# Supporting Information for "Dir-GLM: A Bayesian GLM with data-driven reference distribution" by Entejar Alam, Peter Müller, and Paul J. Rathouz

## Appendix A: Results, Proofs and additional Figures

In this section, we first provide the proofs of all the results stated in "Dir-GLM: A Bayesian GLM with data-driven reference distribution" [AMR, 2024].

**Result 1.** *The exponentially tilted Dirichlet distribution, tDir$(c, \mu, s)$ with $c = \{\alpha H(s_\ell);\ \ell = 1, \ldots, k\}$, $s = \{s_\ell;\ \ell = 1, \ldots, k\}$, is outside the Dirichlet parametric family; however, it remains a valid probability distribution defined over the $(k-1)$ dimensional simplex. In the context of Eq. (1) in [AMR, 2024], exponential tilting transforms $f_0(s_\ell)$ to $f(s_\ell \mid x) \propto \exp(\theta s_\ell) f_0(s_\ell)$.*

*Proof.* As per Eq. (4) in [AMR, 2024], $[f_0(s_1), \ldots, f_0(s_k)] \sim \text{Dir}\left[\alpha H(s_1), \ldots, \alpha H(s_k)\right]$. Let $a_\ell = \alpha H(s_\ell) > 0$ and $f_{0,\ell} = f_0(s_\ell)$, $\ell = 1, \ldots, k$. Upon exponential tilting $f_{0,\ell}$ as per Eq. (1) in [AMR, 2024], it transforms to $f(s_\ell \mid x) \propto \exp(\theta_x s_\ell) f_{0,\ell}$. We use $f_0^{(x)}$ to represent the tilted random vector $[f(s_1 \mid x), \ldots,\ \ f(s_k \mid x)]$.

We proof the result using contradiction. We then assume that $f_0^{(x)}$ follows a Dirichlet distribution. So, its marginal should be a beta distribution. For simplicity, we first take $k = 2$, and denote $Z_\ell = f_{0,\ell}$, for $\ell = 1, 2$. Thus, $Z_1$ follows a beta distribution with shape parameters $(a_1, a_2)$, and $Z_2 = 1 - Z_1$. Writing $Z_\ell^x = f_{0,\ell}^x$, we have

$$Z_1^x = \frac{\exp(\theta_x s_1) Z_1}{\exp(\theta_x s_1) Z_1 + \exp(\theta_x s_2) Z_2} = \frac{\exp(\theta_x s_1) Z_1}{\exp(\theta_x s_2) + [\exp(\theta_x s_1) - \exp(\theta_x s_2)] Z_1} = h(Z_1).$$

For notational simplicity, let $u = \exp(\theta_x s_1)$ and $v = \exp(\theta_x s_2)$. Then $Z_1^x$ can be expressed as $Z_1^x = \frac{uZ_1}{v+(u-v)Z_1}$, and $Z_1 = h^{-1}\left(Z_1^x\right) = \frac{vZ_1^x}{u-Z_1^x(u-v)}$. The density of $Z_1$ is $f_{Z_1}(z) = \frac{1}{B(a_1,a_2)} z^{a_1-1}(1-z)^{a_2-1}$, $\quad 0 < z < 1$, where $B(a_1, a_2)$ is the beta function. The Jacobian of the transformation is $J = \left|\frac{dz_1}{dz_1^x}\right| = \frac{uv}{[u-z_1^x(u-v)]^2}$. Applying the change of variables formula, we get

$$\begin{aligned} f_{Z_1^x}(z) &= f_{Z_1}(h^{-1}(z)) \cdot |J| \\ &= \frac{1}{B(a_1, a_2)} \cdot \left[\frac{vz}{A - z(u-v)}\right]^{a_1 - 1} \cdot \left[1 - \frac{vz}{u - z(u-v)}\right]^{a_2-1} \cdot \frac{uv}{[u - z(u-v)]^2} \end{aligned}$$

Let $w = v/u$. After simplification, the density of $Z_1^x$ is

$$f_{Z_1^x}(z) = \frac{w^{a_1}}{B(a_1, a_2)} \cdot \frac{z^{a_1-1} \cdot (1-z)^{a_2-1}}{[1-(1-w)z]^{a_1+a_2}}, \tag{1}$$

for $0 < z < 1$, where $w = \exp[\theta_x(s_2 - s_1)]$. Based on the density in Eq. (1), we can say that $Z_1^x$ does not follow a beta distribution. A similar argument can be made for $k > 2$, with $h$ representing a multivariate mapping. This contradiction implies that the exponentially tilted Dirichlet random vector $f_0^{(x)}$ is outside the Dirichlet parametric family. Nevertheless, since $f_{0,\ell}^{(x)} \in [0, 1]$ for all $s_\ell$, $\ell = 1, \ldots, k$ and $\sum_{\ell=1}^{k} f_{0,\ell}^{(x)} = 1$, it remains a valid probability distribution defined over the standard $(k-1)$ dimensional simplex. □

**Result 2.** *Under the Dir-GLM with prior model (4) in [AMR, 2024], with the additional constraint that $\mu$ is bounded away from the two endpoints $s_1$ and $s_k$, and $f_0 \in \mathcal{F}^\star = \{f_0^\star \in \mathcal{F} : E_{f_0^\star}(y) = \mu_0\}$, if $g$ and $\mu_0$ are chosen such that as $||x||_2 \to 0$, $\mu = g^{-1}(\eta) \xrightarrow{P} \mu_0$, then the derived parameter $\theta = \theta(x; \beta, f_0)$ has the following properties*

*(a) $\theta \xrightarrow{P} 0$, and*

*(b) $\theta$ is asymptotically normal with mean zero.*

*Proof.* The derived parameter $\theta = \theta(x; \beta, f_0)$ is a solution from Eq. (3) in [AMR, 2024]. If possible let, $\theta \xrightarrow{P} \theta_0 (\neq 0)$ as $||x||_2 \to 0$. Using continuous mapping theorem, we have

$$b'(\theta) \xrightarrow{P} b'(\theta_0) \text{ [as } b' \text{ is a continuous function]} \tag{2}$$

From Eq. (3) in [AMR, 2024], we have

$$b'(\theta) = \frac{\int_{\mathcal{Y}} y \exp(\theta y) f_0(y) dy}{\int_{\mathcal{Y}} \exp(\theta y) f_0(y) dy} = g^{-1}(\eta) \xrightarrow{P} \mu_0 \text{ [as } \mu = g^{-1}(\eta) \xrightarrow{P} \mu_0, \text{ as } ||x||_2 \to 0] \tag{3}$$

Note that $b'$ is a one-to-one function since $b''(\theta) = \text{Var}_\theta(y) > 0$. Using Eqs. (2) and (3), $b'(\theta_0) = \mu_0$. A trivial solution is $\theta_0 = 0$, as $b'(0) = \frac{\int_{\mathcal{Y}} y f_0(y) dy}{\int_{\mathcal{Y}} f_0(y) dy} = \mu_0$. By assumption $\theta_0 \neq 0$. Hence, there are at least two values of $\theta_0$ which satisfies $b'(\theta_0) = \mu_0$. This implies that $b'$ is not a one-to-one function, which is not true. Hence, proof by contradiction and we have $\theta \xrightarrow{P} 0$ as $||x||_2 \to 0$. This completes part (a).

Define a function $h : \mathcal{R} \to \mathcal{R}$ such that $h = b'^{-1} \circ g^{-1}$. First consider proving the following required properties: $h(0) = 0$ and $h'(0) \neq 0$. If possible let, $h(0) = \theta_0 (\neq 0)$. We have $lim_{||x||_2 \to 0} g^{-1}(\eta) = \mu_0$. Since $g^{-1}$ is a continuous function, $g^{-1}(0) = \mu_0$. Based on Eq. (3) in [AMR, 2024], we have $b'(\theta_0) = b'(h(0)) = g^{-1}(0) = \mu_0$. Also, $b'(0) = \mu_0$. Since $\theta_0 \neq 0$, $b'$ is not a one-to-one function, which is not true. Hence, proof by contradiction and we have $h(0) = 0$. For the next one, first note that $h$ is differentiable by definition since composite of two differentiable functions is also differentiable. Upon differentiating $g(b'(h(\eta))) = \eta$ both sides with respect to $\eta$, we have

$$g'(b'(h(\eta))) \cdot b''(h(\eta)) \cdot h'(\eta) = 1 \tag{4}$$

Eq. (4) implies that $h'(\eta)$ can not be zero for any choices of $\eta$. Hence, $h'(0) \neq 0$. In part (b), we are to show that $\theta$ is asymptotically mean-zero normal as $||x||_2 \to 0$. Due to the sequential definition of limit, equivalently we can show the result for an arbitrary sequence $\{x_m\}_{m=1}^\infty$ with $lim_{m\to\infty} ||x_m||_2 \to 0$, where $x_m = \left(\frac{x_1^\star}{\sqrt{m_1}}, \ldots, \frac{x_p^\star}{\sqrt{m_p}}\right)$ with $m = \min_j m_j \to \infty$. We consider the following two exhaustive cases.

Case 1. Take $m_j = k_j m + r_j$ for some fixed constants $k_j \neq 0$ and $r_j$. Define, $\widetilde{x}_j = \frac{x_j^\star}{\sqrt{k_j}}$ and $\frac{r_j}{k_j} = \widetilde{r}_j$ for all $j$. Then, we can express $x_m$ as

$$x_m = \left(\frac{x_1^\star}{\sqrt{m_1}}, \ldots, \frac{x_p^\star}{\sqrt{m_p}}\right) = \left(\frac{\widetilde{x}_1}{\sqrt{m + \widetilde{r}_1}}, \ldots, \frac{\widetilde{x}_p}{\sqrt{m + \widetilde{r}_p}}\right)$$

We have, $\eta_m = g(b'(\theta_m)) = x_m^t \beta = \sum_{j=1}^p \frac{\widetilde{x}_j}{\sqrt{m + \widetilde{r}_j}} \beta_j$, which implies $\sqrt{m}\eta_m = \sum_{j=1}^p \frac{\widetilde{x}_j}{\sqrt{1 + \widetilde{r}_j/m}} \beta_j \xrightarrow{D} N(0, ||\widetilde{x}||_2^2)$ (using Slutsky's theorem as $1 + \widetilde{r}_j/m \to 1$). Using delta method,

$$\sqrt{m}(\eta_m - 0) \sim N(0, ||\widetilde{x}||_2^2) \Rightarrow \sqrt{m}\{h(\eta_m) - h(0)\} \xrightarrow{D} N(0, \{h'(0)\}^2 ||\widetilde{x}||_2^2), \tag{5}$$

since $h'(0)$ exists and is non-zero. Note that $\theta_m = b'^{-1}(g^{-1}(\eta_m)) = h(\eta_m)$. Since $h(0) = 0$, from Eq. (5) we have $\sqrt{m}\theta_m \xrightarrow{D} N(0, \{h'(0)\}^2 ||\widetilde{x}||_2^2) \equiv$ mean-zero normal.

Case 2. Take $m_j = k_j m^{\alpha_j} + r_j$ for some fixed constants $k_j \neq 0$ and $r_j$, where $\alpha_j = 1$ for some $j$ and $\alpha_j > 1$ for others. WOLG let's assume that $\alpha_j = 1$ for $j = 1$ and $m = \min_j m_j = m_1$. Define, $\widetilde{x}_j = \frac{x_j^\star}{\sqrt{k_j}}$ and $\widetilde{r}_j = \frac{r_j}{k_j}$ for all $j$. Then, we can express $x_m$ as

$$x_m = \left(\frac{\widetilde{x}_1}{\sqrt{m}}, \frac{\widetilde{x}_2}{\sqrt{m^{\alpha_2} + \widetilde{r}_2}}, \ldots, \frac{\widetilde{x}_p}{\sqrt{m^{\alpha_p} + \widetilde{r}_p}}\right)$$

We have, $\eta_m = g(b'(\theta_m)) = x_m^t\beta = \frac{\widetilde{x}_1}{\sqrt{m}}\beta_1 + \sum_{j=2}^{p} \frac{\widetilde{x}_j}{\sqrt{m^{\alpha_j} + \widetilde{r}_j}}\beta_j \Rightarrow \sqrt{m}\eta_m = \widetilde{x}_1\beta_1 + \sum_{j=2}^{p} \frac{\widetilde{x}_j}{\sqrt{m^{\alpha_j - 1} + \widetilde{r}_j/m}}\beta_j$. Here, $\widetilde{x}_1\beta_1 \sim N(0, \widetilde{x}_1^2)$ and $\sum_{j=2}^{p} \frac{\widetilde{x}_j}{\sqrt{m^{\alpha_j - 1} + \widetilde{r}_j/m}}\beta_j \xrightarrow{P} 0$ as $\alpha_j - 1 > 0$ for $j = 2, \ldots, n$. Hence, using Slutsky's theorem, we can say that $\sqrt{m}\eta_m \xrightarrow{D} N(0, \widetilde{x}_1^2) + 0 \equiv N(0, \widetilde{x}_1^2)$. Using delta method and following similar steps as in case 1, we get

$$\sqrt{m}\theta_m \xrightarrow{D} N(0, \{h'(0)\}^2\widetilde{x}_1^2) \equiv \text{mean-zero normal}$$

In case 1 and 2, we have shown that for any arbitrary sequence $\{x_m\}_{m=1}^{\infty}$ with $lim_{m\to\infty}||x_m||_2 \to 0$, $\sqrt{m}\theta_m$ converges in distribution to a mean-zero normal. Hence, as $||x||_2 \to 0$, the derived parameter $\theta$ is asympotically normal with zero mean. □

**Additional Figures.** See Section 3.1 in [AMR, 2024]. We here show the cdf of $f_0^\star \in \mathcal{F}^\star$ for two choices of $f_0 \in \mathcal{F}$ in Figure 1. As we increase $\mu_0$, the distribution $f_0^\star$ becomes more and more left-skewed to achieve the required mean.

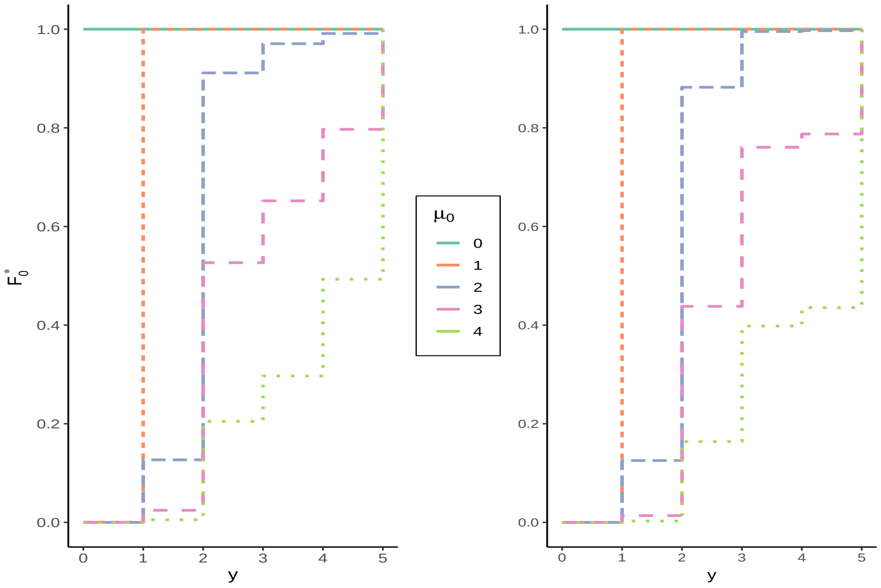

Figure 1: For two choices of $f_0$ (blue line, with $\mu_0 = 2$), the figure shows the equivalent (with respect to likelihood identifiability) exponentially tilted $f_0^\star$ with fixed $\mu_0$, with different color lines showing the cdf for $f_0^\star$ for different $\mu_0$.

See Section 3.2 in [AMR, 2024]. Using the same setup as in Figure 2 in [AMR, 2024], a kernel density estimate of the prior density on $\theta$ for different values of $x$ is displayed in Figure 2.

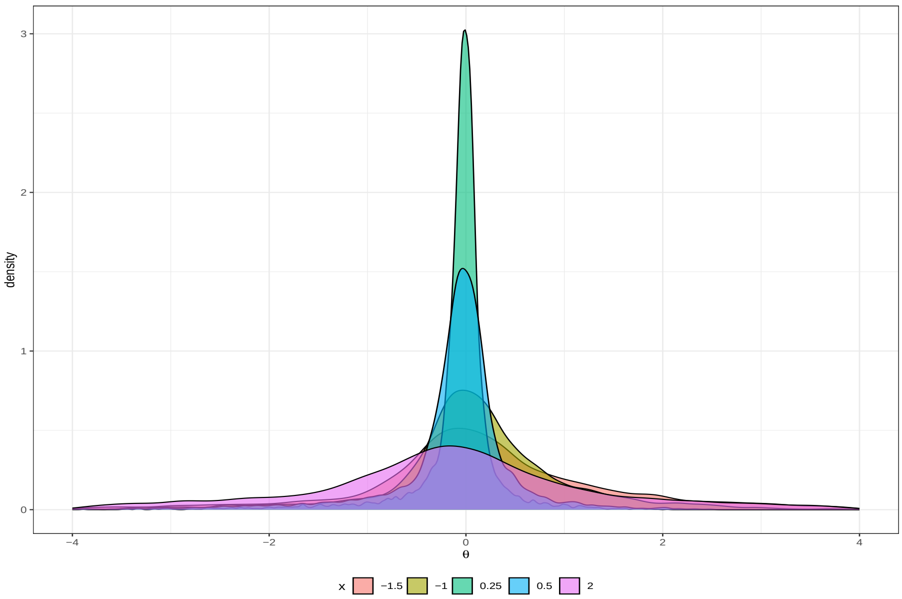

Figure 2: Kernel density estimate of the prior density on $\theta$, for various values of $x$.

# Appendix B: Simulation results

This section provides some additional plots for the regression parameters. Figure 3 represents the Rao–Blackwellized marginal posterior density estimate and its uncertainty quantification, in terms of 95% quantile-based credible intervals, for the regression parameters. The estimates and credible intervals are averaged across 10 replicates based on the first simulation scenario.

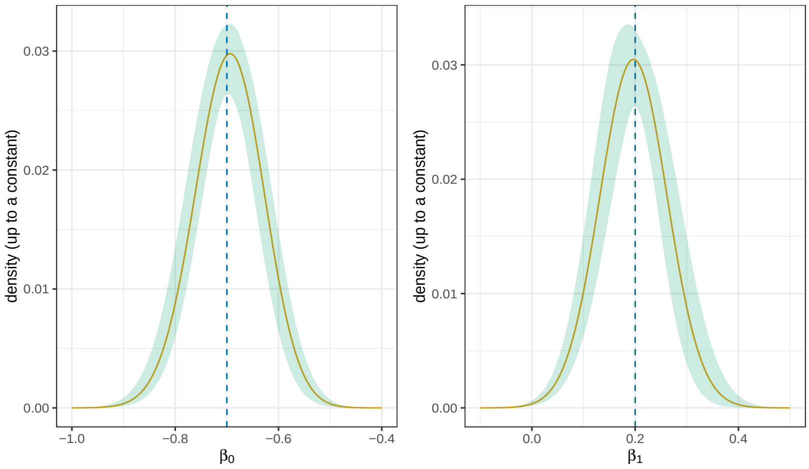

Figure 3: Simulation study. The solid brown curve represents the Rao–Blackwellized marginal posterior density estimate and the shaded area (in light green) surrounding it indicates the corresponding 95% credible band for the regression parameters. The blue dotted line represents the true value.

Figure 4 shows the box-plots of Dir-GLM estimates as a function of sample size ($n$), for the regression parameters. With the increase in sample size, the estimate converges to the truth — ensures consistency. This figure is based on $1,000$ replicates using the first simulation scenario.

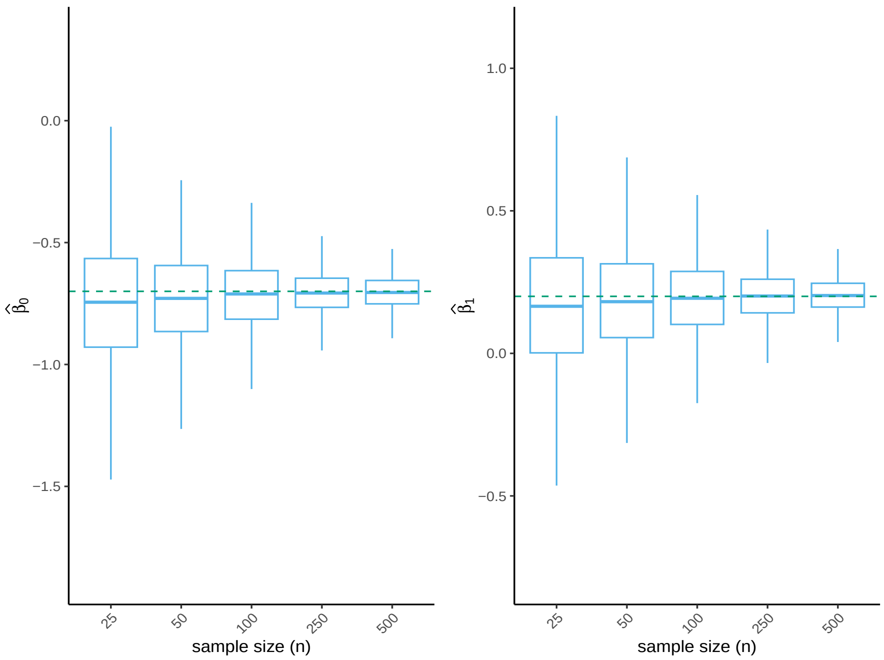

Figure 4: Simulation study. Box-plots of Dir-GLM estimates as a function of sample size, for the regression parameters. The green dotted line represents the true value.

In Figures 5 and 6, we compare estimation of the selected exceedance probabilities, $p(y \geq y_0 \mid x)$ with $y_0 = 2$ and 4, under GLDRM and Dir-GLM models.

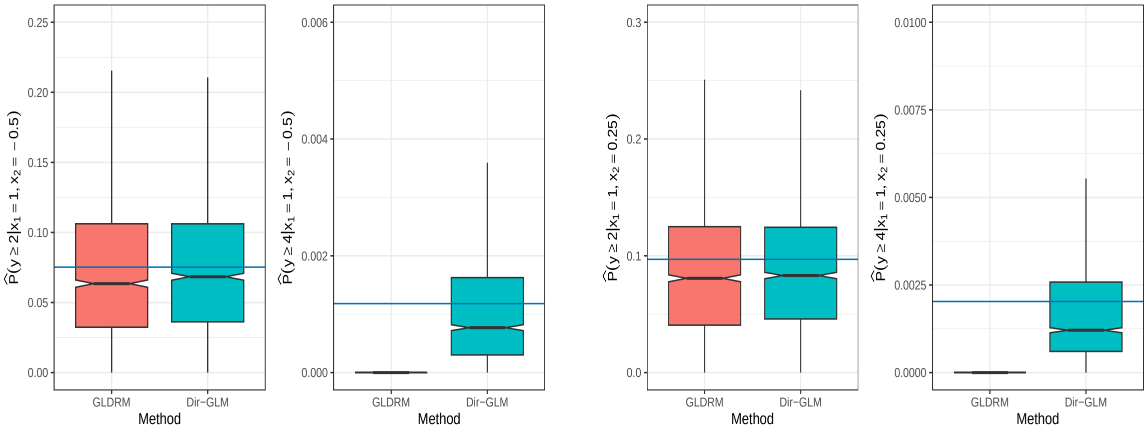

Figure 5: Simulation study. Exceedance probability estimates for GLDRM and Dir-GLM models replicating over 2, 000 simulated data sets. Blue horizontal line denotes the true value. This figure is based on the first simulation scenario.

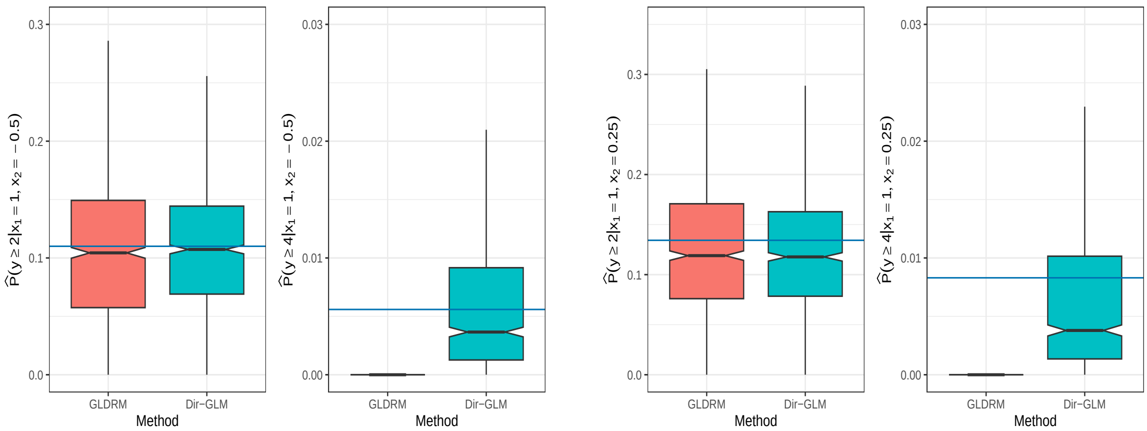

Figure 6: Simulation study. Exceedance probability estimates for GLDRM and Dir-GLM models replicating over $2,000$ simulated data sets. Blue horizontal line denotes the true value. This figure is based on the second simulation scenario.

# Appendix C: Large sample simulation results for $f_0$

Table 1: Simulation study. Same as Table 2 in [AMR, 2024], for $n = 250$.

| n | DGM | Parm | Method | True value | Est$_a$ | RRMSE$_a$ | Est$_m$ | RRMSE$_m$ | CP |
|---|---|---|---|---|---|---|---|---|---|
| 250 | 1 | $f_0(0)$ | GLDRM | 0.367 | 0.359 | 1.00 | 0.359 | 1.00 | N/A |
| | | | Dir-GLM | | 0.370 | 0.92 | 0.370 | 0.93 | 0.95 |
| | | $f_0(1)$ | GLDRM | 0.368 | 0.373 | 1.00 | 0.372 | 1.00 | N/A |
| | | | Dir-GLM | | 0.369 | 0.97 | 0.368 | 0.99 | 0.95 |
| | | $f_0(2)$ | GLDRM | 0.185 | 0.191 | 1.00 | 0.188 | 1.00 | N/A |
| | | | Dir-GLM | | 0.183 | 0.91 | 0.181 | 0.94 | 0.93 |
| | | $f_0(3)$ | GLDRM | 0.062 | 0.062 | 1.00 | 0.062 | 1.00 | N/A |
| | | | Dir-GLM | | 0.056 | 0.92 | 0.056 | 0.93 | 0.93 |
| | | $f_0(4)$ | GLDRM | 0.015 | 0.014 | 1.00 | 0.000 | 1.00 | N/A |
| | | | Dir-GLM | | 0.016 | 0.63 | 0.014 | 0.45 | 0.97 |
| | | $f_0(5)$ | GLDRM | 0.003 | 0.000 | 1.00 | 0.000 | 1.00 | N/A |
| | | | Dir-GLM | | 0.006 | 1.70 | 0.004 | 0.48 | 0.97 |
| | 2 | $f_0(0)$ | GLDRM | 0.471 | 0.465 | 1.00 | 0.466 | 1.00 | N/A |
| | | | Dir-GLM | | 0.471 | 0.91 | 0.472 | 0.93 | 0.94 |
| | | $f_0(1)$ | GLDRM | 0.232 | 0.235 | 1.00 | 0.234 | 1.00 | N/A |
| | | | Dir-GLM | | 0.233 | 0.99 | 0.232 | 1.00 | 0.95 |
| | | $f_0(2)$ | GLDRM | 0.172 | 0.177 | 1.00 | 0.174 | 1.00 | N/A |
| | | | Dir-GLM | | 0.173 | 0.95 | 0.172 | 0.97 | 0.94 |
| | | $f_0(3)$ | GLDRM | 0.085 | 0.087 | 1.00 | 0.087 | 1.00 | N/A |
| | | | Dir-GLM | | 0.082 | 0.93 | 0.084 | 0.94 | 0.94 |
| | | $f_0(4)$ | GLDRM | 0.031 | 0.030 | 1.00 | 0.029 | 1.00 | N/A |
| | | | Dir-GLM | | 0.028 | 0.86 | 0.026 | 0.99 | 0.91 |
| | | $f_0(5)$ | GLDRM | 0.009 | 0.006 | 1.00 | 0.000 | 1.00 | N/A |
| | | | Dir-GLM | | 0.012 | 0.87 | 0.008 | 0.54 | 0.95 |

# Appendix D: Predictive inference for the AHEAD data

We perform a small sample study on the AHEAD data in Section 6.2 in [AMR, 2024]. The competing models GLDRM and Dir-GLM are fitted based on a small training data set of size $n = 100$, randomly sampled from the complete AHEAD data of size $6,441$. We then assess prediction accuracy on the held-out test data set of size 6,341. For comparison, we focus on estimating probabilities of exceedance events at moderate and severe difficulty in daily activities, i.e. $p(y \geq y_0 \mid x)$ with $y_0 = 2$ and 4, respectively. The comparison is summarized in Figure 7.

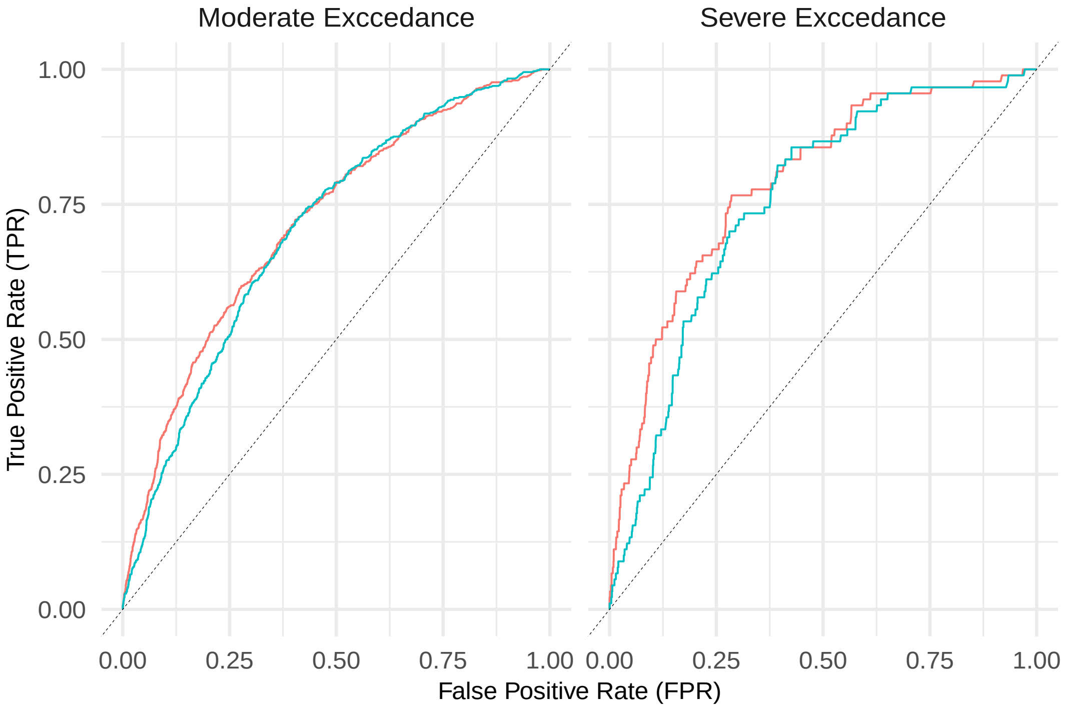

Figure 7: AHEAD study. ROC plot for comparing GLDRM (cyan) and Dir-GLM (coral) model accuracy in predicting moderate exceedance events (left) and severe exceedance events (right) on the held-out test data set.

Figure 8 presents some additional analysis plots. The competing models GLDRM and Dir-GLM are again fitted based on a small training data set of size $n = 100$, randomly sampled from the complete AHEAD data of size $6,441$. However here we access the prediction accuracy in estimating exceedance probabilities based on five test data sets, of size $1,000$, randomly sampled from the held aside test data set of size $6,341$.

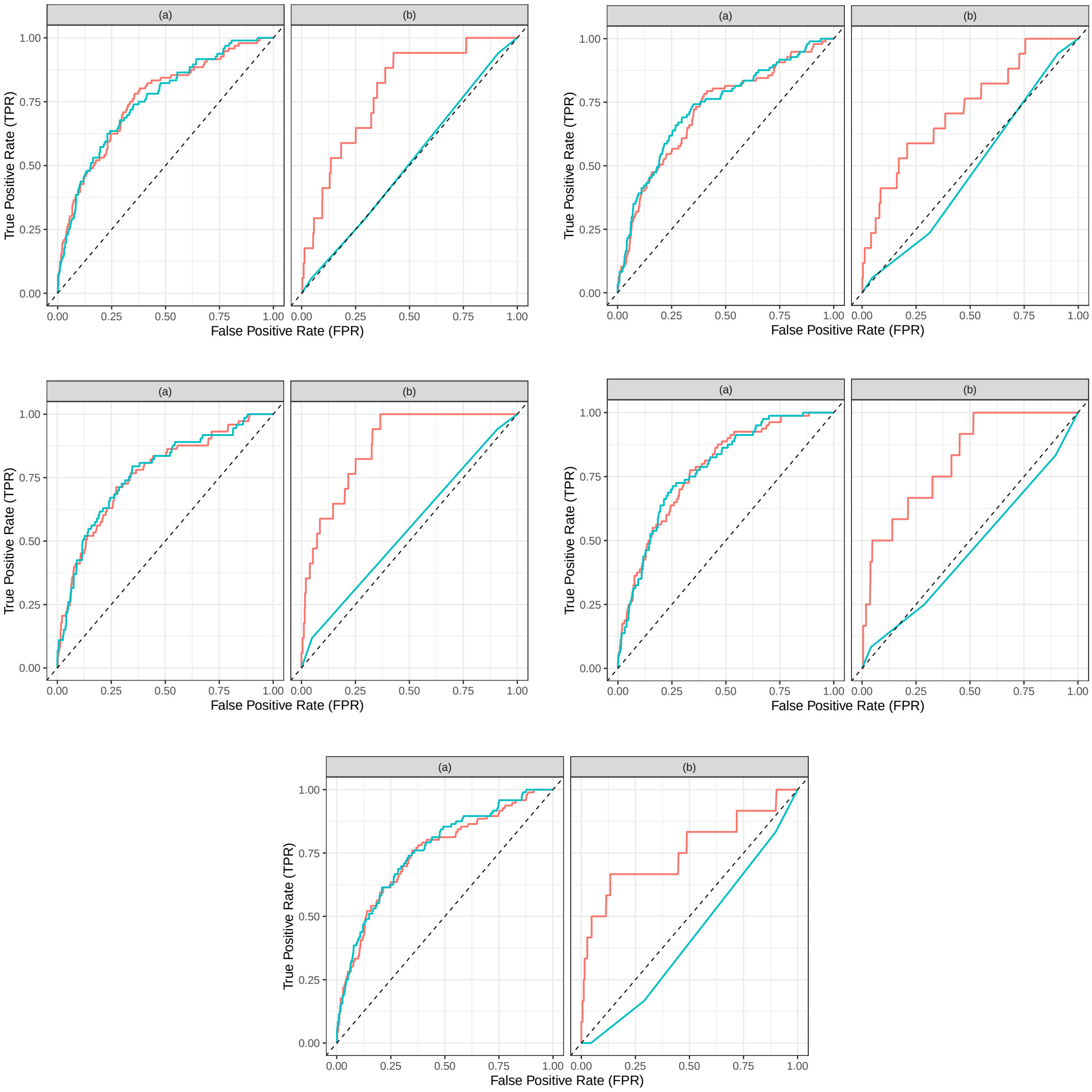

Figure 8: AHEAD study. ROC plot for comparing GLDRM (cyan) and Dir-GLM (coral) model accuracy in exceedance probabiities, $p(y \geq y_0 \mid x)$, estimation with $y_0 = 2$ (a) and $y_0 = 4$ (b).

The above figures in 8 highlight the limitation of maximum likelihood based inference in sparse or small data scenarios. We have a sparse-data situation for $y_0 = 4$ and hence, only a limited number of observations are available for estimating $p(y \geq y_0|x)$, which is not the case for $y_0 = 2$.

# Appendix E: AHEAD study results for the model parameters

Table 2: AHEAD study. Small and large training sample results for regression parameters. Abbreviations: TSS – Training sample size; Par – Parameter; Est – Estimate; CI – 95% credible or confidence intervals.

| TSS | Par | Method | Est | CI | CI length |
|---|---|---|---|---|---|
| Small | $\beta_0$ | GLDRM | -0.10 | [-1.37, 1.17] | 2.54 |
| | | Dir-GLM | -0.69 | [-1.53, 0.09] | 1.62 |
| | $\beta_1$ | GLDRM | 0.14 | [-0.29, 0.58] | 0.87 |
| | | Dir-GLM | 0.20 | [-0.20, 0.62] | 0.82 |
| | $\beta_2$ | GLDRM | -0.29 | [-1.12, 0.54] | 1.66 |
| | | Dir-GLM | -0.24 | [-0.97, 0.57] | 1.54 |
| | $\beta_3$ | GLDRM | -0.19 | [-0.72, 0.34] | 1.06 |
| | | Dir-GLM | -0.22 | [-0.70, 0.28] | 0.98 |
| | $\beta_4$ | GLDRM | -0.72 | [-2.00, 0.55] | 2.55 |
| | | Dir-GLM | -0.24 | [-1.22, 0.78] | 2.00 |
| | $\beta_5$ | GLDRM | -1.82 | [-3.65, 0.00] | 3.65 |
| | | Dir-GLM | -0.94 | [-2.09, 0.22] | 2.31 |
| | $\beta_6$ | GLDRM | -1.11 | [-2.41, 0.19] | 2.60 |
| | | Dir-GLM | -0.55 | [-1.53, 0.38] | 1.91 |
| | $\beta_7$ | GLDRM | -1.16 | [-2.76, 0.43] | 3.19 |
| | | Dir-GLM | -0.53 | [-1.61, 0.57] | 2.18 |
| Large | $\beta_0$ | GLDRM | -0.70 | [-0.86, -0.55] | 0.31 |
| | | Dir-GLM | -0.71 | [-0.87, -0.57] | 0.30 |
| | $\beta_1$ | GLDRM | 0.28 | [0.23, 0.32] | 0.09 |
| | | Dir-GLM | 0.28 | [0.23, 0.32] | 0.09 |
| | $\beta_2$ | GLDRM | 0.16 | [0.06, 0.26] | 0.20 |
| | | Dir-GLM | 0.15 | [0.05, 0.26] | 0.21 |
| | $\beta_3$ | GLDRM | -0.39 | [-0.44, -0.34] | 0.10 |
| | | Dir-GLM | -0.39 | [-0.44, -0.34] | 0.10 |
| | $\beta_4$ | GLDRM | -0.26 | [-0.41, -0.10] | 0.31 |
| | | Dir-GLM | -0.25 | [-0.40, -0.09] | 0.31 |
| | $\beta_5$ | GLDRM | -0.45 | [-0.61, -0.29] | 0.32 |
| | | Dir-GLM | -0.44 | [-0.59, -0.28] | 0.31 |
| | $\beta_6$ | GLDRM | -0.69 | [-0.85, -0.53] | 0.32 |
| | | Dir-GLM | -0.69 | [-0.83, -0.51] | 0.32 |
| | $\beta_7$ | GLDRM | -0.76 | [-0.94, -0.59] | 0.35 |
| | | Dir-GLM | -0.76 | [-0.91, -0.57] | 0.34 |

Table 3: AHEAD study. Small and large training sample results for baseline distribution $f_0$. Abbreviations: TSS – Training sample size; Par – Parameter; Est – Estimate; CI – 95% credible or confidence intervals; N/A – Not available.

| TSS | Par | Method | Est | CI |
|---|---|---|---|---|
| Small | $f_0(0)$ | GLDRM | 0.805 | N/A |
| | | Dir-GLM | 0.815 | [0.764, 0.859] |
| | $f_0(1)$ | GLDRM | 0.136 | N/A |
| | | Dir-GLM | 0.126 | [0.069, 0.198] |
| | $f_0(2)$ | GLDRM | 0.023 | N/A |
| | | Dir-GLM | 0.021 | [0.004, 0.057] |
| | $f_0(3)$ | GLDRM | 0.021 | N/A |
| | | Dir-GLM | 0.020 | [0.003, 0.044] |
| | $f_0(4)$ | GLDRM | 0.009 | N/A |
| | | Dir-GLM | 0.010 | [0.001, 0.028] |
| | $f_0(5)$ | GLDRM | 0.006 | N/A |
| | | Dir-GLM | 0.009 | [0.001, 0.026] |
| Large | $f_0(0)$ | GLDRM | 0.725 | N/A |
| | | Dir-GLM | 0.725 | [0.719, 0.731] |
| | $f_0(1)$ | GLDRM | 0.187 | N/A |
| | | Dir-GLM | 0.186 | [0.176, 0.196] |
| | $f_0(2)$ | GLDRM | 0.059 | N/A |
| | | Dir-GLM | 0.059 | [0.054, 0.064] |
| | $f_0(3)$ | GLDRM | 0.022 | N/A |
| | | Dir-GLM | 0.022 | [0.019, 0.025] |
| | $f_0(4)$ | GLDRM | 0.006 | N/A |
| | | Dir-GLM | 0.006 | [0.005, 0.008] |
| | $f_0(5)$ | GLDRM | 0.002 | N/A |
| | | Dir-GLM | 0.002 | [0.001, 0.002] |

# Appendix F: Uncertainty in exceedance probability estimates for the AHEAD data

See Section 6.3 in [AMR, 2024].

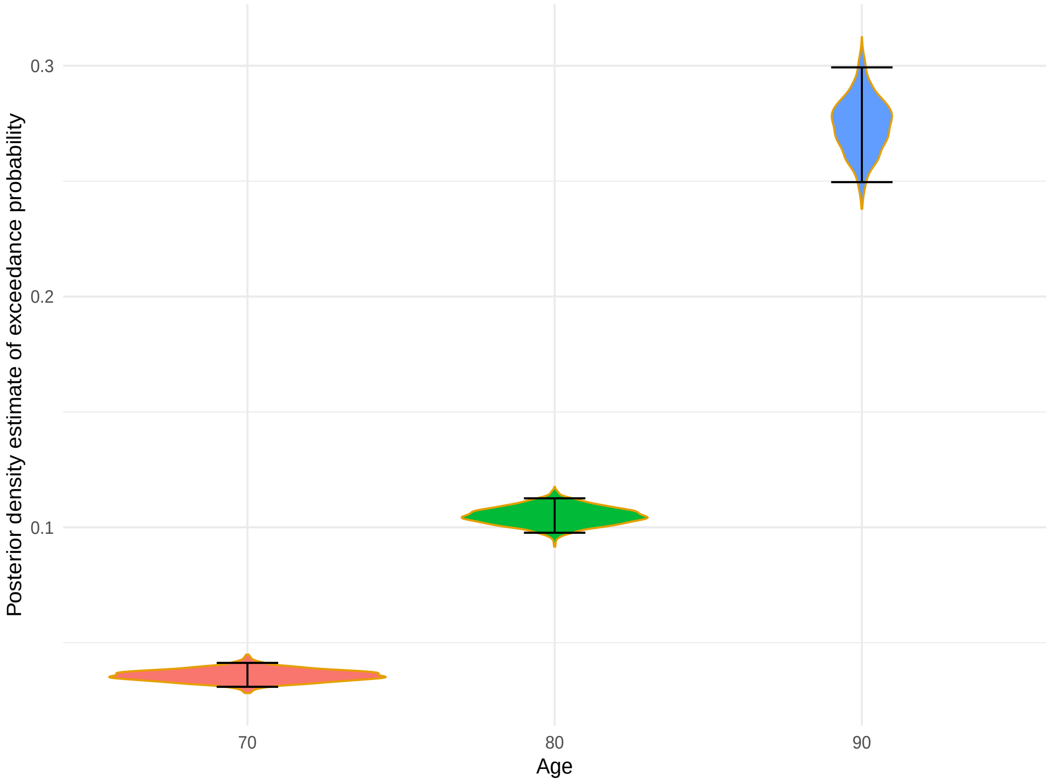

Figure 9: AHEAD study. Moderate exceedance probabilities $\pi = p(y \geq 2 \mid x_{age}, \mathcal{D}_n)$ for $x_{age} = 70, 80,$ and 90. The vertically plotted density summarizes uncertainty about the probability of the exceedance event as a function of model parameters $\phi = (\beta, f_0)$. That is, let $\pi_\phi = p(y \geq 2 \mid x_{age}, \phi)$ (with $\pi = \int \pi_\phi dp(\phi \mid \mathcal{D}_n)$). The violin plots show posterior distribution of $\pi_\phi$; a peaked density indicates less posterior uncertainty about $\pi_\phi$.

# Appendix G: Computation time

We carried out a simulation experiment for assessing scalability of the proposed methodology, in terms of number of covariates ($P$), and also the same for number of observations ($N$). Figures 10 and 11 illustrate that the methodology scales linearly with $N$ and $P$ over a range of reasonable values.

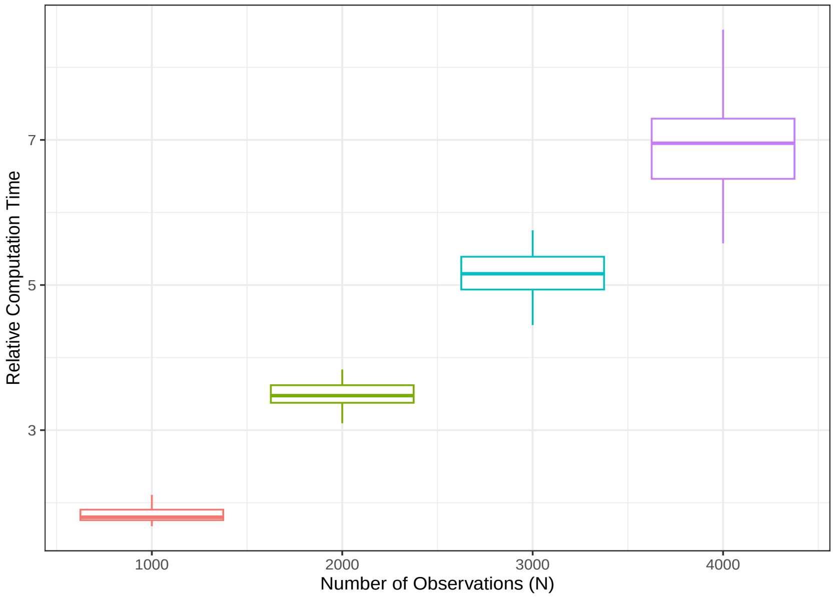

Figure 10: Relative computation time as a function of the number of observations ($N$) for 25 data replicates. We take $N = 500, 1000, 2000, 3000, 4000$ and $P = 10$. Relative computation time is calculated relative to $N = 500$. The absolute computation time with $N = 500$ is approx 1 second per iteration.

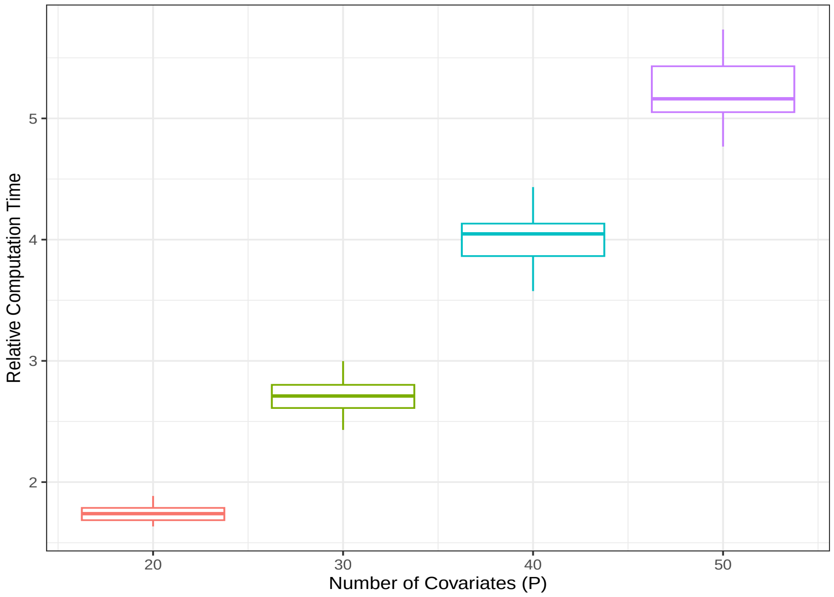

Figure 11: Relative computation time as a function of the number of covariates ($P$) for 25 data replicates. We take $N = 500$ and $P = 10, 20, 30, 40, 50$. Relative computation time is calculated relative to $P = 10$. The absolute computation time with $P = 10$ is approx 1 second per iteration.

# Appendix H: $\theta_i$ solving algorithm

We summarize the discussion on solving $\theta_i$ from Wurm and Rathouz (2018). We use the following notations:

- $n$: sample size
- $S = \{s_1, \ldots, s_K\}$: observed support, with respective frequencies $\{f_1, \ldots, f_k\}$;
- $m = min(S)$; $M = max(S)$
- $\mu_i = E(y_i \mid x_i)$
- $\varepsilon$: convergence threshold, typically $10^{-10}$

We solve for $\theta_i$ such that $b'(\theta_i) = \mu_i = g^{-1}(\eta_{x_i})$, where $g$ is a user-chosen link function and e.g., $\eta_{x_i} = x_i^T\beta$. Equivalently, $g_l\{b'(\theta_i)\} = g_l(\mu_i)$, where $g_l(u) = \text{logit}\left(\frac{u-m}{M-m}\right) = \log\left(\frac{u-m}{M-u}\right)$. The transformation $g_l$ stabilizes the solution. Define $t(\theta_i) = g_l\{b'(\theta_i)\} - g_l(\mu_i)$, which implies $t'(\theta_i) = \frac{M-m}{\{b'(\theta_i)-m\}\{M-b'(\theta_i)\}} b''(\theta_i)$, with $b''(\theta_i) = \sum_{k=1}^{K}\{s_k - b'(\theta_i)\}^2 f_k \exp\{\theta_i s_k - b(\theta_i)\}$. We use Newton-Raphson iterative procedure for finding the root $\theta_i$ of the equation $t(\theta_i) = g_l\{b'(\theta_i)\} - g_l(\mu_i)$, see Algorithm 1. Note that as $\mu_i \to M$ from the left, $\theta_i \to +\infty$. Likewise, as $\mu_i \to m$ from the right, $\theta_i \to -\infty$. To prevent the numerical instability when $\mu_i$ is at or near these boundaries $(m, M)$, we cap $|\theta_i|$ at a maximum value (we use 500 by default).

**Algorithm 1:** Newton-Raphson Procedure for Solving $\theta_i$

```
Initialization: ($\theta_i^{(0)} = 0$, for $i = 1, \ldots, n$)
r ← 0;
repeat
    $\theta^{(r+1)} \leftarrow \theta_i^{(r)} - \{t'(\theta_i^{(r)})\}^{-1} t(\theta_i^{(r)})$;
    $r \leftarrow r + 1$;
until $|t(\theta_i^{(r)})| < \varepsilon$;
$\theta_i \leftarrow \theta_i^{(r)}$;
return $\theta_i$
```

# References

Entejar Alam, Peter Müller, and Paul J. Rathouz. Dir-GLM: A Bayesian GLM with data-driven reference distribution. 2024.

Michael J Wurm and Paul J Rathouz. Semiparametric generalized linear models with the gldrm package. *The R journal*, 10(1):288, 2018.