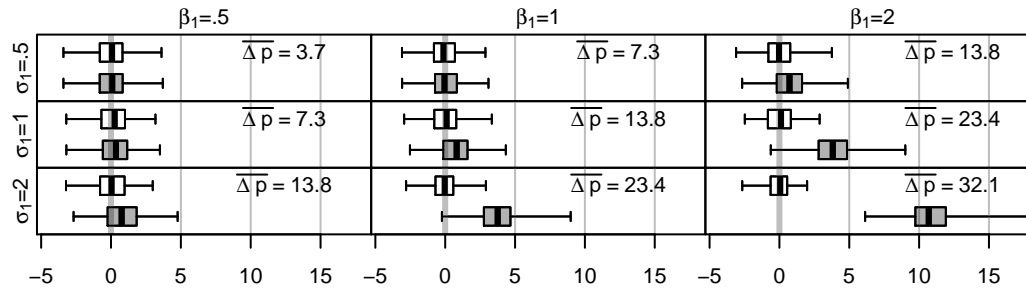
## Acknowledgment

I thank Love Christensen, Juraj Medzihorsky, two anonymous reviewers and editor Valentina Hlebec for helpful comments and suggestions, and Central European University for institutional support. All remaining errors are my own.

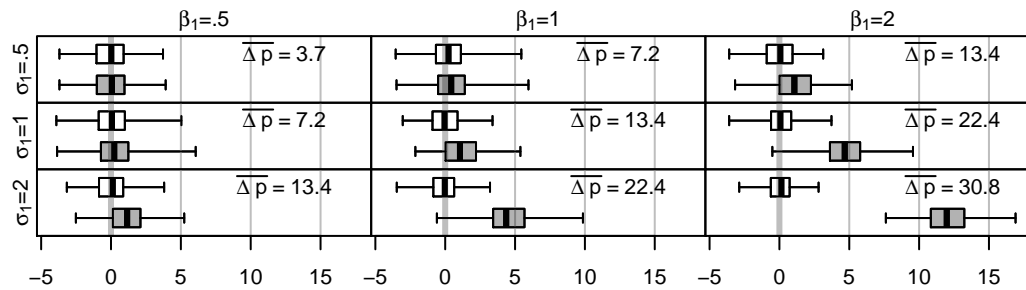
## References

- [1] Aronow, P. M. and Samii, C. (2016): Does regression produce representative estimates of causal effects? *American Journal of Political Science*, **60**(1), 250–267.
- [2] Berry, W. D., DeMeritt, J. H. R., and Esarey, J. (2010): Testing for interaction in binary logit and probit models: Is a product term essential? *American Journal of Political Science*, **54**(1), 248–266.
- [3] Brambor, T., Clark, W. R., and Golder, M. (2006): Understanding interaction models: Improving empirical analyses. *Political Analysis*, **14**(1), 63–82.
- [4] Gelman, A. and Pardoe, I. (2007): Average predictive comparisons for models with nonlinearity, interactions, and variance components. *Sociological Methodology*, **37**, 23–51.
- [5] Hanmer, M. J. and Ozan Kalkan, K. (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.
- [6] Herron, M. C. (2000): Postestimation uncertainty in limited dependent variable models? *Political Analysis*, **8**(1), 83–98.
- [7] King, G. and Zeng, L. (2006): The dangers of extreme counterfactuals. *Political Analysis*, **14**(2), 131–159.
- [8] King, G., Tomz, M., and Wittenberg, J. (2000): Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, **44**(2), 347–361.
- [9] Liao, T. F. (1994): *Interpreting Probability Models: Logit, Probit, and other Generalized Linear Models*. Thousand Oaks: Sage.
- [10] Long, J. S. (1997): *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.

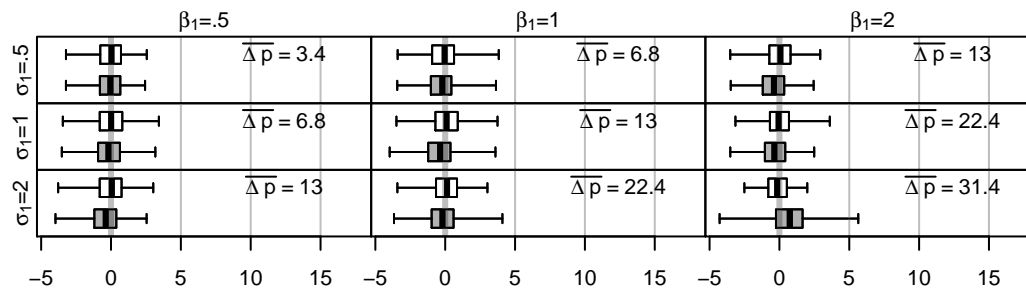
## Appendix



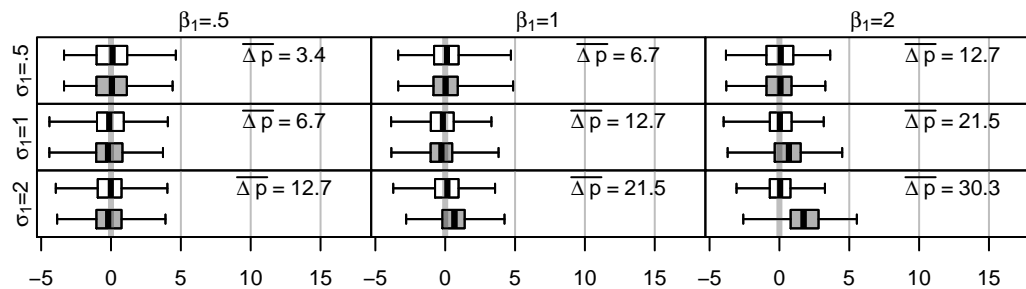
(a) Scenarios 1-9,  $\rho(x_1, x_2) = 0$ ,  $\alpha = 0$ .



(b) Scenarios 10-18,  $\rho(x_1, x_2) = .5$ ,  $\alpha = 0$ .



(c) Scenarios 19-27,  $\rho(x_1, x_2) = 0$ ,  $\alpha = 1$ .



(d) Scenarios 28-36,  $\rho(x_1, x_2) = .5$ ,  $\alpha = 1$ .

**Figure 4:** Estimation error in Predicted Probability Changes based on Standard Deviation Differences. Each plot corresponds to one scenario and 250 simulations. The average population parameter is indicated in the top-right corner. The boxes (interquartile range) and whiskers (full range) summarize the distribution of estimation errors for the IOVA (white-shaded box) and OVA estimates (grey-shaded box).

**Table 3:** Bias and Precision of all Estimates and Scenarios

Scenario	Bias						Precision (S.E.)					
	Interquartile		First difference		Standard deviation		Interquartile		First difference		Standard deviation	
	OVA	IOVA	OVA	IOVA	OVA	IOVA	OVA	IOVA	OVA	IOVA	OVA	IOVA
<b>Total</b>	4.960	0.096	1.922	0.080	1.481	0.069	0.366	0.423	0.291	0.300	0.319	0.336
1	-0.028	-0.058	-0.076	-0.078	-0.015	-0.031	0.084	0.084	0.124	0.124	0.062	0.062
2	0.109	-0.078	-0.124	-0.131	0.047	-0.059	0.166	0.168	0.245	0.245	0.124	0.125
3	1.219	-0.079	-0.144	-0.064	0.694	-0.027	0.323	0.337	0.474	0.475	0.249	0.256
4	0.351	0.157	0.240	0.136	0.249	0.144	0.171	0.173	0.128	0.13	0.129	0.13
5	1.348	0.039	0.791	0.070	0.775	0.053	0.327	0.34	0.251	0.258	0.251	0.258
6	7.218	-0.025	3.798	0.020	3.804	0.027	0.561	0.640	0.474	0.506	0.474	0.507
7	1.381	0.059	0.531	0.061	0.800	0.084	0.323	0.338	0.126	0.131	0.249	0.256
8	7.308	-0.019	2.781	0.045	3.767	-0.013	0.567	0.646	0.252	0.273	0.479	0.511
9	25.630	-0.168	10.703	-0.025	10.798	-0.041	0.683	1.034	0.474	0.520	0.802	0.888
10	0.002	-0.078	-0.050	-0.124	-0.009	-0.063	0.083	0.084	0.124	0.124	0.062	0.063
11	0.685	0.306	0.674	0.431	0.485	0.241	0.174	0.178	0.257	0.259	0.130	0.133
12	1.888	0.068	0.669	0.071	1.108	0.013	0.323	0.342	0.473	0.482	0.247	0.258
13	0.408	0.037	0.302	0.064	0.292	0.053	0.169	0.172	0.127	0.129	0.127	0.129
14	1.862	0.015	1.128	0.007	1.112	-0.009	0.327	0.346	0.251	0.261	0.25	0.261
15	8.660	0.110	4.774	0.123	4.750	0.096	0.561	0.650	0.474	0.513	0.474	0.513
16	1.871	0.031	0.690	0.024	1.159	0.053	0.323	0.342	0.125	0.132	0.248	0.259
17	8.191	-0.182	3.128	-0.020	4.526	-0.078	0.559	0.647	0.247	0.271	0.473	0.510
18	27.416	-0.040	11.450	0.014	12.077	0.068	0.683	1.043	0.477	0.52	0.806	0.892
19	-0.025	-0.047	-0.346	-0.081	-0.087	-0.033	0.087	0.087	0.13	0.129	0.065	0.065
20	-0.001	-0.146	-1.154	-0.196	-0.274	-0.109	0.172	0.173	0.258	0.254	0.130	0.129

continued ...

...continued

Scenario	Bias						Precision (S.E.)					
	Interquartile		First difference		Standard deviation		Interquartile		First difference		Standard deviation	
	OVA	IOVA	OVA	IOVA	OVA	IOVA	OVA	IOVA	OVA	IOVA	OVA	IOVA
21	1.116	0.081	-3.565	0.105	-0.386	0.044	0.340	0.349	0.501	0.492	0.264	0.264
22	0.109	-0.040	-0.186	-0.010	-0.192	-0.016	0.173	0.174	0.131	0.130	0.131	0.130
23	1.091	0.043	-0.393	0.021	-0.386	0.029	0.338	0.348	0.262	0.263	0.263	0.263
24	5.946	-0.135	-0.322	0.003	-0.351	-0.035	0.594	0.652	0.501	0.515	0.501	0.515
25	0.935	-0.101	0.101	-0.020	-0.425	-0.029	0.335	0.344	0.131	0.133	0.261	0.261
26	6.400	0.090	1.384	0.027	-0.163	0.065	0.599	0.657	0.265	0.275	0.504	0.518
27	23.159	-0.228	6.129	-0.026	0.726	-0.115	0.751	1.036	0.499	0.518	0.807	0.886
28	0.159	0.089	-0.120	0.125	0.033	0.063	0.089	0.090	0.134	0.133	0.067	0.067
29	0.447	0.149	-0.597	0.206	0.062	0.129	0.177	0.180	0.267	0.263	0.135	0.135
30	1.475	-0.019	-2.976	0.025	-0.067	0.014	0.338	0.351	0.501	0.497	0.264	0.266
31	0.128	-0.147	-0.169	-0.092	-0.173	-0.096	0.171	0.174	0.130	0.130	0.130	0.130
32	1.145	-0.270	-0.256	-0.176	-0.275	-0.198	0.332	0.345	0.259	0.261	0.258	0.261
33	7.371	0.029	0.685	0.076	0.675	0.064	0.589	0.655	0.499	0.516	0.499	0.516
34	1.317	-0.169	0.236	-0.053	-0.150	-0.090	0.338	0.350	0.132	0.135	0.263	0.265
35	7.259	0.079	1.817	0.099	0.659	0.151	0.594	0.659	0.264	0.276	0.502	0.520
36	24.905	0.045	6.709	0.031	1.755	0.048	0.756	1.047	0.502	0.519	0.811	0.893

Note: Bias indicates mean difference between estimate,  $\widehat{\Delta p}$  and population parameter,  $\Delta p$ ; total bias indicates mean of all scenario biases (absolute values). Precision is indicated by the standard error of the estimate,  $SD(\widehat{\Delta p})/\sqrt{n}$ , and total precision indicates the mean standard error of all scenario.