# ON THE EQUIVALENCE OF OFFSET AND WEIGHTED GLMS AND GBMS

Thomas Bury

Internal Presentation – September 2025

# VOCABULARY

# VOCABULARY & NOTATION

- Data:
  - $n_i \in \{0, 1, 2, \ldots\}$: observed number of claims for cell $i$.
  - $\omega_i > 0$: exposure (e.g. policy-years) for cell $i$.
  - $x_i \in \mathbb{R}^p$: vector of covariates (risk factors).
- Model parameters:
  - $\beta \in \mathbb{R}^p$: regression coefficients.
  - $\eta_i = x_i^\top \beta$: linear predictor.
  - $\mu_i = \exp(\eta_i)$: frequency per unit exposure.
  - $\lambda_i = \omega_i \mu_i$: expected count of claims.
- Tariff cells:
  - Defined by discretization (binning) of risk factors.
  - Within each cell, exposure $\omega_i$ aggregates homogeneous risks.

# POISSON: WEIGHTS AND OFFSET

# TWO WAYS TO INCORPORATE EXPOSURE

## Poisson regression with offset

$$N_i \sim \text{Poisson}(\lambda_i), \quad \lambda_i = \omega_i \mu_i, \quad \log \mu_i = x_i^\top \beta.$$

Equivalent to

$$\log \lambda_i = \log \omega_i + x_i^\top \beta,$$

where $\log \omega_i$ enters as a fixed offset.

## Quasi-likelihood on rates with weights

Define rate $r_i = n_i / \omega_i$. Objective:

$$L_{\text{wt}}(\beta) = \sum_i \omega_i \left[ r_i \eta_i - e^{\eta_i} \right].$$

This is algebraically identical to the offset likelihood (see later proof).

# INTUITION: OFFSET VS WEIGHTS

- **Offset view:** We predict counts. Exposure acts as a known multiplicative factor in *intensity* $\lambda_i$.
- **Weight view:** We normalize by exposure to work with frequencies. To preserve the correct variance structure, we re-weight each observation by $\omega_i$.
- Mathematically: both give the same score and Hessian $\Rightarrow$ identical $\hat{\beta}$.
- Practically: choice is about *implementation* (e.g., GLM vs. LightGBM).

# STEP 1 — POISSON WITH LOG-EXPOSURE OFFSET

**True Poisson log-likelihood (constants in $n_i!$ dropped)**

$$\ell_{\text{off}}(\beta) = \sum_{i=1}^{N}\Big[n_i \log\lambda_i - \lambda_i\Big] = \sum_{i=1}^{N}\Big[n_i(\log\omega_i + \eta_i) - \omega_i e^{\eta_i}\Big].$$

Drop terms independent of $\beta$ to get the optimization objective:

$$\ell_{\text{off}}(\beta) = \sum_{i=1}^{N}\Big[n_i\eta_i - \omega_i e^{\eta_i}\Big].$$

**Score and Hessian**

$$\nabla\ell_{\text{off}} = \sum_{i=1}^{N}(n_i - \omega_i e^{\eta_i})\,x_i, \qquad \nabla^2\ell_{\text{off}} = -\sum_{i=1}^{N}\omega_i e^{\eta_i}\,x_i x_i^{\top}.$$

- Canonical link log $\Rightarrow$ strict concavity in $\beta$ $\Rightarrow$ unique maximizer $\hat{\beta}$.

# STEP 2 — "RATES + WEIGHTS": DEFINE A QUASI-LIKELIHOOD

**Key caveat**

If $N_i$ is Poisson, the rate $R_i \equiv N_i/\omega_i$ is not Poisson. There is no valid "likelihood of the rates."

**Construct a *weighted quasi-likelihood* per unit exposure**

Define $r_i = n_i/\omega_i$ and consider

$$\ell_{\mathrm{wt}}(\beta) = \sum_{i=1}^{N} \omega_i \left[ r_i \eta_i - e^{\eta_i} \right] = \sum_{i=1}^{N} \left[ n_i \eta_i - \omega_i e^{\eta_i} \right].$$

**Score and Hessian (identical to offset)**

$$\nabla \ell_{\mathrm{wt}} = \sum_{i=1}^{N} (n_i - \omega_i e^{\eta_i})\, x_i, \qquad \nabla^2 \ell_{\mathrm{wt}} = -\sum_{i=1}^{N} \omega_i e^{\eta_i}\, x_i x_i^\top.$$

- **Estimator equivalence:** $\ell_{\mathrm{wt}}$ and $\ell_{\mathrm{off}}$ have the same score/Hessian $\Rightarrow$ same $\hat{\beta}$ and same fitted means.

# STEP 3 — WHY THE OPTIMIZERS COINCIDE

1. Offset likelihood objective:

$$\ell_{\text{off}}(\beta) = \sum_i \left[ n_i \eta_i - \omega_i e^{\eta_i} \right].$$

2. Quasi-likelihood (rates + weights):

$$\ell_{\text{wt}}(\beta) = \sum_i \omega_i \left[ (n_i/\omega_i)\eta_i - e^{\eta_i} \right] = \sum_i \left[ n_i \eta_i - \omega_i e^{\eta_i} \right].$$

3. Therefore $\ell_{\text{off}} \equiv \ell_{\text{wt}}$ as functions of $\beta$ (up to constants), with identical score/Hessian.
4. Strict concavity under log link $\Rightarrow$ unique maximizer $\hat{\beta}$ in both cases.

**What this statement *is***

Equality of *optimization objectives*/estimators (likelihood vs. quasi-likelihood), not equality of a true likelihood for $R_i$.

# CONDITIONS FOR EQUIVALENCE & WHERE IT FAILS

**Holds when**

- Poisson GLM with canonical log link.
- Objective is the sum of contributions (no hidden normalization changes).
- No extra penalties, or penalties scaled consistently across parameterizations.

**Does *not* hold when**

- Using non-canonical links or different distributions (e.g., NB with logit of mean-variance).
- Using squared error on rates (MSE on $r_i$) instead of the Poisson objective.
- Computing likelihood-based ICs (AIC) from the “rates” objective (it is not a true likelihood).

# PRACTICAL CAVEATS — POISSON GBM (LIGHTGBM)

**Two equivalent ways to encode exposure in LightGBM**

1. **Sample weights:** use `objective="poisson"`, labels $y_i = r_i$, and `sample_weight` $= \omega_i$.
2. **Offset via init_score:** labels $y_i = n_i$, set `init_score` $= \log \omega_i$ (per-row constant added to raw score), weights $= 1$.

Both yield identical per-point gradients/Hessians:

$$g_i = \omega_i e^{f_i} - n_i, \qquad h_i = \omega_i e^{f_i},$$

with $f_i$ the current raw score (including `init_score` when used).

# PRACTICAL CAVEATS — POISSON GBM (LIGHTGBM)

**Which should we use in a dataset class?**

- Default to `sample_weight` $= \omega_i$ for LightGBM Poisson: simplest, robust across CV/early stopping/metrics.
- Use `init_score` $= \log \omega_i$ when you need an explicit offset (e.g., to align with GLM formulations, or to pass the same offset into scoring on holdout).
- Provide a toggle and test that the two modes produce numerically equivalent training curves on the same seed.

# PRACTICAL CAVEATS — METRICS, REGULARIZATION, STABILITY

- **Tiny exposure** small exposure values can break the optimization or prevent convergence.
- **Metrics & early stopping:** ensure validation metrics are computed with the *same* weights $\omega_i$ if using `sample_weight`. Mixing weighted training with unweighted early stopping biases model selection.
- **Regularization scaling:** many libs minimize averaged loss. If you switch between offset and weights, the effective strength of $\ell_1/\ell_2$ penalties may change. Scale penalties with total weight.
- **Prediction reconstruction:** with `sample_weight`, model learns $\eta_i$; compute $\hat{\lambda}_i = \omega_i \exp(\hat{\eta}_i)$. With `init_score`, the offset is already baked into the raw score.
- **Large exposures:** extremely large $\omega_i$ can dominate gradients. Consider capping exposures or grouping to stabilize training; always inspect leverage points.
- **Don't use MSE on rates:** it breaks the Poisson equivalence; stick to the Poisson objective for count data with exposure.

# BOTTOM LINE & REFERENCES

- For Poisson/log-link GLMs, **log-exposure offset** and **rates + weights** yield the **same estimator** (identical score/Hessian).
- In LightGBM, prefer `sample_weight` for simplicity; support `init_score` when an explicit offset is required. Validate equivalence in unit tests.

Statement noted in [3] (no proof). See also GLM treatments for the Exponential Dispersion Family under canonical links in [1].

# TWEEDIE: OFFSET VS WEIGHTED RATES

# TWEEDIE DISTRIBUTION (SETUP)

Let $Y_i$ be aggregate claims for risk $i$, with exposure $\omega_i > 0$ and covariates $x_i$. We define:

$$R_i = \frac{Y_i}{\omega_i}, \qquad \mu_i = \mathbb{E}[R_i], \qquad \tilde{\mu}_i = \mathbb{E}[Y_i] = \omega_i \mu_i.$$

We use a Tweedie GLM with power index $p \in (1, 2]$ and a log link:

$$\eta_i = x_i^\top \beta, \quad \mu_i = e^{\eta_i}, \quad \log \tilde{\mu}_i = \log \omega_i + \eta_i.$$

**Tweedie exponential dispersion family**

The density (up to a normalising constant $c$) is

$$f(y_i; \tilde{\mu}_i, \phi, p) = \exp\left\{ \frac{y_i \tilde{\mu}_i^{1-p}}{(1-p)\,\phi} - \frac{\tilde{\mu}_i^{2-p}}{(2-p)\,\phi} + c(y_i, \phi, p) \right\},$$

with variance function

$$\mathrm{Var}(Y_i) = \phi\, \tilde{\mu}_i^{p}.$$

# ITERATIVELY REWEIGHTED LEAST SQUARES (IRLS)

In a GLM, the coefficients $\beta$ are updated by solving a weighted least squares problem. At iteration $t$:

$$\beta^{(t+1)} = (X^\top W^{(t)} X)^{-1} X^\top W^{(t)} z^{(t)}.$$

**Working quantities**

$$\eta_i^{(t)} = x_i^\top \beta^{(t)} \qquad \text{(linear predictor)}$$

$$\mu_i^{(t)} = g^{-1}(\eta_i^{(t)}) \qquad \text{(fitted mean)}$$

$$D_i^{(t)} = \left.\frac{d\mu}{d\eta}\right|_{\eta=\eta_i^{(t)}} \qquad \text{(derivative of link)}$$

$$z_i^{(t)} = \eta_i^{(t)} + \frac{y_i - \mu_i^{(t)}}{D_i^{(t)}} \qquad \text{(working response)}$$

$$w_i^{(t)} = \frac{1}{\phi} \frac{(D_i^{(t)})^2}{V(\mu_i^{(t)})} \qquad \text{(working weight)}$$

# TWEEDIE IRLS (LOG LINK)

For Tweedie GLM with $V(\mu) = \mu^p$ and $g(\mu) = \log \mu$:

$$\mu_i = e^{\eta_i}, \qquad \frac{d\mu_i}{d\eta_i} = \mu_i.$$

**Offset form** (aggregate mean $\tilde{\mu}_i = \omega_i \mu_i$)

$$z_i^{\text{off}} = \log \tilde{\mu}_i + \frac{y_i - \tilde{\mu}_i}{\tilde{\mu}_i}, \qquad w_i^{\text{off}} = \frac{1}{\phi}(\tilde{\mu}_i)^{2-p} = \frac{1}{\phi}(\omega_i \mu_i)^{2-p}.$$

**Rates + weights form** ($r_i = y_i/\omega_i$, weight $s_i = \omega_i^{2-p}$)

$$z_i^{\text{wt}} = \eta_i + \frac{r_i - \mu_i}{\mu_i}, \qquad w_i^{\text{wt}} = \frac{s_i}{\phi}\,\mu_i^{2-p}.$$

# TWEEDIE SCORE WITH LOG-EXPOSURE OFFSET (QUASI-LIKELIHOOD)

## Quasi log-likelihood (ignoring $y$-only constants)

$$\ell_{\text{off}}(\beta) \;\propto\; -\sum_{i=1}^{N} \frac{1}{\phi}\left[\frac{(\tilde{\mu}_i)^{2-p}}{2-p} - \frac{y_i(\tilde{\mu}_i)^{1-p}}{1-p}\right], \qquad \tilde{\mu}_i = \omega_i e^{\eta_i}.$$

## Score function (log link)

Let $\tilde{\eta}_i = \log \tilde{\mu}_i = \log \omega_i + \eta_i$. Then

$$\nabla \ell_{\text{off}} = \sum_{i=1}^{N} \frac{y_i - \tilde{\mu}_i}{\phi\,(\tilde{\mu}_i)^p} \frac{d\tilde{\mu}_i}{d\tilde{\eta}_i} x_i = \sum_{i=1}^{N} \frac{y_i - \tilde{\mu}_i}{\phi\,(\tilde{\mu}_i)^p} \tilde{\mu}_i x_i = \sum_{i=1}^{N} \frac{y_i - \tilde{\mu}_i}{\phi} (\tilde{\mu}_i)^{1-p} x_i.$$

# TWEEDIE SCORE WITH PER-EXPOSURE NORMALIZATION (QUASI-LIKELIHOOD)

Define rates $R_i \equiv Y_i/\omega_i$ so that $\mathbb{E}[R_i] = \mu_i$ and

$$\mathrm{Var}(R_i) = \frac{\mathrm{Var}(Y_i)}{\omega_i^2} = \frac{\phi\,(\tilde{\mu}_i)^p}{\omega_i^2} = \frac{\phi}{\omega_i^{2-p}}\,\mu_i^p.$$

Thus the **effective dispersion** for $R_i$ is $\phi_r = \phi/\omega_i^{2-p}$.
prior/sample weight $s_i = \omega_i^{2-p} \quad \Longleftrightarrow \quad \phi_r = \phi/s_i$.
Quasi-log-likelihood (up to $r_i$-only constants): $\ell_{\mathrm{wt}}(\beta) \propto -\sum_{i=1}^N \frac{s_i}{\phi}\left[\frac{\mu_i^{2-p}}{2-p} - \frac{r_i\mu_i^{1-p}}{1-p}\right]$.

## Score (log-link)

With $\eta_i = \log\mu_i$,

$$\nabla\ell_{\mathrm{wt}} = \sum_{i=1}^N \frac{s_i}{\phi}\,\frac{r_i-\mu_i}{\mu_i^p}\,\frac{d\mu_i}{d\eta_i}\,x_i = \sum_{i=1}^N \frac{s_i}{\phi}\,(r_i-\mu_i)\,\mu_i^{1-p}x_i.$$

# TWEEDIE EQUIVALENCE (LOG LINK)

**Key result**

With $s_i = \omega_i^{2-p}$ and $r_i = y_i/\omega_i$:

$$\nabla \ell_{\text{off}}(\beta) = \nabla \ell_{\text{wt}}(\beta), \quad W_i^{\text{off}} = W_i^{\text{wt}}.$$

Thus the two approaches give identical estimating equations and Hessian. Special cases:

- $p = 1$ (Poisson): $s_i = \omega_i$.
- $p = 2$ (Gamma): $s_i = 1$, exposure affects only the mean.

**Conclusion**

With a **log link** and **Tweedie variance** $V(\mu) = \mu^p$, the offset and rates+weights formulations yield identical score functions and IRLS weights. Hence, when the GLM normal equations have a unique solution (full column rank), they produce the **same estimator** $\hat{\beta}$ and the same fitted means.

# CONDITIONS FOR EQUIVALENCE & WHEN IT FAILS (TWEEDIE)

## Holds when

- **Log link** on the mean: $\log \tilde{\mu} = \log \omega + x^\top \beta$ so exposure is a true additive offset.
- Tweedie GLM with $V(\mu) = \mu^p$ and $p \in (1, 2]$ (Poisson $p$=1, Gamma $p$=2 as limits).
- In *rates* formulation, use **weights** $s_i = \omega_i^{2-p}$.

## Fails or changes when

- **Non-log links:** $g(\omega\mu) \neq g(\mu) + \text{const}$, so exposure cannot be a fixed offset; equivalence breaks.
- **Wrong weights on rates:** using $s_i = \omega_i$ (Poisson-style) when $p \neq 1$ distorts the variance and the estimator.
- **MSE on rates:** replacing the Tweedie objective by squared error destroys the equivalence.
- **Penalties/averaging:** if losses are averaged, rescale $\ell_1/\ell_2$ by total weight to keep regularization comparable across parameterizations.

**REM:** $p$=1 (Poisson): $s_i = \omega_i$ and $p$=2 (Gamma): $s_i = \omega_i^0 = 1$ (exposure offset only; no reweighting).

# PRACTICAL CAVEATS — TWEEDIE GBM (LIGHTGBM)

LightGBM uses a log link for `objective="tweedie"` with index $p \in (1,2)$. Let the raw score be $f_i$ and $\hat{\mu}_i = \exp(f_i)$.

## Two equivalent encodings of exposure

1. **Offset on totals:** train on $y_i$ with

$$\texttt{init\_score}_i = \log \omega_i, \quad \texttt{sample\_weight}_i = 1.$$

2. **Weighted rates:** train on $r_i = y_i/\omega_i$ with

$$\texttt{init\_score}_i = 0, \quad \texttt{sample\_weight}_i = \omega_i^{2-p}.$$

# PRACTICAL CAVEATS — TWEEDIE GBM (LIGHTGBM)

**Gradient/Hessian equality ($\phi$ acts as a dispersion constant and cancels in tree growth.)**

Per point (lightgbm uses first/second derivatives of the loss):

$$g_i \propto (\hat{\tilde{\mu}}_i)^{2-p} - y_i\,(\hat{\tilde{\mu}}_i)^{1-p}, \qquad h_i \propto (2-p)\,(\hat{\tilde{\mu}}_i)^{2-p} - (1-p)\,y_i\,(\hat{\tilde{\mu}}_i)^{1-p}.$$

With Option 1, $\hat{\tilde{\mu}}_i = \omega_i e^{\eta_i}$. With Option 2, $g_i, h_i$ are multiplied by $\omega_i^{2-p}$ and $y_i$ is replaced by $r_i = y_i/\omega_i$, yielding the **same** $g_i, h_i$ numerically. Hence identical trees (same seed) and the same fitted means.

- **Metrics/early stopping:** If training on rates, compute validation metrics with the same weights $\omega^{2-p}$ (or reconstruct totals).
- **Large exposures:** leverage grows like $\omega^{2-p}$; inspect/cap extreme exposures to stabilize training.

# INTUITION AND TAKEAWAYS (TWEEDIE)

- **Offset view:** exposure scales the *mean* linearly; variance grows as $(\omega\mu)^p$.
- **Rates view:** dividing by $\omega$ attenuates variance by $\omega^{2-p}$; weighting by $\omega^{2-p}$ restores the correct contribution to the score/Hessian.
- **Unification:** Poisson ($p$=1) $\Rightarrow$ weights $\omega$; Gamma ($p$=2) $\Rightarrow$ no reweighting; in-between smoothly interpolates.
- **Rule of thumb:** *Log link* + *weights* = $\omega^{2-p}$ is the exact Tweedie analogue of the Poisson offset/weights equivalence.

# TWEEDIE REGRESSION

# EXPONENTIAL DISPERSION FAMILY (EDF) SETUP

- General EDF density:

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)\right\},$$

  with mean $\mu = b'(\theta)$ and variance $\mathrm{Var}(Y) = \phi V(\mu)$.

- For GLMs:

$$g(\mu_i) = x_i^\top \beta, \qquad g(\cdot) : \text{link function.}$$

- Exposure $\omega_i$ enters as:

$$\theta_i \mapsto \theta_i + \log \omega_i \quad \text{(offset formulation).}$$

# TWEEDIE DISTRIBUTION

- The **Tweedie** family: variance function

$$V(\mu) = \mu^p, \qquad 1 < p < 2 \quad \Rightarrow \quad \text{compound Poisson–Gamma.}$$

- Mean/variance structure:

$$\mathbb{E}[Y_i] = \mu_i, \qquad \mathrm{Var}(Y_i) = \phi\mu_i^p.$$

- Used in insurance for claim cost modeling:
  - $p = 1 \Rightarrow$ overdispersed Poisson (frequency).
  - $p = 2 \Rightarrow$ Gamma (severity).
  - $1 < p < 2 \Rightarrow$ compound Poisson–Gamma (aggregate claims).

# OFFSET VS WEIGHTS IN TWEEDIE GLMS

## Offset approach

$$Y_i \sim \mathrm{Tweedie}(\mu_i \omega_i, \phi, p), \qquad \log \mu_i = x_i^\top \beta.$$

$$\Rightarrow \log \mathbb{E}[Y_i] = \log \omega_i + x_i^\top \beta.$$

## Rates + weights approach

$$R_i = \frac{Y_i}{\omega_i}, \qquad \mathbb{E}[R_i] = \mu_i.$$

Variance:

$$\mathrm{Var}(R_i) = \frac{\phi}{\omega_i^{2-p}} \mu_i^p.$$

**Weighting by** $\omega_i^{2-p}$ restores correct variance scaling.

# GENERAL EQUIVALENCE CONDITIONS

- For Poisson ($p = 1$):

$$\mathrm{Var}(R_i) = \frac{\mu_i}{\omega_i} \quad \Rightarrow \quad \text{weight } \omega_i.$$

  Matches the offset equivalence.

- For Tweedie $1 < p < 2$:

$$\mathrm{Var}(R_i) = \frac{\phi}{\omega_i^{2-p}} \mu_i^p.$$

  Requires weight $\omega_i^{2-p}$ to match offset formulation.

- Thus:

**Equivalence principle**

Offset and weighted-rates formulations yield identical score/Hessian *only if* weights are scaled as $\omega_i^{2-p}$.

# INTUITION IN TWEEDIE REGRESSION

- **Offset view:** exposure scales expected aggregate claim cost linearly.
- **Weights view:** normalizing by exposure alters variance structure; weighting by $\omega_i^{2-p}$ corrects this.
- **Poisson special case:** $p = 1$, weight reduces to $\omega_i$ (what we already proved).
- **Gamma special case:** $p = 2$, weight reduces to 1 — exposure has no variance effect (modeling average cost).
- Practical note: in software libraries (e.g. LightGBM, XGBoost), Tweedie objectives implicitly assume offset parameterization; manual reweighting may be needed if working with normalized rates.

# REFERENCES


Ohlsson, Esbjörn and Johansson, Björn

Non Non-life insurance pricing with generalized linear models

Springer, 2010

Kaivanipour, Kivan

Non-Life Insurance Pricing Using the Generalized Additive Model, Smoothing Splines and L-Curves

2015

Anderson, D and Feldblum, S and Modlin, C and Schirmacher, D and Schirmacher, E and Thandi, N

A Practitioner's Guide to Generalized Linear Models: A Foundation for Theory

Interpretation and Application, Watson Wyatt Worldwide, London

2005