ARTIFICIAL EXAMPLE 2:

INTERPRETATION OF PARAMETERS IN THE COUNT-COMPONENT

Suppose we are interested in the effects of a binary predictor x_1 (e.g., education level being high or low), a continuous predictor x_2 (e.g., a measure of anxiety with values between 50 and 110, noted x_2), and its interaction on a count variable Y (e.g., the number of UPB-perpetrations). We illustrate that it can be misleading to interpret the parameters $\boldsymbol{\gamma}$ as predictors for the change in the mean of Y.

Under models (1), (2), (5) and (6) of the paper we generated a sample of size 1000, with $\boldsymbol{x}_{\mathbf{i}}^{t}\boldsymbol{\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i}$ and $\boldsymbol{x}_{\mathbf{i}}^{t}\boldsymbol{\gamma} = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{1i} x_{2i}$, with the following specific choices for the parameters $\beta_0 = \log 2$, $\beta_1 = -2\beta_0$, $\beta_2 = \log(5)/100$, $\beta_3 = -2\beta_2$, $\gamma_0 = \log(2)$, $\gamma_1 = -\beta_0$, $\gamma_2 = \log 10/100$ and $\gamma_3 = -\beta_2$. Parameter values were chosen such that $E(Y_i \mid X_{1i} = 0, X_{2i} = x)$ equals $E(Y_i \mid X_{1i} = 1, X_{2i} = x)$ for all x. The latter implies that looking at the mean of Y will not reveal a differential of x_2 by levels of x_1.

We fitted a ZIP-model, assuming models (1), (2), (5) and (6). The upper panel of table 1 reveals no significant effects on the logistic part but significant main effects for x_1 and x_2, and a significant interaction on the count part of the ZIP-model at the 5% significance level (without any multiplicity adjustment). Based on the estimated parameters on the count part, the erroneous interpretation could be made here that for example a 10-point increase in x_2 corresponds to a $(\exp(0.023 * 10) - 1) = 25\%$ increase in mean count of Y for $x_1 = 0$-group, and only a $(\exp((0.023 - 0.014) * 10) - 1) = 8\%$ increase in mean count of Y for $x_1 = 1$-group.

Looking at the mean predicted counts as a function of x_2 for both levels of x_1 in the right panel of figure 1, the aforementioned moderated effect of x_2 between x_1–levels is not visible. The tricky issue in the interpretation of μ_i made here is that the mean under a ZIP-distribution is given by $(1 - p_i)\mu_i$ and not by μ_i. The correct interpretation is that an increase in x_2 from a score of 0 to 10 corresponds to a $\exp(0.023 * 10) * (1 + \exp(0.998))/(1 + \exp(0.998 + 0.014 * 10)) - 1 = 13.4\%$ increase in mean counts of Y for $x_1 = 0$ and a $\exp((0.023-0.014)*10) * (1 + \exp(0.998 - 2.559))/(1 + \exp(0.998 - 2.559 + (0.014 - 0.017) * 10)) - 1 = 10.0\%$ increase in mean count of Y for $x_1 = 1$. Alternatively, following the latent class interpretation (as in Karazsia & van Dulmen, 2008), one could say here that among the subjects in the 'not always zero group' a 10-point increase in x_2 corresponds to a 25% increase for the $x_1 = 0$-

group and an 8% increase in x_1 = 1-group.

For this example parameter choices were selected such that the difference between $E(Y_i \mid X_{1i} = 0, X_{2i} = x)$ and $E(Y_i \mid X_{1i} = 1, X_{2i} = x)$ is zero.Hence, the lack of interaction in the right panel of figure 1 is not surprising. The presence or absence of interactions observed in the fitted model can be seen on the individual components of the mixture model (left and middle panel of figure 1), but not on the mean. The left panel of figure 2 shows an alternative presentation of the data. The area of rectangle formed by the estimated μ_i (on the X-axis) and $(1 - p_i)$ (on the Y-axis) corresponds to mean of Y_i (note that to illustrate the effect of continuous predictor X_2 estimated values were shown at particular values of X_2, e.g. 1 standard deviation below or above the mean). The right panel of figure 3 shows the model-predicted distribution of Y for several levels of x_1 and x_2. The lower panel of table 1 shows the results from the fitted PLH-model.

Estimated parameters for the count component in the PLH-model are very similar to these in the ZIP-model, and as for the ZIP-model graphical presentations like figure 1 and figure 2 may help to understand the estimated effects (see R-code).

	Logistic portion			Counts portion			Joint
Variable	β	SE β	Z	γ	SE γ	Z	LRT - χ^2
ZIP-model							
Intercept	0.998	0.692	1.442	0.602	0.215	2.796**	
x_1	-2.560	1.351	-1.894	-0.729	0.311	-2.346*	
x_2	0.014	0.009	1.592	0.023	0.003	9.272***	
x1*x2	-0.017	0.016	-1.063	-0.015	0.004	-3.995***	16.13***(df =2)
Hurdle-model							
Intercept	1.000	0.692	1.446	0.602	0.215	2.797**	
x_1	-0.906	0.832	-1.088	-0.730	0.312	-2.343*	
x_2	0.013	0.009	1.589	0.023	0.003	9.273***	
*x1*x2*	-0.027	0.010	-2.552*	-0.015	0.004	-3.983***	22.29***(df=2)

Table 1: Hypothetical Example 2: Estimated Parameters under Zero-Inflated Poisson Model and Poisson Logit Hurdle Model (*** $p \le 0.001$, ** $p \le 0.01$, * $p \le 0.05$)

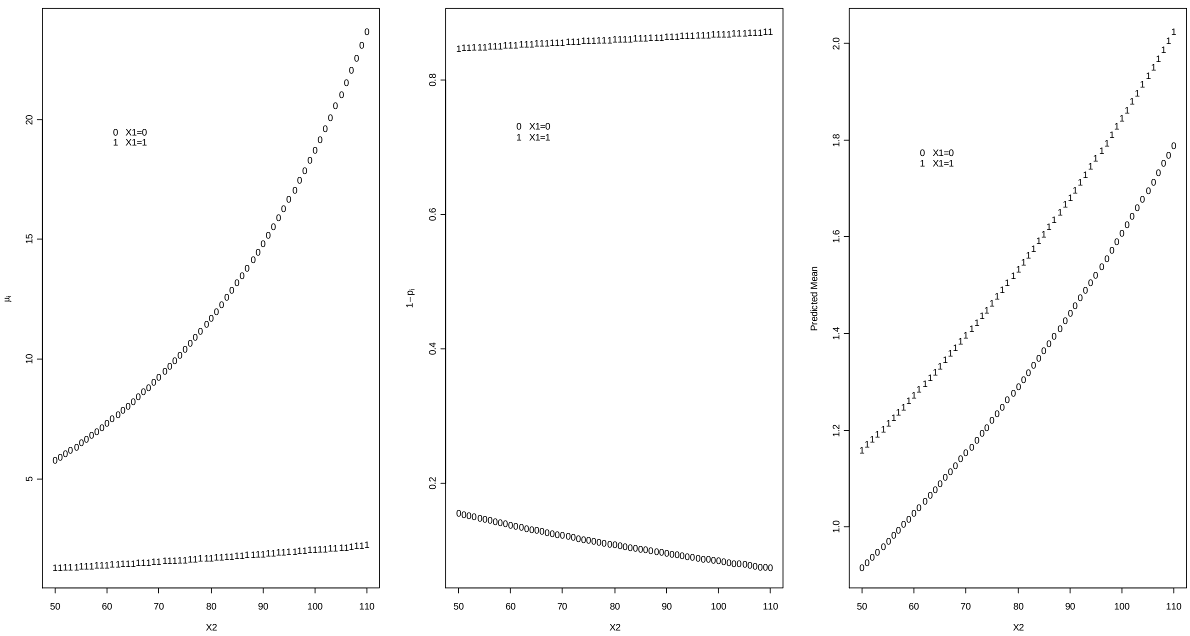

Figure 1: : Effect of x_2 by x_1-level on both components and the mean of the ZIP-model

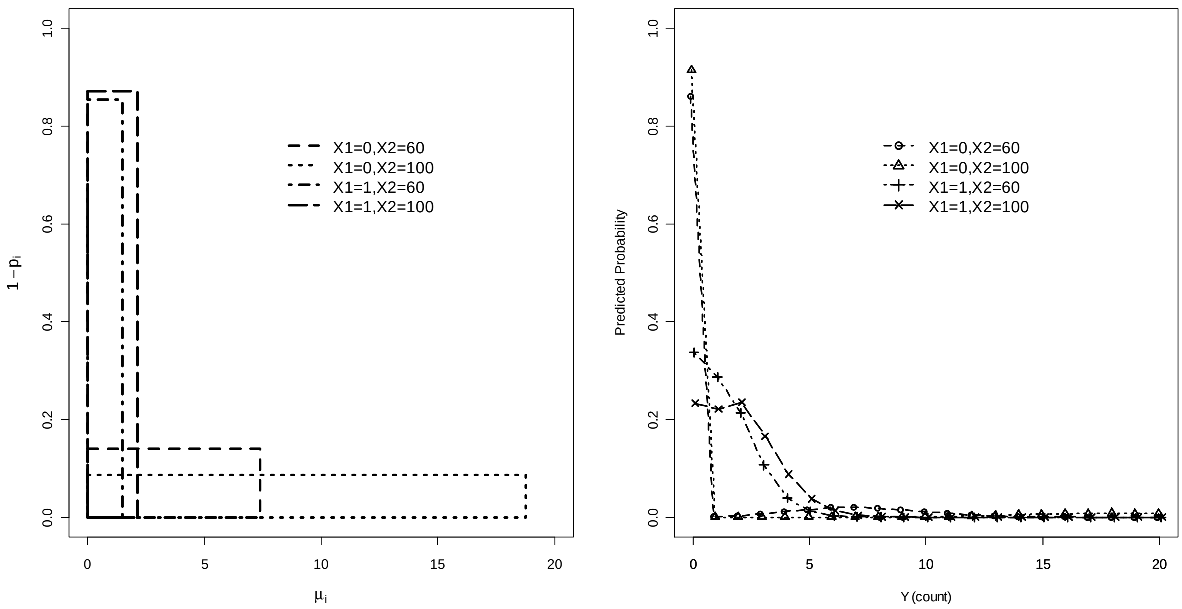

Figure 2: Alternative graphical presentations of interaction between x_1-level and x_2-level on Y