Robust penalized empirical likelihood estimation method for linear regression (1-22).

✅ **Prerequisites:**

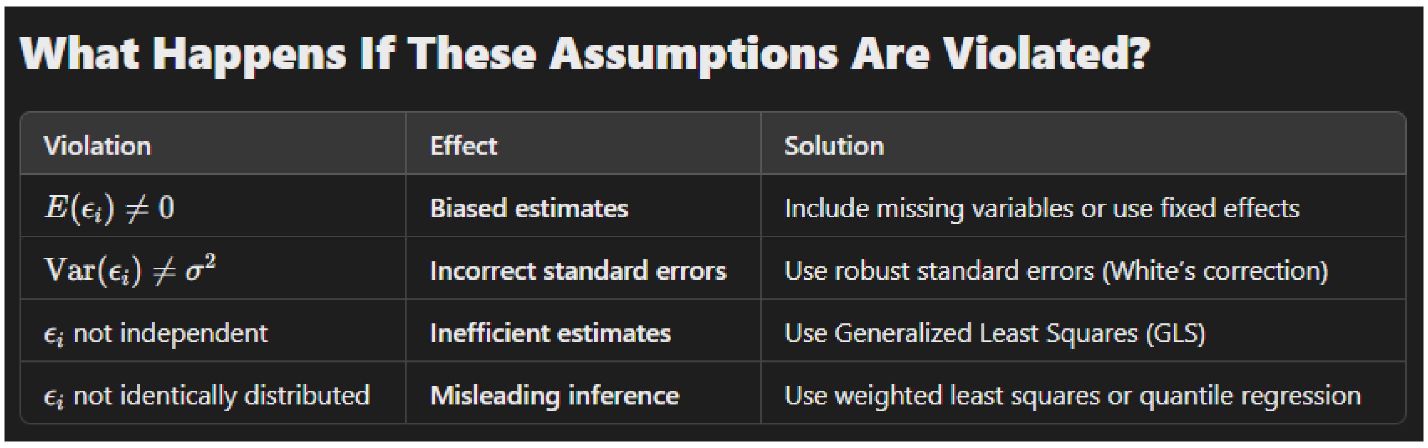
1. Linear Regression
2. Ordinary Least Squares
3. Maximum Likelihood Estimation
4. Empirical Likelihood
5. Robust Estimation
6. MM-Estimators
7. Penalized Regression
8. Asymptotic Properties

[INTRODUCTION] Normal Linear Regression Model:

$Y_i = X_i^T\beta + \epsilon_i$, for i=1, 2, → n.

$Y_i \in R \;\; and \;\; X_i \in R^k \;\; and \;\; \beta \in R^k$
Y → response variable.
X → k-dimensional vector of explanatory variables.
Beta → Unknown parameter vector we want to estimate.
Epsilon → Error Term with:

$E(\epsilon_i) = 0 \;\; and \;\; Var(\epsilon_i) = \sigma^2.$



What Happens If These Assumptions Are Violated?

Violation	Effect	Solution
$E(\epsilon_i) \neq 0$	**Biased estimates**	Include missing variables or use fixed effects
$\mathrm{Var}(\epsilon_i) \neq \sigma^2$	**Incorrect standard errors**	Use robust standard errors (White's correction)
ϵ_i not independent	**Inefficient estimates**	Use Generalized Least Squares (GLS)
ϵ_i not identically distributed	**Misleading inference**	Use weighted least squares or quantile regression

→ The goal is regression is to estimate the parameter vector β and to select the significant explanatory variables (variable selection).

Two classical methods for estimating regression parameters:
1. Least Squares Estimation.
2. Maximum Likelihood Estimation.

1. Least Squares Estimation:

The least squares method minimizes the sum of squared differences (residuals) between the observed Y_i and the predicted $X_i^T\beta$.

$\hat{\beta_{LS}} = arg\ min_{\beta}\ \sum_{i=1}^{n}(Y_i - X_i^T\beta)^2$.
Optimal under Gauss-Markov assumptions,
1. Errors Follow:
$E(\epsilon_i) = 0\ \ and\ \ Var(\epsilon_i) = \sigma^2$
2. Errors are uncorrelated.
3.
X_i are non-random and linearly independent.

2.
Maximum Likelihood Estimation:

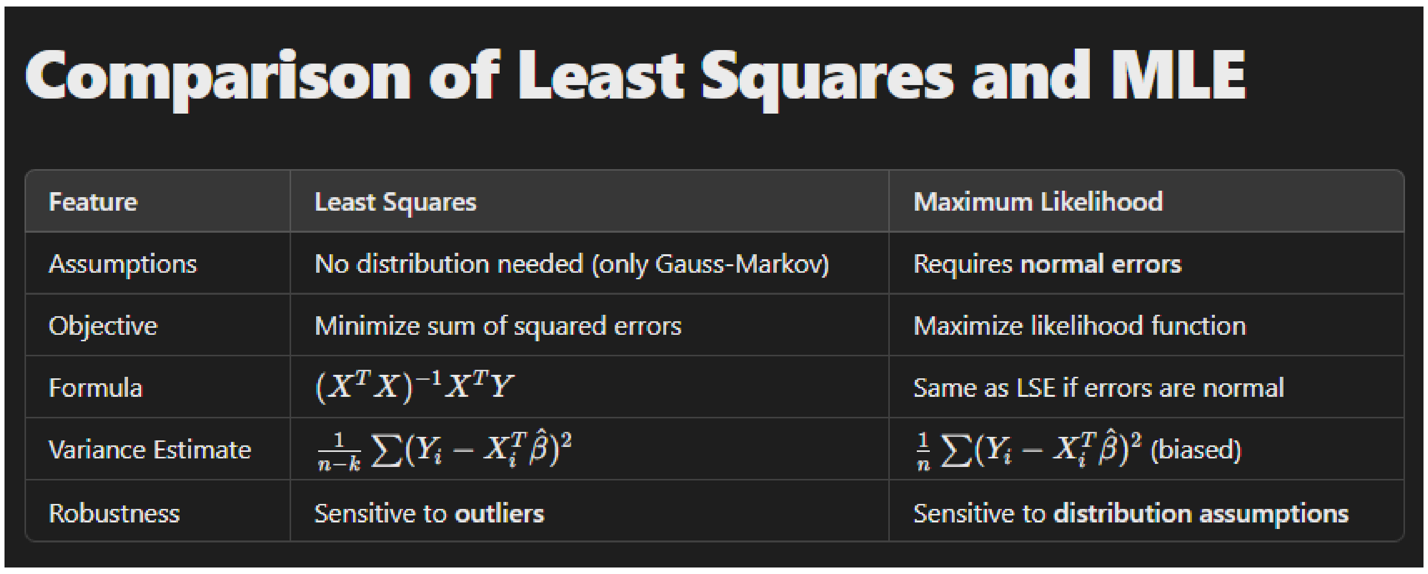
Maximum Likelihood estimation is a probabilistic method that finds values of β that maximize the likelihood of observing the given data.

Assuming the errors ϵ_i are normally distributed, $\epsilon_i \sim N(0, \sigma^2)$.

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(Y_i - X_i^T\beta)^2}{2\sigma^2}).$$

The ML estimator's beta and sigma are obtained by maximizing this likelihood function. Under normality. The LS and ML estimators for beta are the same.

The paper points out that the LS and ML estimators are sensitive to outliers and non-normality of the error terms. This motivates the need for robust estimation methods.



Comparison of Least Squares and MLE

Feature	Least Squares	Maximum Likelihood
Assumptions	No distribution needed (only Gauss-Markov)	Requires **normal errors**
Objective	Minimize sum of squared errors	Maximize likelihood function
Formula	$(X^TX)^{-1}X^TY$	Same as LSE if errors are normal
Variance Estimate	$\frac{1}{n-k}\sum(Y_i - X_i^T\hat{\beta})^2$	$\frac{1}{n}\sum(Y_i - X_i^T\hat{\beta})^2$ (biased)
Robustness	Sensitive to **outliers**	Sensitive to **distribution assumptions**

Error term in regression:

Residual or Disturbance Term, is the difference between observed response and predicted response.

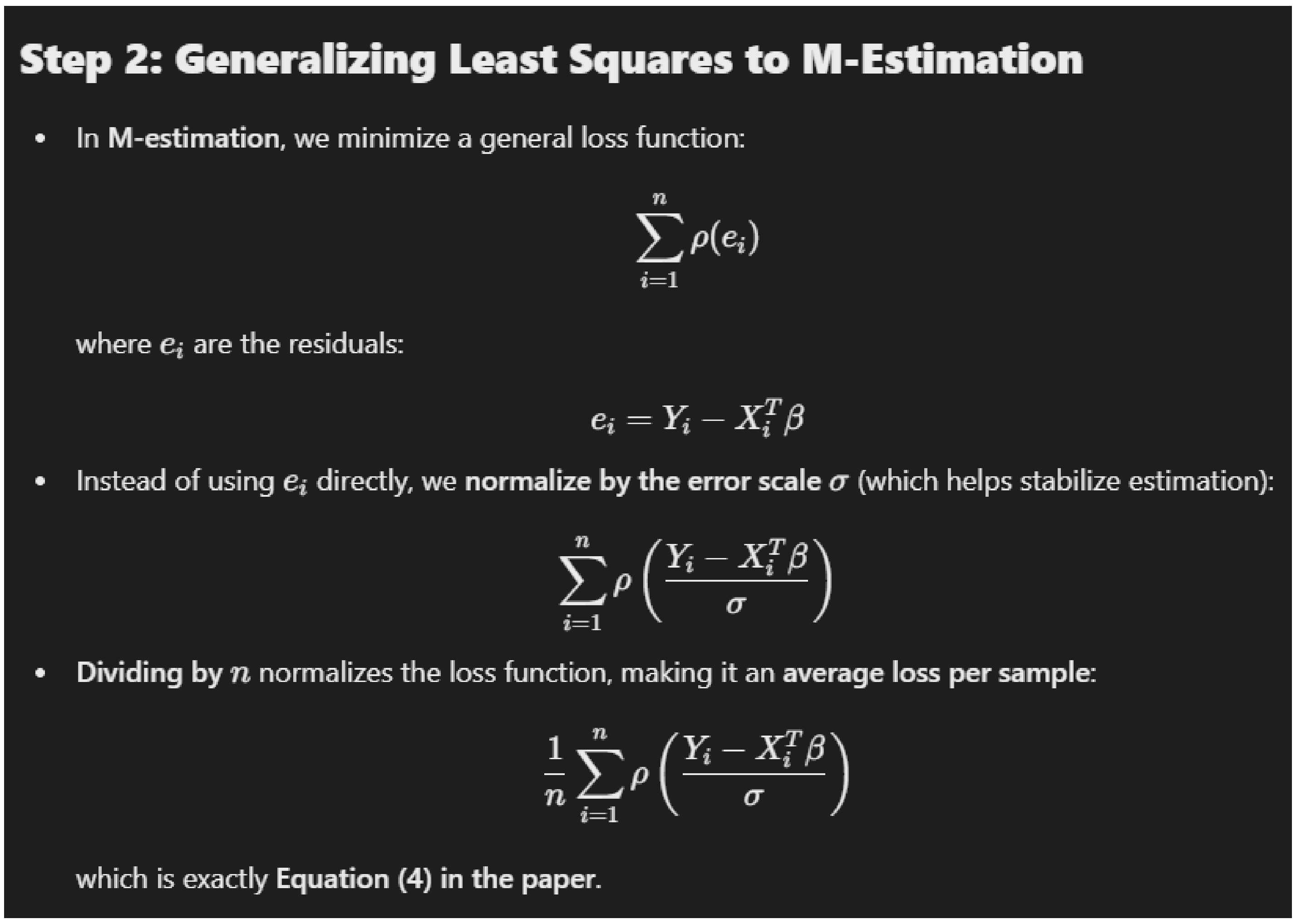
$\epsilon_i = Y_i - X_i^T\beta$.
It accounts for Unobserved factors, Measurement errors and Random variations.
Assumptions:
1.
Zero Mean Assumption: On average the errors are zero, the model $X_i^T\beta$ correctly captures the systematic part of the relationship between X_i and Y_i.
2.
Homoscedasticity (Constant Variance): The variance of of the errors is constant across all observations. This means that the spread of the errors is the same for all values of X_i.
3.
No Autocorrelation (Independence of errors):
4.
Normality of Errors (Optional for LS, Required for ML):

Problem of Outliers:
1. Least Squares Estimation: SSR gives disproportionate weight to large residuals, pulling regression line towards the outlier - leading to biased estimates.
2. Maximum Likelihood Estimation: If errors are not normally distributed, the outliers can distort the likelihood function - leading to biased estimates.

Robust Parameter Estimation Methods:
The first mentioned method is the M-estimation method; (Huber's Method)
M-Estimation Generalizes OLS*

$\frac{1}{n}\sum_{i=1}^{n}\rho(\frac{Y_i-X_i^T\beta}{\sigma}),$

ρ is a loss function that is less sensitive to outliers.



Step 2: Generalizing Least Squares to M-Estimation

- In **M-estimation**, we minimize a general loss function:

$$\sum_{i=1}^{n}\rho(e_i)$$

where e_i are the residuals:

$$e_i = Y_i - X_i^T\beta$$

- Instead of using e_i directly, we **normalize by the error scale** σ (which helps stabilize estimation):

$$\sum_{i=1}^{n}\rho\left(\frac{Y_i - X_i^T\beta}{\sigma}\right)$$

- **Dividing by** n normalizes the loss function, making it an **average loss per sample**:

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i - X_i^T\beta}{\sigma}\right)$$

which is exactly **Equation (4) in the paper**.

ρ is a non-negative, non-decreasing function $\rho(0) = 0$ and σ is a scale parameter. Common choices for ρ are:

Huber's function and Tukey's bi-square function.

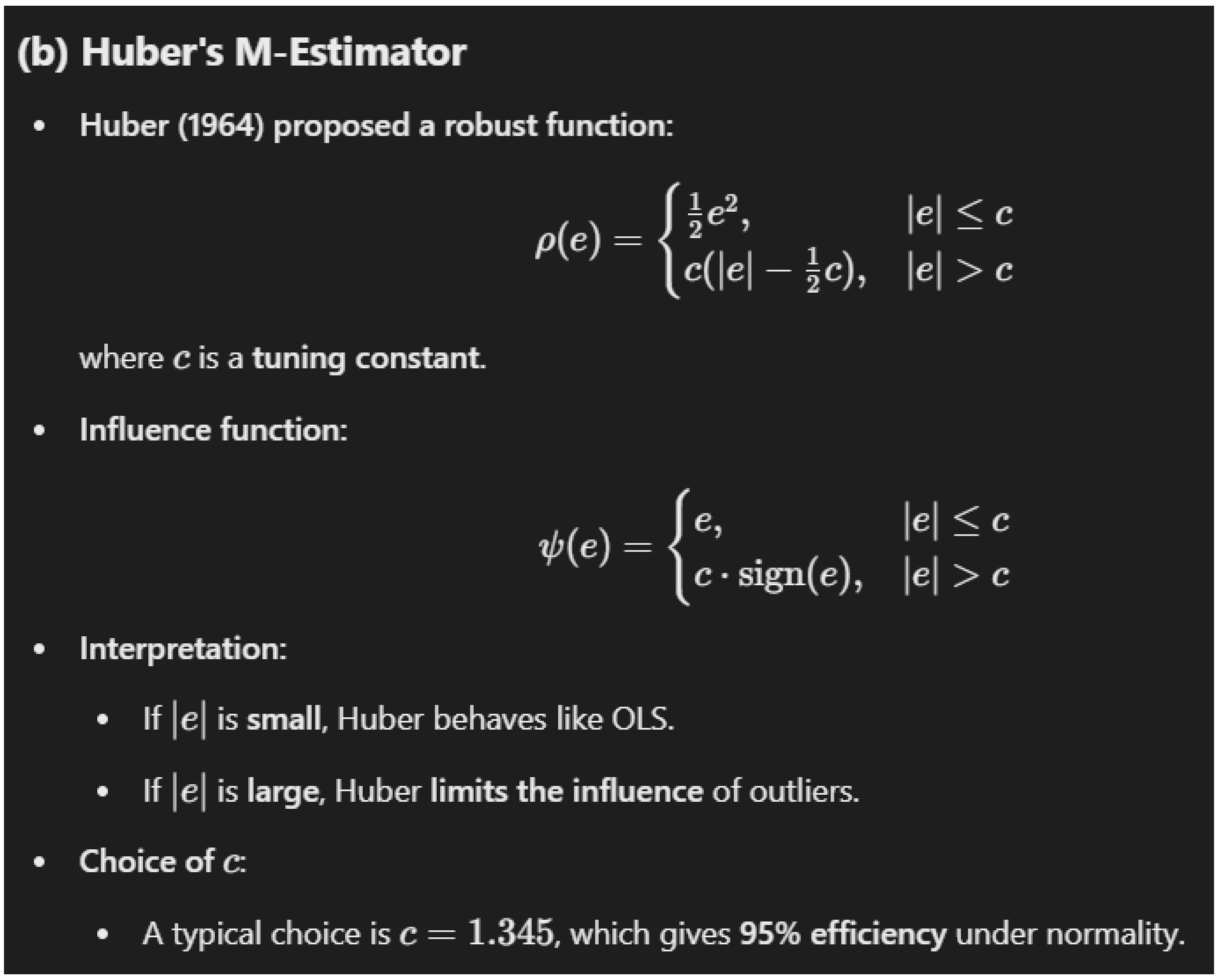
The Huber ρ function is a piecewise monotone increasing function for positive arguments, but it increases less rapidly than the squared function (loss function of the OLS method or log-likelihood function under normality). The derivative of Huber ρ function is a bounded function. The Tukey ρ
function is a bounded function and its derivative is a redescending function.

Key Points about M-Estimation:

Robustness: The ρ function is chosen to reduce the effect of the outliers - Common choices are:

Huber's function: Combines quadratic and linear behavior to limit the influence of large residuals.

Tukey's Bi Square: Bounded, meaning it completely ignores the influence of large residuals.



(b) Huber's M-Estimator

- **Huber (1964) proposed a robust function:**

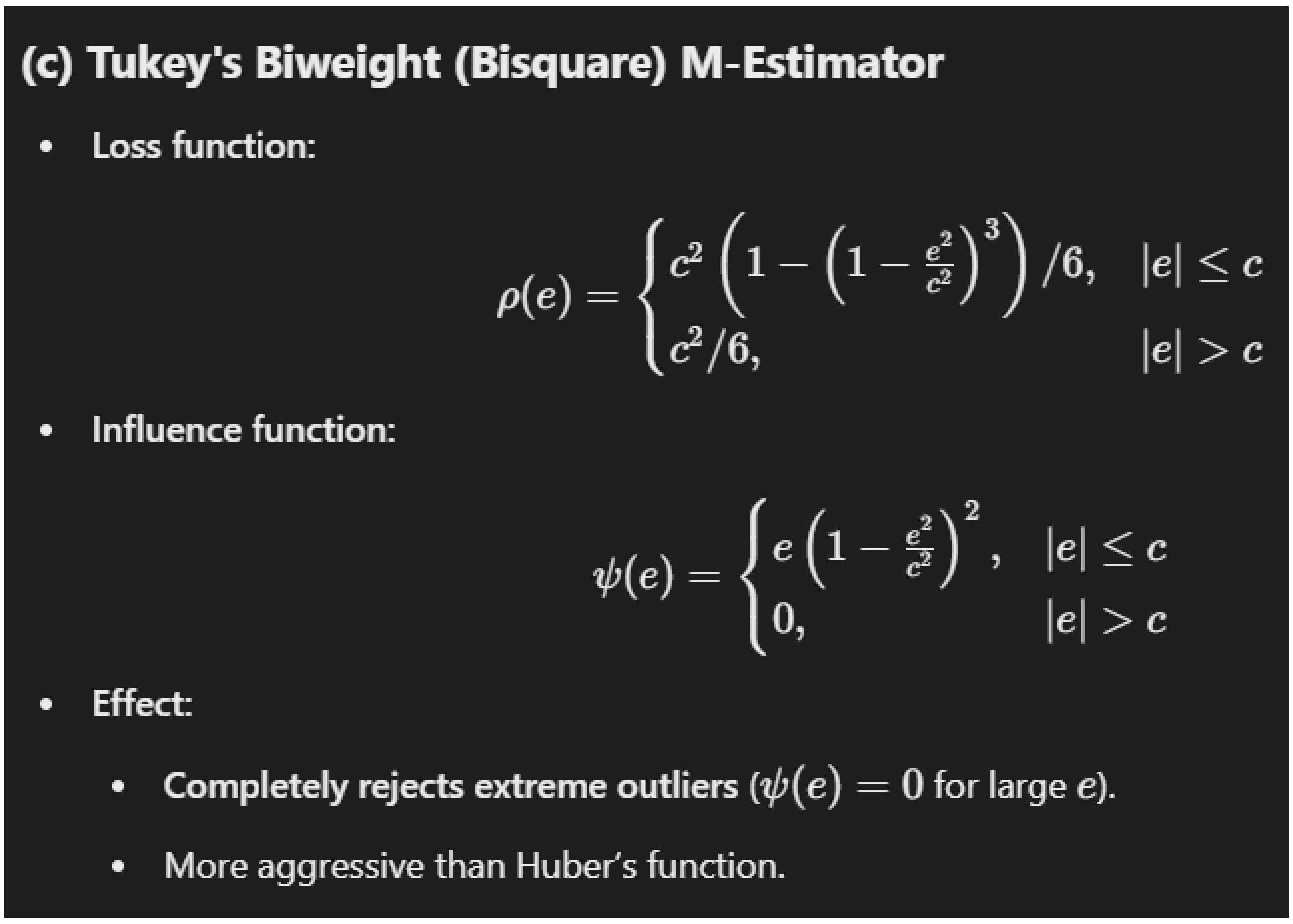
$$\rho(e) = \begin{cases} \frac{1}{2}e^2, & |e| \leq c \\ c(|e| - \frac{1}{2}c), & |e| > c \end{cases}$$

where c is a **tuning constant**.

- **Influence function:**

$$\psi(e) = \begin{cases} e, & |e| \leq c \\ c \cdot \text{sign}(e), & |e| > c \end{cases}$$

- **Interpretation:**
  - If $|e|$ is **small**, Huber behaves like OLS.
  - If $|e|$ is **large**, Huber **limits the influence** of outliers.
- **Choice of c:**
  - A typical choice is $c = 1.345$, which gives **95% efficiency** under normality.



(c) Tukey's Biweight (Bisquare) M-Estimator

- **Loss function:**

$$\rho(e) = \begin{cases} c^2\left(1 - \left(1 - \frac{e^2}{c^2}\right)^3\right)/6, & |e| \leq c \\ c^2/6, & |e| > c \end{cases}$$

- **Influence function:**

$$\psi(e) = \begin{cases} e\left(1 - \frac{e^2}{c^2}\right)^2, & |e| \leq c \\ 0, & |e| > c \end{cases}$$

- **Effect:**
  - **Completely rejects extreme outliers** ($\psi(e) = 0$ for large e).
  - More aggressive than Huber's function.

Scale Parameter: The M-estimator depends on the scale parameter σ , which must be estimated robustly. Two approaches are mentioned:

1.

Two Step Estimation: First estimate σ robustly. (Using median absolute deviation), then estimate β.

2.

Simultaneous Estimation: Estimate σ and β simultaneously using an iterative procedure.

What is the scale parameter -

σ represents the spread of the error terms. In **Ordinary Least Squares (OLS):**

$\hat{\sigma}^2 \; = \; \frac{1}{n-k}\sum_{i=1}^{n}(Y_i - X_i^T\hat{\beta})^2.$ → Not robust to outliers.

If

σ too small → residuals become too large, making the estimation sensitive to small changes.
If
σ too large → The model becomes insensitive to variation.

1.
Two Step Estimation:
Step 1:

$$\hat{\sigma} \;=\; 1.4826 \;*\; (|Y_i - X_i^T \hat{\beta^{(0)}}|)$$

MAD is
based on medians, which are resistant to **outliers**. The constant **1.4826** scales the MAD to match the standard deviation if the data is normally distributed.

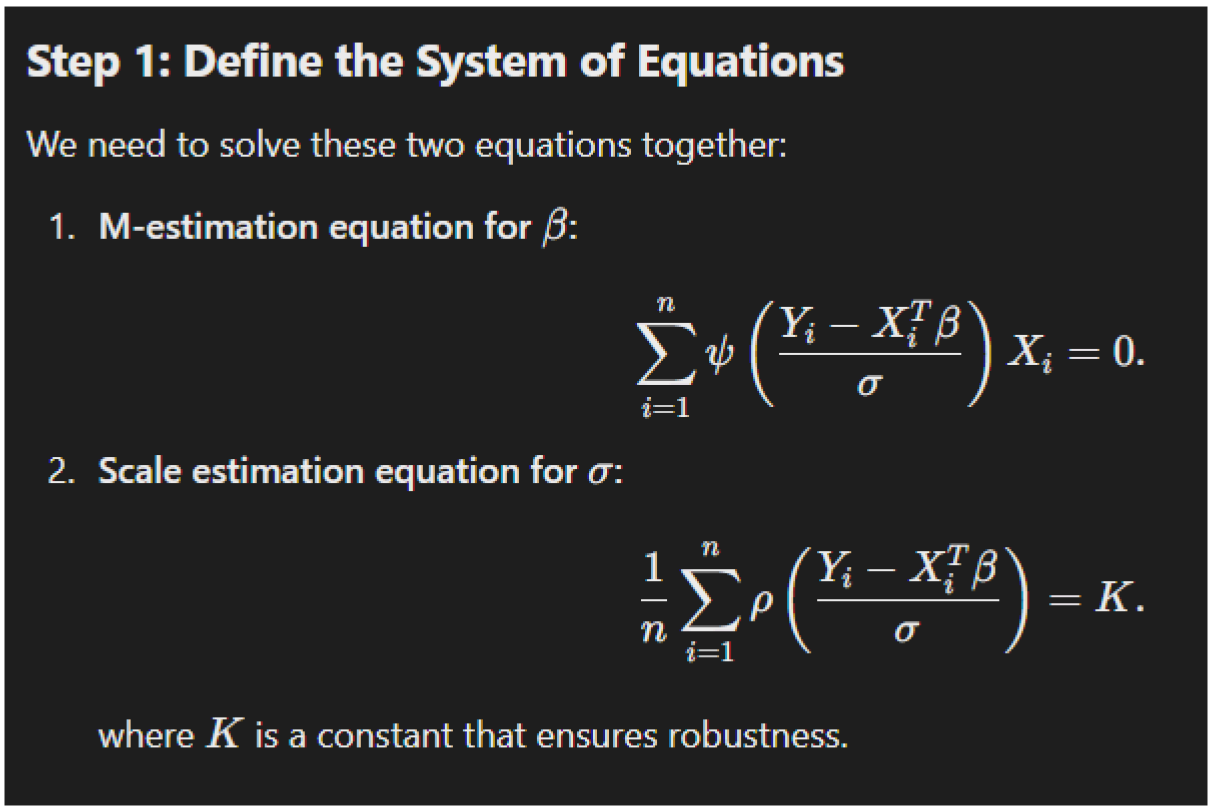
Step 2:
$\sum_{i=1}^{n} \psi(\frac{Y_i - X_i^T \beta}{\hat{\sigma}}) X_i \;=\; 0.$
This ensures
β is robustly estimated.

$\psi(e)$ is the influence function, derived from the derivative of robust loss function $\rho(e)$.

$$\psi(e) \;=\; \frac{d}{de}\rho(e)$$

It controls how much weight each residual contributes when estimating
β.
In two step estimation,
$\psi(e)$ is applied after robustly estimating σ.

2.
Simultaneous Estimation:
Step 1:



Step 1: Define the System of Equations

We need to solve these two equations together:

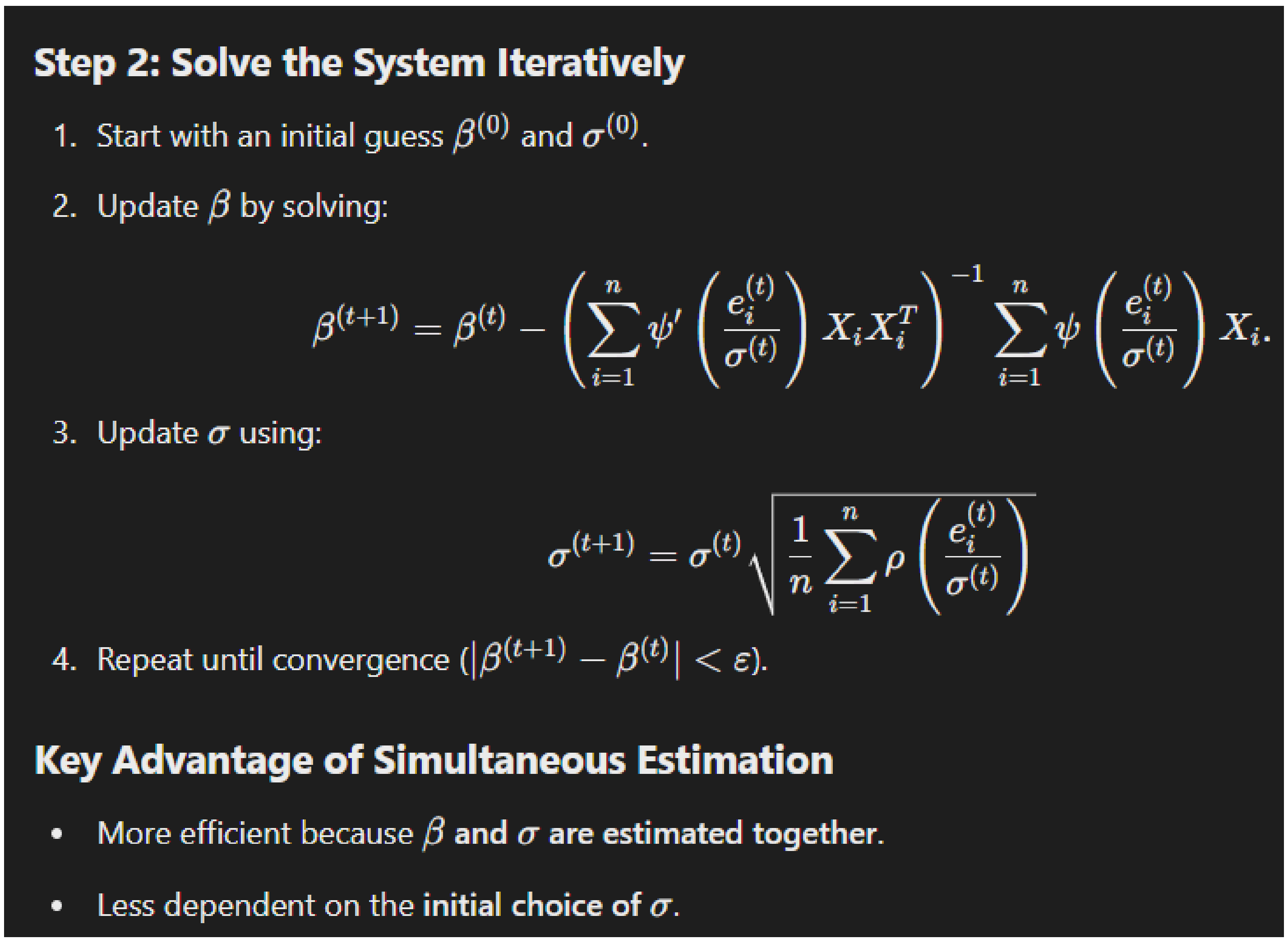
1. **M-estimation equation for β:**

$$\sum_{i=1}^{n} \psi\left(\frac{Y_i - X_i^T \beta}{\sigma}\right) X_i = 0.$$

2. **Scale estimation equation for σ:**

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{Y_i - X_i^T \beta}{\sigma}\right) = K.$$

where K is a constant that ensures robustness.



Step 2: Solve the System Iteratively

1. Start with an initial guess $\beta^{(0)}$ and $\sigma^{(0)}$.
2. Update β by solving:

$$\beta^{(t+1)} = \beta^{(t)} - \left(\sum_{i=1}^{n} \psi' \left(\frac{e_i^{(t)}}{\sigma^{(t)}} \right) X_i X_i^T \right)^{-1} \sum_{i=1}^{n} \psi \left(\frac{e_i^{(t)}}{\sigma^{(t)}} \right) X_i.$$

3. Update σ using:

$$\sigma^{(t+1)} = \sigma^{(t)} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{e_i^{(t)}}{\sigma^{(t)}} \right)}$$

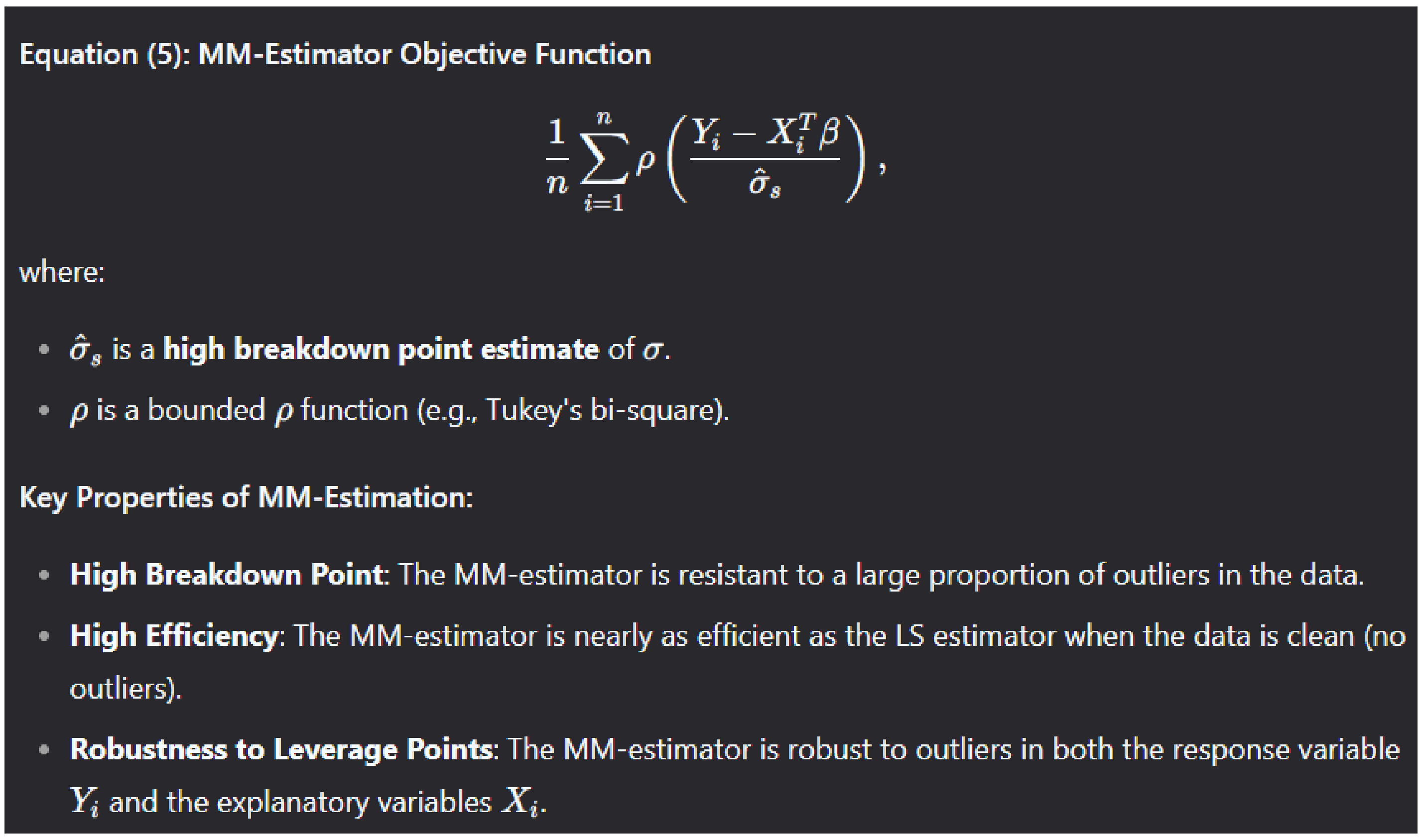
4. Repeat until convergence ($|\beta^{(t+1)} - \beta^{(t)}| < \varepsilon$).

Key Advantage of Simultaneous Estimation

- More efficient because β **and** σ **are estimated together**.
- Less dependent on the **initial choice of** σ.

Limitations of M-Estimation: M-estimators are robust to outliers in the response variable Y_i but can still be influenced by outliers in the explanatory variable X_i (leverage points). The breakdown point of M-estimators is relatively low, thus can be affected by a large proportion of outliers.

The second mentioned mention is MM-Estimation, which combines high breakdown point estimation with high efficiency.



Equation (5): MM-Estimator Objective Function

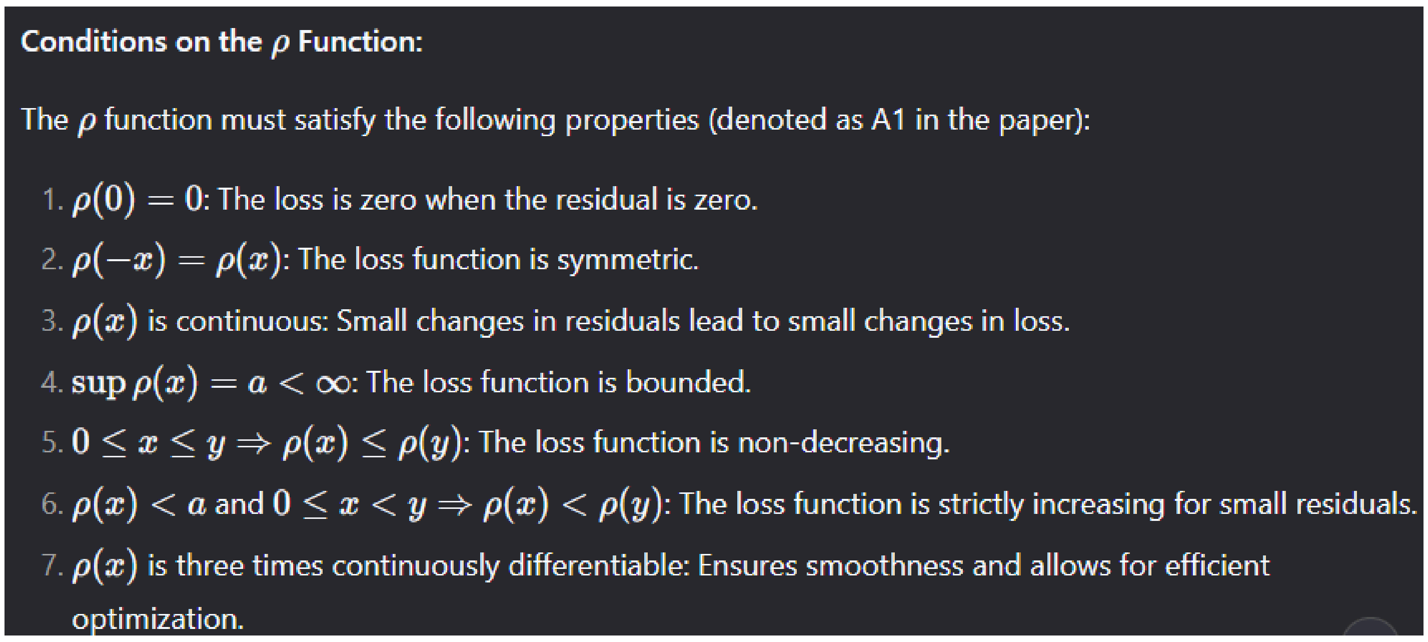
$$\frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{Y_i - X_i^T \beta}{\hat{\sigma}_s} \right),$$

where:

- $\hat{\sigma}_s$ is a **high breakdown point estimate** of σ.
- ρ is a bounded ρ function (e.g., Tukey's bi-square).

Key Properties of MM-Estimation:

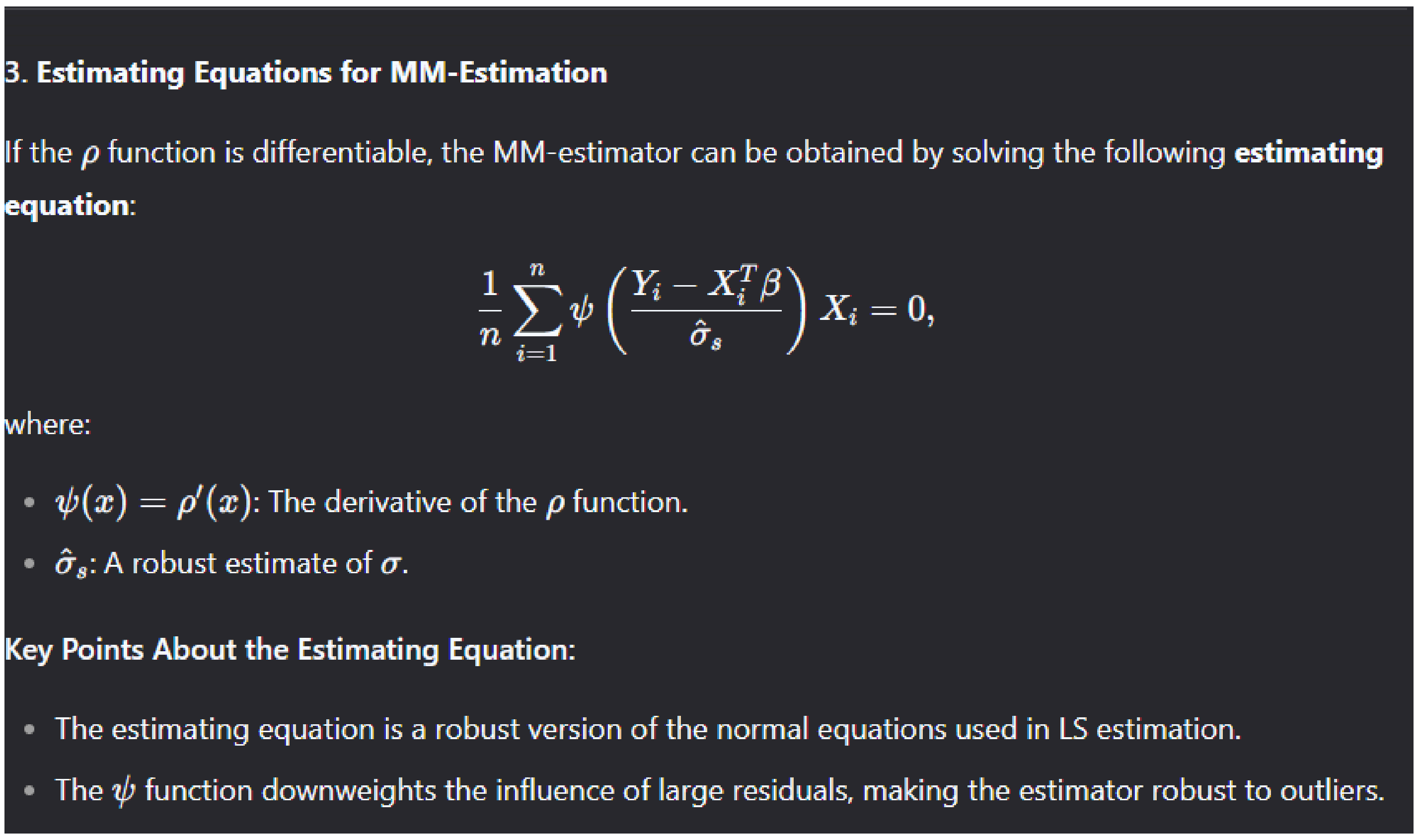
- **High Breakdown Point**: The MM-estimator is resistant to a large proportion of outliers in the data.
- **High Efficiency**: The MM-estimator is nearly as efficient as the LS estimator when the data is clean (no outliers).
- **Robustness to Leverage Points**: The MM-estimator is robust to outliers in both the response variable Y_i and the explanatory variables X_i.



Conditions on the ρ Function:

The ρ function must satisfy the following properties (denoted as A1 in the paper):

1. $\rho(0) = 0$: The loss is zero when the residual is zero.
2. $\rho(-x) = \rho(x)$: The loss function is symmetric.
3. $\rho(x)$ is continuous: Small changes in residuals lead to small changes in loss.
4. $\sup \rho(x) = a < \infty$: The loss function is bounded.
5. $0 \leq x \leq y \Rightarrow \rho(x) \leq \rho(y)$: The loss function is non-decreasing.
6. $\rho(x) < a$ and $0 \leq x < y \Rightarrow \rho(x) < \rho(y)$: The loss function is strictly increasing for small residuals.
7. $\rho(x)$ is three times continuously differentiable: Ensures smoothness and allows for efficient optimization.



3. Estimating Equations for MM-Estimation

If the ρ function is differentiable, the MM-estimator can be obtained by solving the following **estimating equation**:

$$\frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right)X_i = 0,$$

where:

- $\psi(x) = \rho'(x)$: The derivative of the ρ function.
- $\hat{\sigma}_s$: A robust estimate of σ.

Key Points About the Estimating Equation:

- The estimating equation is a robust version of the normal equations used in LS estimation.
- The ψ function downweights the influence of large residuals, making the estimator robust to outliers.

Why is a breakdown point important?
It measures how much data corruption an estimator can withstand.
High-Breakdown-Point-Estimators are required in real world noisy data.

Bound vs Unbound Function?
Bounded functions stay within a horizontal band. Unbounded functions do not.

MM-Estimator Starts with:

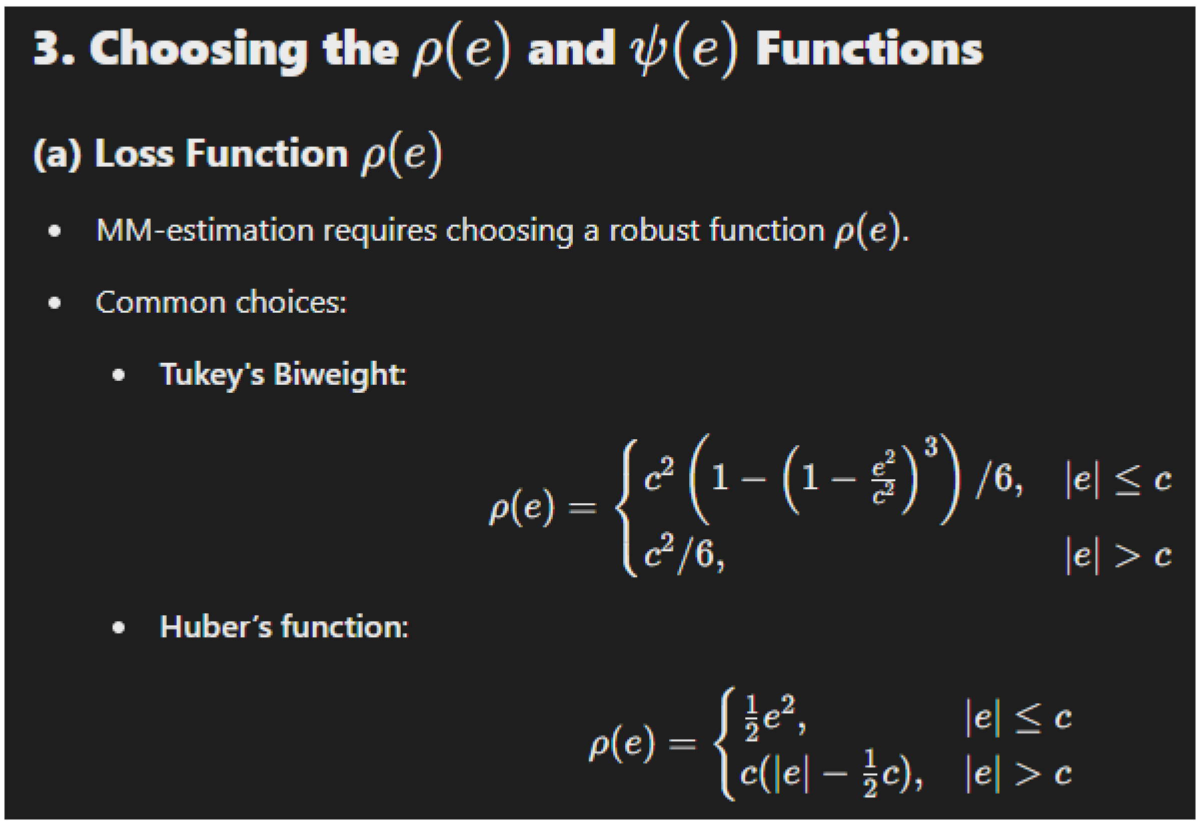
MM-estimators combine the best of both worlds:

-

Step 1: Start with a high breakdown point **S-estimator** for initial robustness.

-

Step 2: Use a M-Estimator with a carefully chose ψ function to improve efficiency.



3. Choosing the $\rho(e)$ and $\psi(e)$ Functions

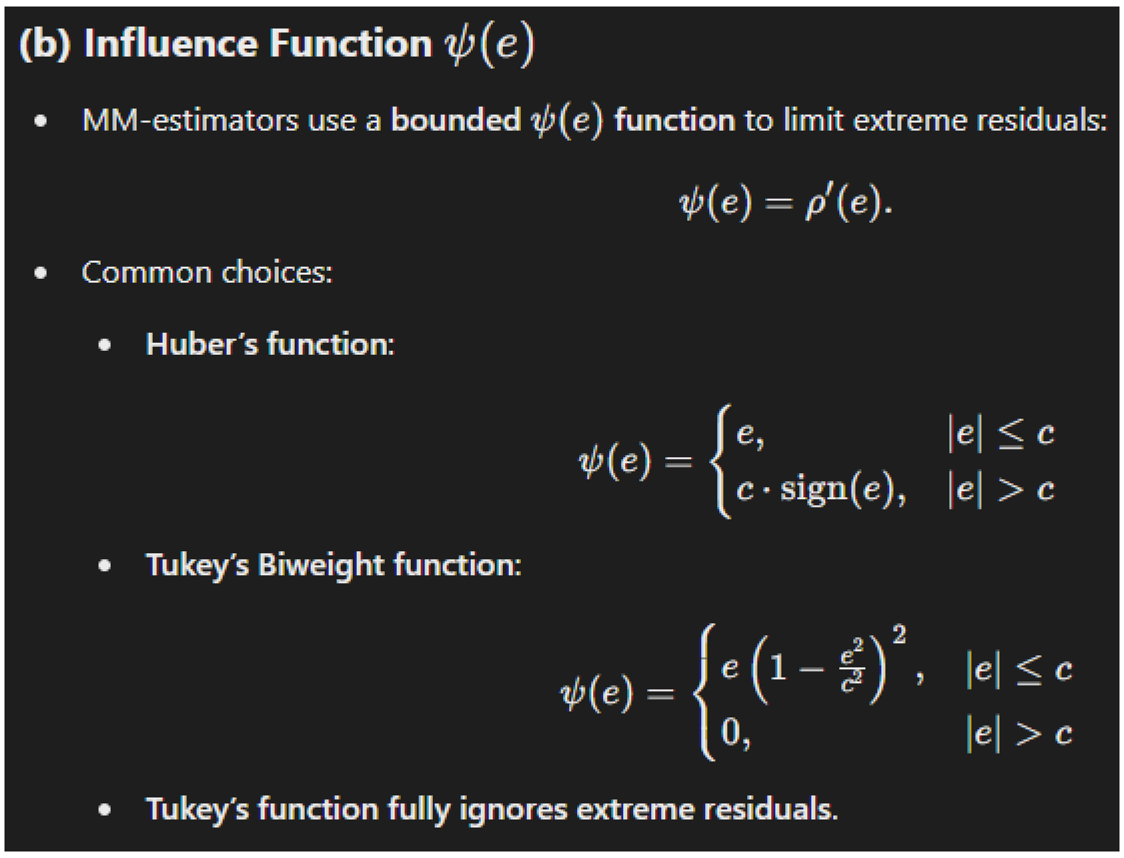
(a) Loss Function $\rho(e)$

- MM-estimation requires choosing a robust function $\rho(e)$.
- Common choices:
  - **Tukey's Biweight:**

$$\rho(e) = \begin{cases} c^2 \left(1 - \left(1 - \frac{e^2}{c^2}\right)^3\right)/6, & |e| \leq c \\ c^2/6, & |e| > c \end{cases}$$

  - **Huber's function:**

$$\rho(e) = \begin{cases} \frac{1}{2}e^2, & |e| \leq c \\ c(|e| - \frac{1}{2}c), & |e| > c \end{cases}$$



(b) Influence Function $\psi(e)$

- MM-estimators use a **bounded $\psi(e)$ function** to limit extreme residuals:

$$\psi(e) = \rho'(e).$$

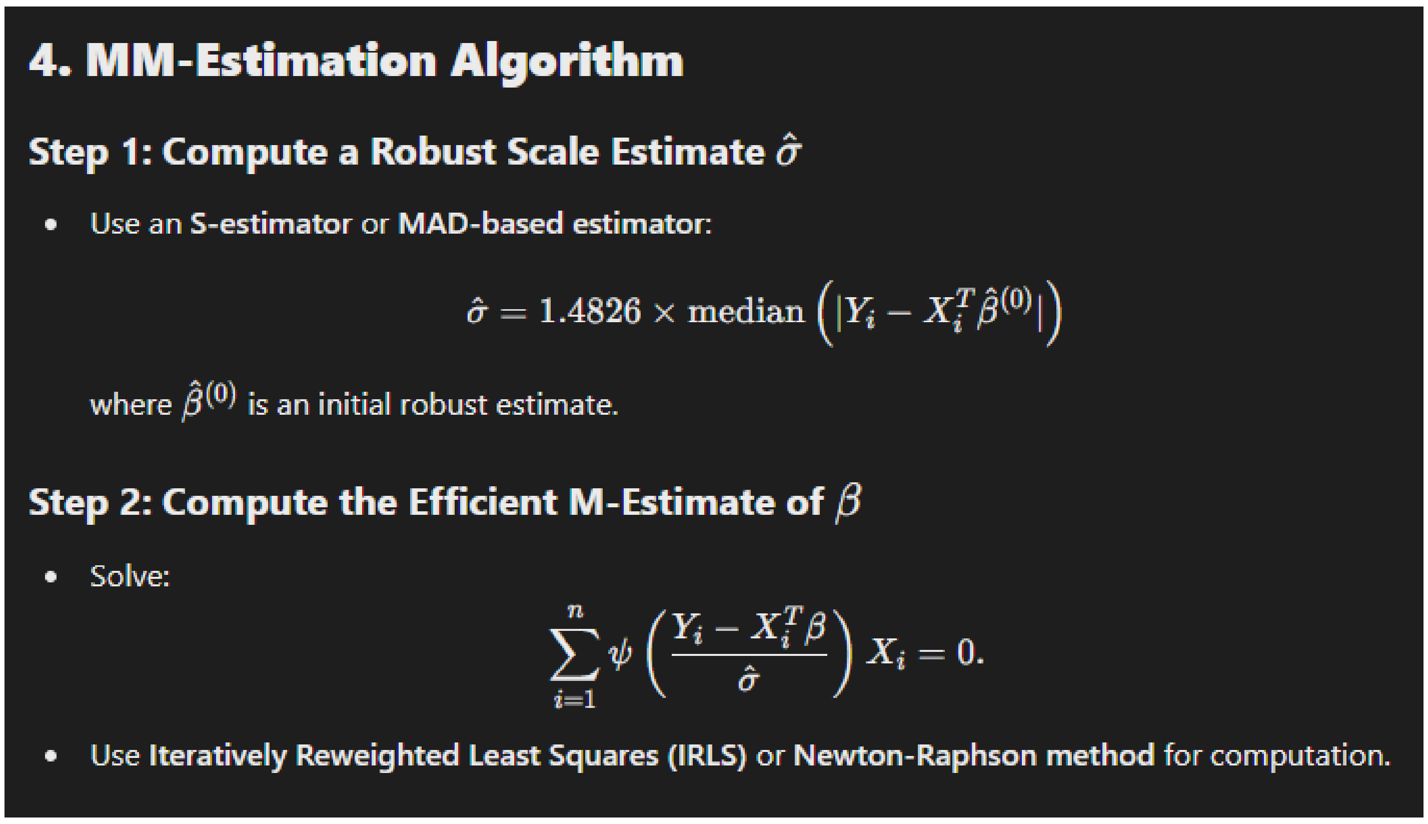
- Common choices:
  - **Huber's function:**

$$\psi(e) = \begin{cases} e, & |e| \leq c \\ c \cdot \text{sign}(e), & |e| > c \end{cases}$$

  - **Tukey's Biweight function:**

$$\psi(e) = \begin{cases} e\left(1 - \frac{e^2}{c^2}\right)^2, & |e| \leq c \\ 0, & |e| > c \end{cases}$$

  - **Tukey's function fully ignores extreme residuals.**



4. MM-Estimation Algorithm

Step 1: Compute a Robust Scale Estimate $\hat{\sigma}$

- Use an **S-estimator** or **MAD-based estimator**:

$$\hat{\sigma} = 1.4826 \times \text{median}\left(|Y_i - X_i^T \hat{\beta}^{(0)}|\right)$$

  where $\hat{\beta}^{(0)}$ is an initial robust estimate.

Step 2: Compute the Efficient M-Estimate of β

- Solve:

$$\sum_{i=1}^{n} \psi\left(\frac{Y_i - X_i^T \beta}{\hat{\sigma}}\right) X_i = 0.$$

- Use **Iteratively Reweighted Least Squares (IRLS)** or **Newton-Raphson method** for computation.

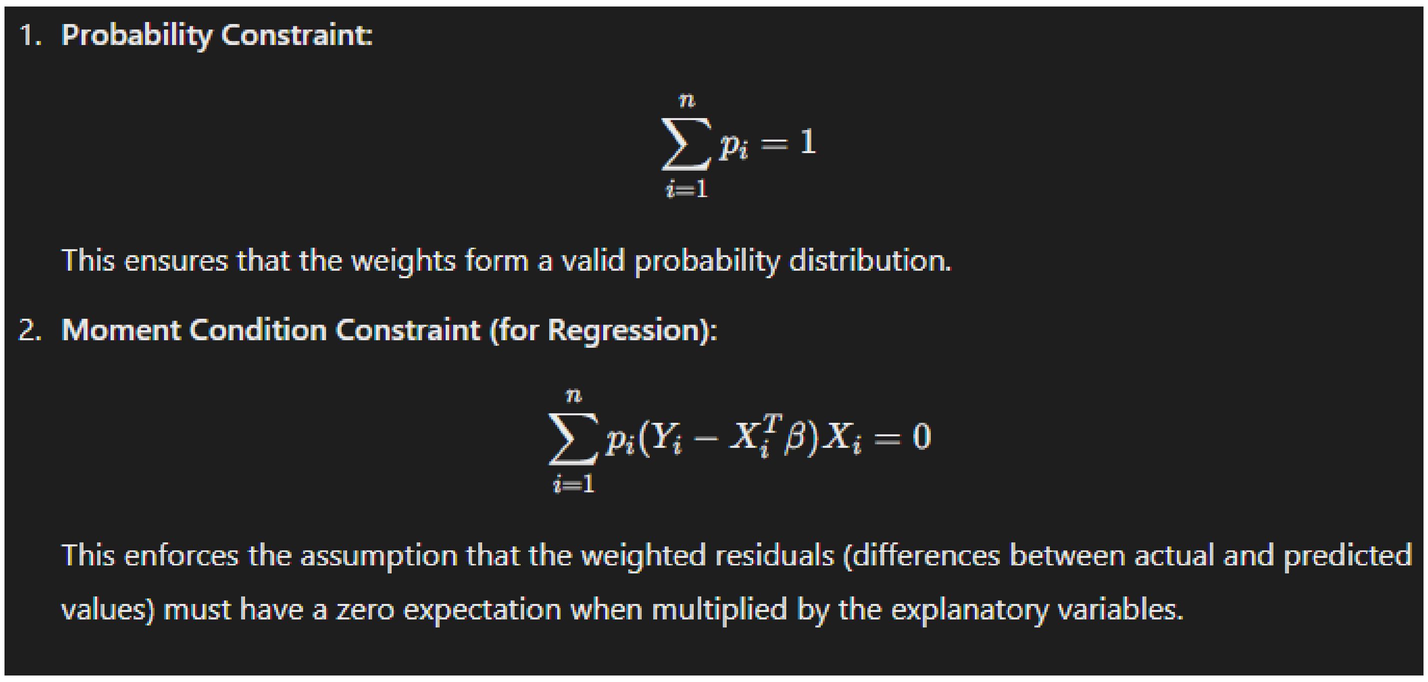
Empirical Likelihood (EL) and Robust EL:

This paper also discusses the Empirical Likelihood Method, which is a non-parametric approach to parameter estimation. The EL method does not require distributional assumptions and assigns probabilities to each observation and maximized the empirical likelihood function.

"parametric" refers to statistical tests that rely on assumptions about the population distribution, usually assuming a normal distribution, while "nonparametric" refers to tests that make no such assumptions about the data distribution, allowing analysis even when the data is not normally distributed

$L(p_1, p_2, p_3, ...p_n) = \prod_{i=1}^{n} p_i$

Constraints:



1. **Probability Constraint:**

$$\sum_{i=1}^{n} p_i = 1$$

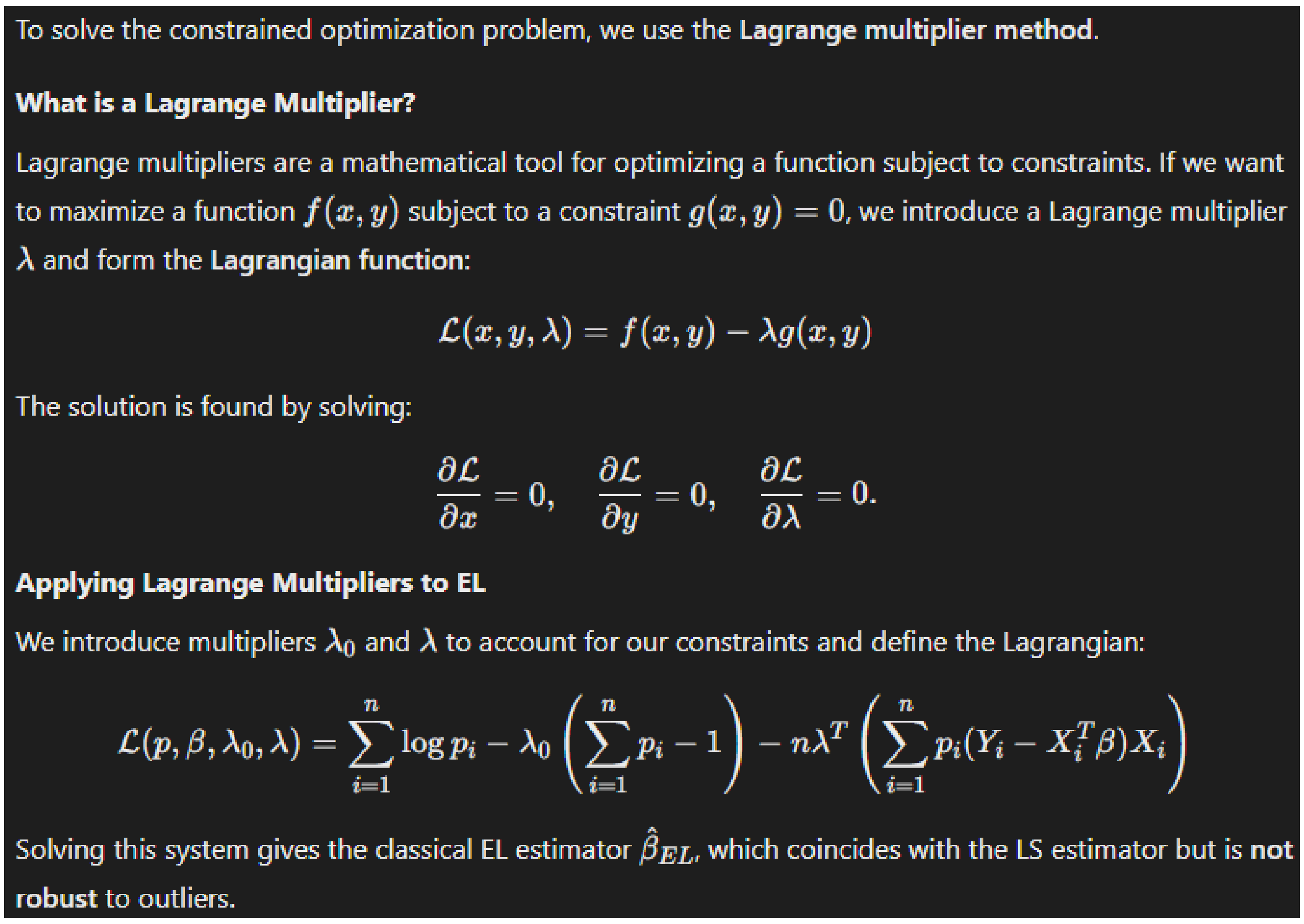
This ensures that the weights form a valid probability distribution.

2. **Moment Condition Constraint (for Regression):**

$$\sum_{i=1}^{n} p_i (Y_i - X_i^T \beta) X_i = 0$$

This enforces the assumption that the weighted residuals (differences between actual and predicted values) must have a zero expectation when multiplied by the explanatory variables.

Solving with Lagrange Multipliers:



To solve the constrained optimization problem, we use the **Lagrange multiplier method**.

What is a Lagrange Multiplier?

Lagrange multipliers are a mathematical tool for optimizing a function subject to constraints. If we want to maximize a function $f(x, y)$ subject to a constraint $g(x, y) = 0$, we introduce a Lagrange multiplier λ and form the **Lagrangian function**:

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

The solution is found by solving:

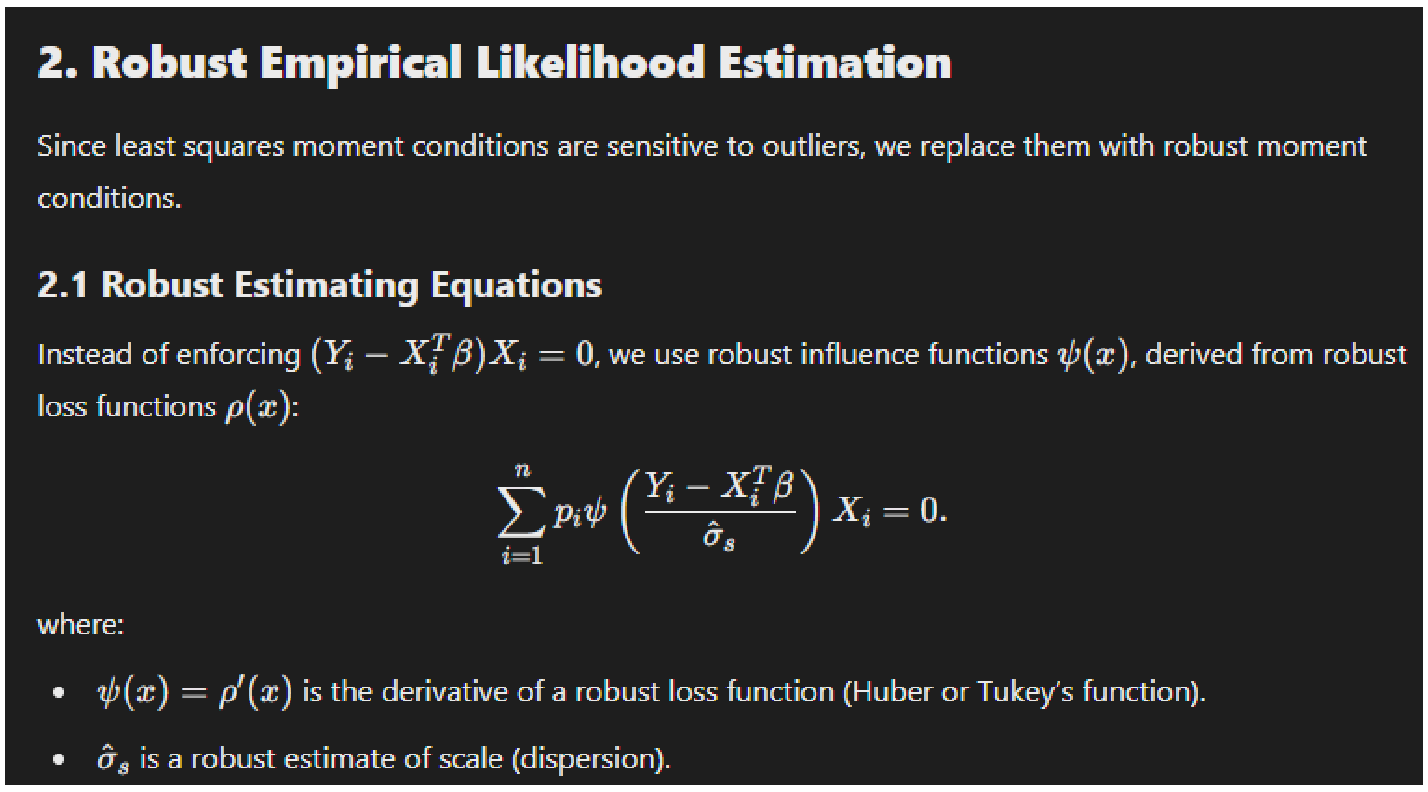
$$\frac{\partial \mathcal{L}}{\partial x} = 0, \quad \frac{\partial \mathcal{L}}{\partial y} = 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0.$$

Applying Lagrange Multipliers to EL

We introduce multipliers λ_0 and λ to account for our constraints and define the Lagrangian:

$$\mathcal{L}(p, \beta, \lambda_0, \lambda) = \sum_{i=1}^{n} \log p_i - \lambda_0 \left(\sum_{i=1}^{n} p_i - 1 \right) - n\lambda^T \left(\sum_{i=1}^{n} p_i (Y_i - X_i^T \beta) X_i \right)$$

Solving this system gives the classical EL estimator $\hat{\beta}_{EL}$, which coincides with the LS estimator but is **not robust** to outliers.



2. Robust Empirical Likelihood Estimation

Since least squares moment conditions are sensitive to outliers, we replace them with robust moment conditions.

2.1 Robust Estimating Equations

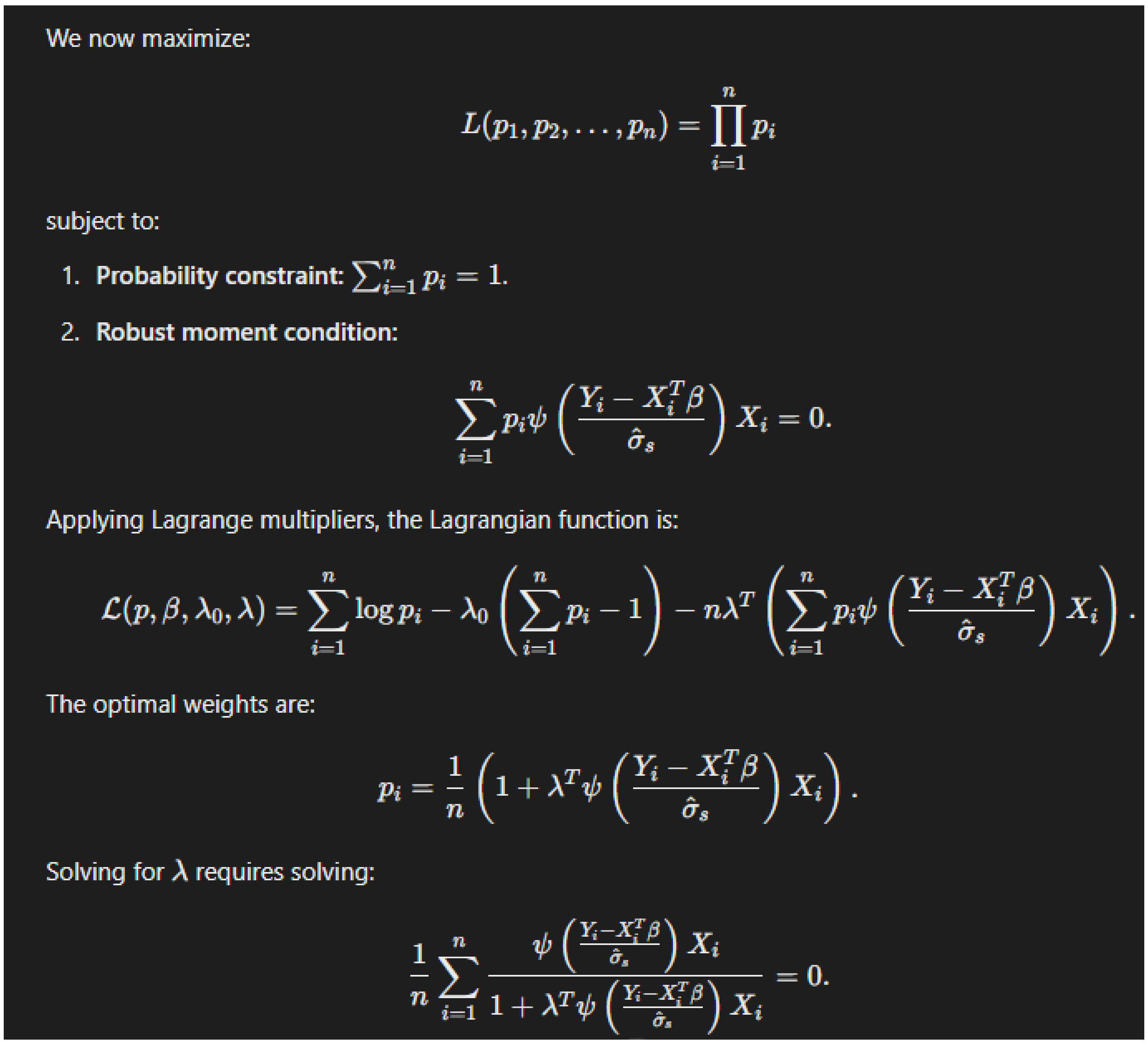
Instead of enforcing $(Y_i - X_i^T\beta)X_i = 0$, we use robust influence functions $\psi(x)$, derived from robust loss functions $\rho(x)$:

$$\sum_{i=1}^{n} p_i \psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i = 0.$$

where:

- $\psi(x) = \rho'(x)$ is the derivative of a robust loss function (Huber or Tukey's function).
- $\hat{\sigma}_s$ is a robust estimate of scale (dispersion).

Options for $\rho(x)\ \ and\ \ \psi(x)$ - same as M-Estimator Huber and Tukey Functions.



We now maximize:

$$L(p_1, p_2, \ldots, p_n) = \prod_{i=1}^{n} p_i$$

subject to:

1. **Probability constraint:** $\sum_{i=1}^{n} p_i = 1$.
2. **Robust moment condition:**

$$\sum_{i=1}^{n} p_i \psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i = 0.$$

Applying Lagrange multipliers, the Lagrangian function is:

$$\mathcal{L}(p, \beta, \lambda_0, \lambda) = \sum_{i=1}^{n} \log p_i - \lambda_0 \left(\sum_{i=1}^{n} p_i - 1\right) - n\lambda^T \left(\sum_{i=1}^{n} p_i \psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i\right).$$

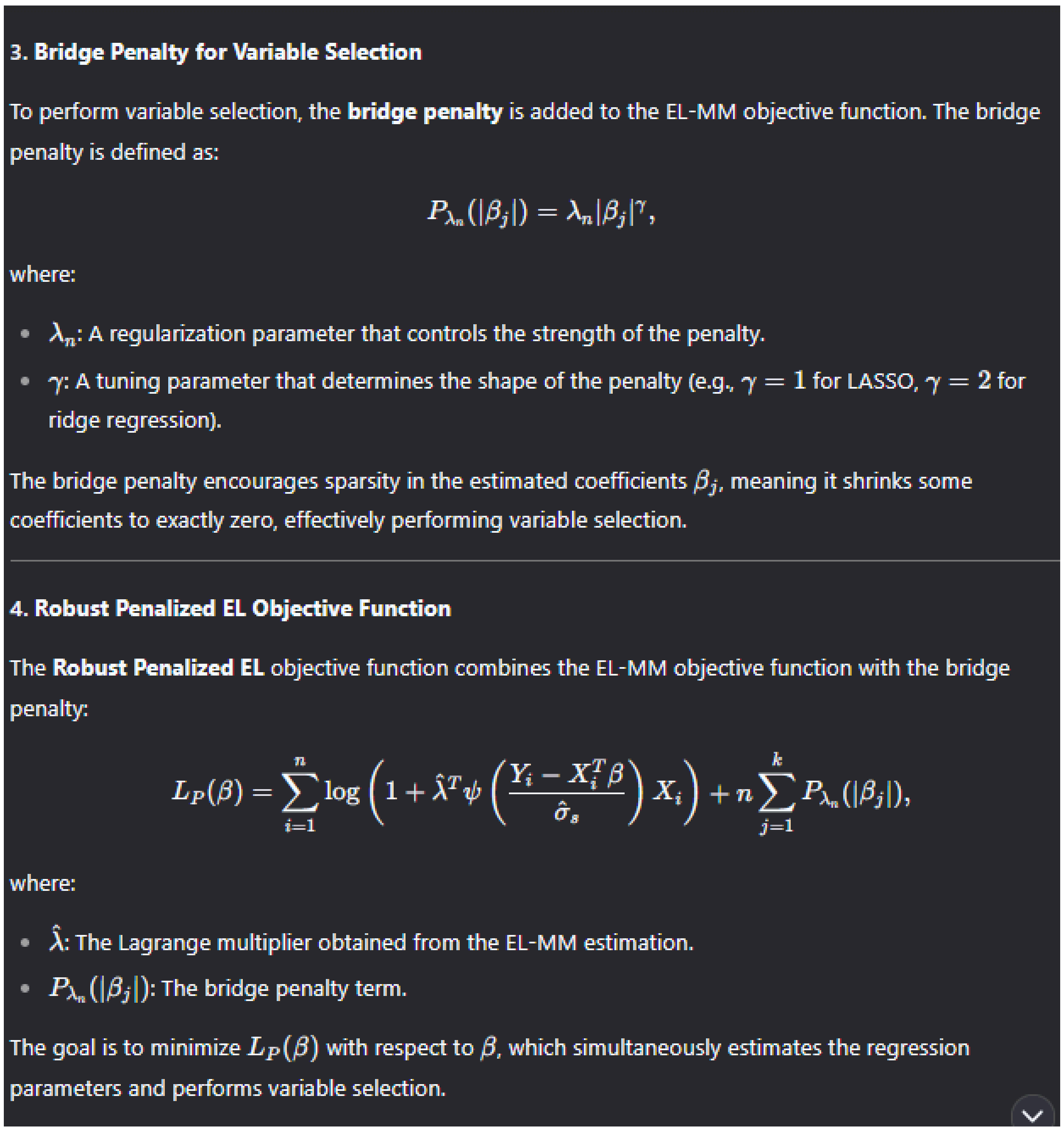
The optimal weights are:

$$p_i = \frac{1}{n}\left(1 + \lambda^T \psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i\right).$$

Solving for λ requires solving:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i}{1 + \lambda^T \psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i} = 0.$$

Robust penalized EL estimation for regression: EL-MM bridge regression estimation:



3. Bridge Penalty for Variable Selection

To perform variable selection, the **bridge penalty** is added to the EL-MM objective function. The bridge penalty is defined as:

$$P_{\lambda_n}(|\beta_j|) = \lambda_n |\beta_j|^\gamma,$$

where:

- λ_n: A regularization parameter that controls the strength of the penalty.
- γ: A tuning parameter that determines the shape of the penalty (e.g., $\gamma = 1$ for LASSO, $\gamma = 2$ for ridge regression).

The bridge penalty encourages sparsity in the estimated coefficients β_j, meaning it shrinks some coefficients to exactly zero, effectively performing variable selection.

4. Robust Penalized EL Objective Function

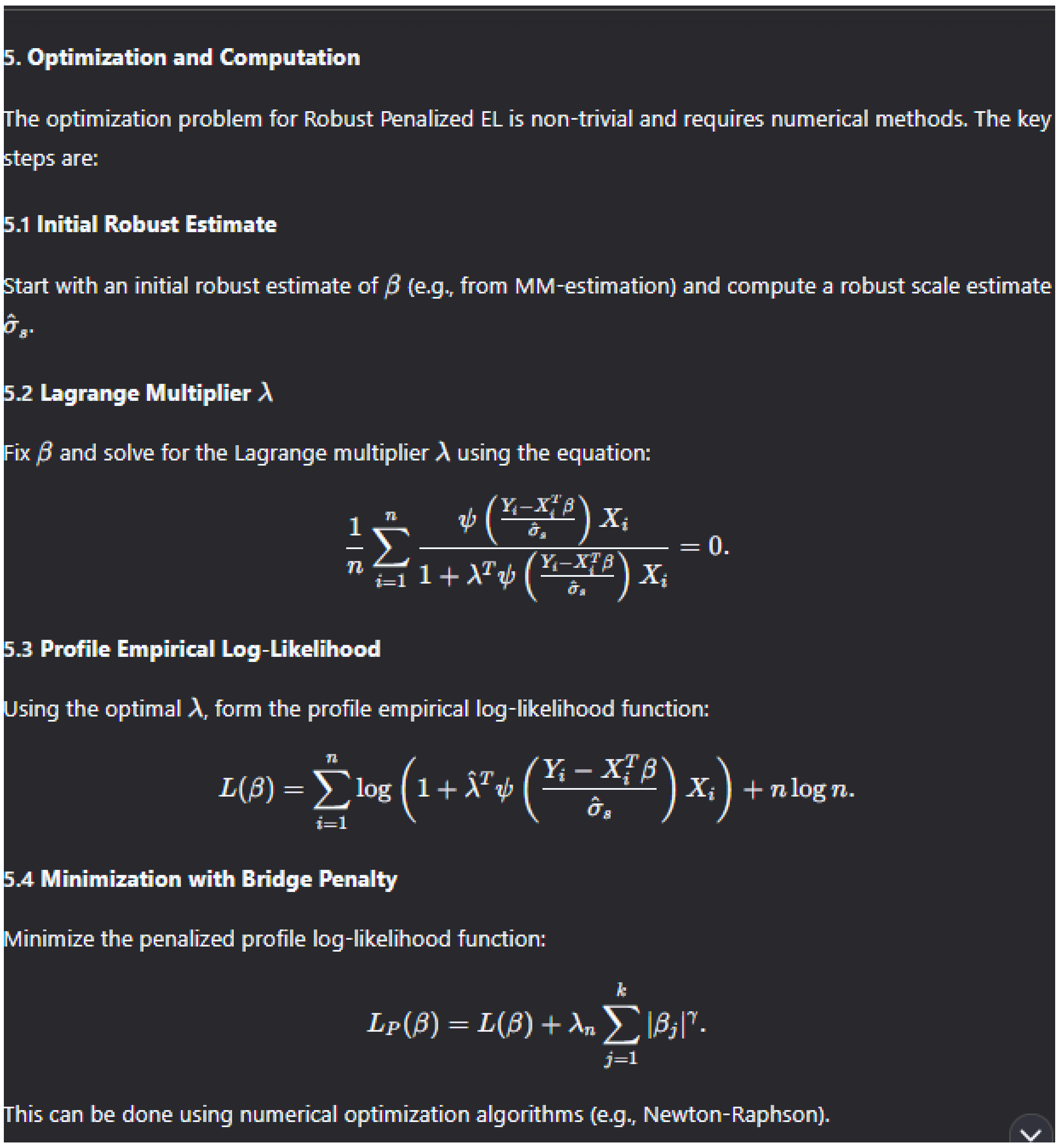
The **Robust Penalized EL** objective function combines the EL-MM objective function with the bridge penalty:

$$L_P(\beta) = \sum_{i=1}^{n} \log\left(1 + \hat{\lambda}^T \psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i\right) + n \sum_{j=1}^{k} P_{\lambda_n}(|\beta_j|),$$

where:

- $\hat{\lambda}$: The Lagrange multiplier obtained from the EL-MM estimation.
- $P_{\lambda_n}(|\beta_j|)$: The bridge penalty term.

The goal is to minimize $L_P(\beta)$ with respect to β, which simultaneously estimates the regression parameters and performs variable selection.



5. Optimization and Computation

The optimization problem for Robust Penalized EL is non-trivial and requires numerical methods. The key steps are:

5.1 Initial Robust Estimate

Start with an initial robust estimate of β (e.g., from MM-estimation) and compute a robust scale estimate $\hat{\sigma}_s$.

5.2 Lagrange Multiplier λ

Fix β and solve for the Lagrange multiplier λ using the equation:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i}{1 + \lambda^T \psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i} = 0.$$

5.3 Profile Empirical Log-Likelihood

Using the optimal λ, form the profile empirical log-likelihood function:

$$L(\beta) = \sum_{i=1}^{n} \log\left(1 + \hat{\lambda}^T \psi\left(\frac{Y_i - X_i^T\beta}{\hat{\sigma}_s}\right) X_i\right) + n \log n.$$

5.4 Minimization with Bridge Penalty

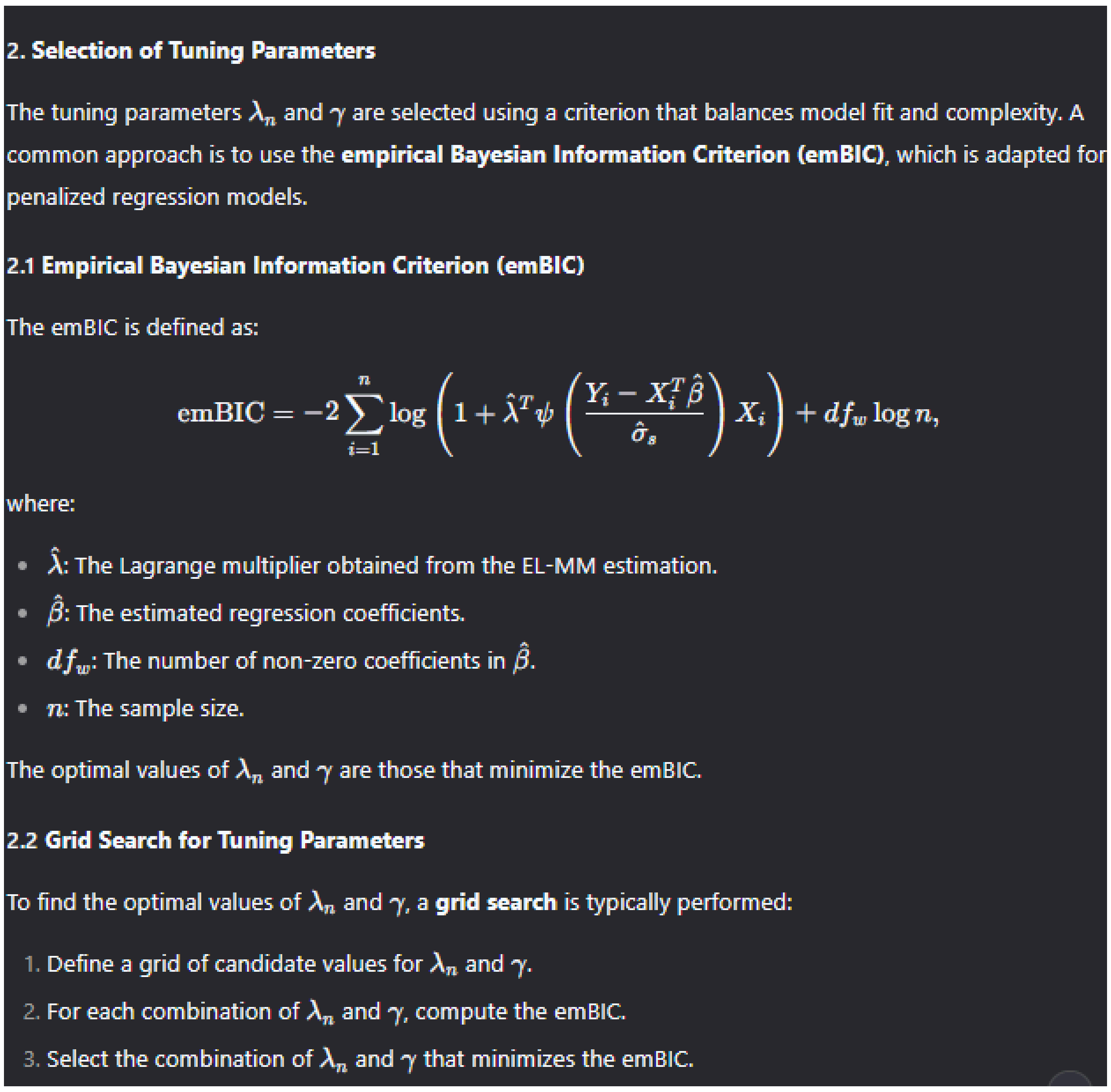
Minimize the penalized profile log-likelihood function:

$$L_P(\beta) = L(\beta) + \lambda_n \sum_{j=1}^{k} |\beta_j|^\gamma.$$

This can be done using numerical optimization algorithms (e.g., Newton-Raphson).

Selection of Tuning Parameters and Computation Steps:

In the **Robust Penalized Empirical Likelihood (EL) Estimation** method, particularly in the **EL-MM Bridge Regression Estimation**, the selection of tuning parameters and the computation steps are crucial for achieving both robust parameter estimation and effective variable selection.



2. Selection of Tuning Parameters

The tuning parameters λ_n and γ are selected using a criterion that balances model fit and complexity. A common approach is to use the **empirical Bayesian Information Criterion (emBIC)**, which is adapted for penalized regression models.

2.1 Empirical Bayesian Information Criterion (emBIC)

The emBIC is defined as:

$$\text{emBIC} = -2\sum_{i=1}^{n} \log\left(1 + \hat{\lambda}^T \psi\left(\frac{Y_i - X_i^T\hat{\beta}}{\hat{\sigma}_s}\right) X_i\right) + df_w \log n,$$

where:

- $\hat{\lambda}$: The Lagrange multiplier obtained from the EL-MM estimation.
- $\hat{\beta}$: The estimated regression coefficients.
- df_w: The number of non-zero coefficients in $\hat{\beta}$.
- n: The sample size.

The optimal values of λ_n and γ are those that minimize the emBIC.

2.2 Grid Search for Tuning Parameters

To find the optimal values of λ_n and γ, a **grid search** is typically performed:

1. Define a grid of candidate values for λ_n and γ.
2. For each combination of λ_n and γ, compute the emBIC.
3. Select the combination of λ_n and γ that minimizes the emBIC.

Summary of Computation Steps [As discussed in the paper]:

1. Initial Robust Estimate: Obtain an initial robust estimate of β and compute $\hat{\sigma}_s$.
2. Lagrange Multiplier
λ: Solve of λ using robust moments conditions.
3. Profile Empirical Log-Likelihood: Form the profile empirical log-likehood function $L(\beta)$.
4. Penalized Profile Log-Likelihood: Minimize
$L_p(\beta) = L(\beta) + \lambda_n \sum_{j=1}^{n} |\beta_j|^\gamma$ using numerical optimization.
5. Tuning parameter selection: Select optimal
λ_n and γ using emBIC and a grid search.

END