Alternative Statistical Model: Weighted Least Square and Generalized Least Square

Xingye Qiao
Dr. Jim Crooks

SAMSI
SAMSI/CRSC Undergraduate Workshop at NCSU

May 22, 2007

Outline

1. **Introduction**
   - Recall of Ordinary Least-Square Regression
   - Current Model
2. **Improved Model**
   - Weighted Least-Square
   - Generalized Least-Square

Outline

1. **Introduction**
   - Recall of Ordinary Least-Square Regression
   - Current Model
2. **Improved Model**
   - Weighted Least-Square
   - Generalized Least-Square

Least Square Regression

- Linear
  - “Linear" is for the parameter(s)
  - e.g. $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

Least Square Regression

- Linear
  - “Linear" is for the parameter(s)
  - e.g. $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
- Non-linear
  - “Non-linear" is for the parameter(s)
  - e.g. $y_i = exp(-\beta_1 x_i) + \alpha cos(\beta_2 x_i) + \varepsilon_i$

Least Square Regression

- Linear
  - “Linear" is for the parameter(s)
  - e.g. $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
- Non-linear
  - “Non-linear" is for the parameter(s)
  - e.g. $y_i = exp(-\beta_1 x_i) + \alpha cos(\beta_2 x_i) + \varepsilon_i$
- Summary
  - $y_i = \eta(x_i; \beta) + \varepsilon_i$
  - $\eta(x; \beta)$ is deterministic function of x, with parameter β
  - Goal: to estimate parameter β

OLS estimation

- Find β to minimize

$$\sum_{i=1}^{m} (y_i - \eta(x_i;\beta))^2,$$

  to give $\hat{\beta}_{OLS}$
- Standard Statistical Assumption:
  - Mean of ε_i is 0 for all i
  - Variance of ε_i is constant for all i, equal to σ^2
  - ε_i,ε_j are independent of each other for all $i \neq j$

OLS Estimation (Cont.)

Property of OLS Estimation

- $\hat{\beta}_{OLS}$ converges to β as n increases

OLS Estimation (Cont.)

Property of OLS Estimation

- $\hat{\beta}_{OLS}$ converges to β as n increases
- Makes efficient use of the data, i.e. has small standard error

OLS Estimation (Cont.)

Property of OLS Estimation

- $\hat{\beta}_{OLS}$ converges to β as n increases
- Makes efficient use of the data, i.e. has small standard error
- These properties hold only when the model is a right model. To be more specific, when the standard statistical assumption holds.

Inverse Problem

- Spring Model:

$$\frac{d^2y(t)}{dt^2} + C\frac{dy(t)}{dt} + Ky(t) = 0$$

Inverse Problem

- Spring Model:

$$\frac{d^2y(t)}{dt^2} + C\frac{dy(t)}{dt} + Ky(t) = 0$$

- For each given C and K, the differential equation has a unique solution given initial value, called $y(t; C, K)$

Inverse Problem

- Spring Model:

$$\frac{d^2y(t)}{dt^2} + C\frac{dy(t)}{dt} + Ky(t) = 0$$

- For each given C and K, the differential equation has a unique solution given initial value, called $y(t; C, K)$
- Target: Estimate C and K based on the observed y_i

Inverse Problem

- Spring Model:

$$\frac{d^2y(t)}{dt^2} + C\frac{dy(t)}{dt} + Ky(t) = 0$$

- For each given C and K, the differential equation has a unique solution given initial value, called $y(t; C, K)$
- Target: Estimate C and K based on the observed y_i
- Minimize the cost function

$$L(C,K) = \sum_{i=1}^{m} (y_i - y(t_i; C, K))^2$$

Underlying Statistical Models

- The above model can be viewed as a regression model

$$y_i = y(t_i; C, K) + \varepsilon_i$$

- Here ε_i are iid(independent identically distributed) from $N(0, \sigma^2)$. That is we suppose the statistical assumptions hold.

Underlying Statistical Models

- The above model can be viewed as a regression model

$$y_i = y(t_i; C, K) + \varepsilon_i$$

- Here ε_i are iid(independent identically distributed) from $N(0, \sigma^2)$. That is we suppose the statistical assumptions hold.
- But is this model a right model?

Violation of Statistical Assumptions

1. Is variance of ε_i constant across time range?

Violation of Statistical Assumptions

1. Is variance of ε_i constant across time range?
2. Are error independent?

Violation of Statistical Assumptions

1. Is variance of ε_i constant across time range?
2. Are error independent?
3. Are error from $N(0, \sigma^2)$?

Violation of Statistical Assumptions

1. Is variance of ε_i constant across time range?
2. Are error independent?
3. Are error from $N(0, \sigma^2)$?

Implication:

Violation of Statistical Assumptions

1. Is variance of ϵ_i constant across time range?
2. Are error independent?
3. Are error from $N(0, \sigma^2)$?

Implication:

- Standard statistical assumptions don’t hold.

Violation of Statistical Assumptions

1. Is variance of ε_i constant across time range?
2. Are error independent?
3. Are error from $N(0, \sigma^2)$?

Implication:

- Standard statistical assumptions don't hold.
- $[\hat{C}, \hat{K}]$ are no longer good estimators for $[C, K]$.

Violation of Statistical Assumptions

1. Is variance of ε_i constant across time range?
2. Are error independent?
3. Are error from $N(0, \sigma^2)$?

Implication:

- Standard statistical assumptions don't hold.
- $[\hat{C}, \hat{K}]$ are no longer good estimators for $[C, K]$.
- We should find a way to remedy this problem.

Outline

1. Introduction
   - Recall of Ordinary Least-Square Regression
   - Current Model
2. **Improved Model**
   - Weighted Least-Square
   - Generalized Least-Square

Assumption

- Instead of constant variance assumption, we deal with nonconstant variance here.

Assumption

- Instead of constant variance assumption, we deal with nonconstant variance here.
- Assume $Var(\varepsilon_i) = \frac{\sigma^2}{w_i}, i = 1, \ldots, m$, for known w_i

Assumption

- Instead of constant variance assumption, we deal with nonconstant variance here.
- Assume $Var(\varepsilon_i) = \frac{\sigma^2}{w_i}, i = 1, \ldots, m$, for known w_i
- What does it mean for (y_i, t_i) if w_i is large?

Assumption

- Instead of constant variance assumption, we deal with nonconstant variance here.
- Assume $Var(\varepsilon_i) = \frac{\sigma^2}{w_i}, i = 1, \ldots, m$, for known w_i
- What does it mean for (y_i, t_i) if w_i is large?
    - $\Leftrightarrow$ This observation is of high quality.

Assumption

- Instead of constant variance assumption, we deal with nonconstant variance here.
- Assume $Var(\varepsilon_i) = \frac{\sigma^2}{w_i}, i = 1, \ldots, m$, for known w_i
- What does it mean for (y_i, t_i) if w_i is large?
    - $\Leftrightarrow$ This observation is of high quality.
    - $\Leftrightarrow$ This observation is of importance

Solve Weighted Least Square in Linear Case

Consider linear context,

$$y_i = x_i^T \beta + \varepsilon_i.$$

Solve Weighted Least Square in Linear Case

Consider linear context,

$$y_i = x_i^T \beta + \varepsilon_i.$$

Denote

$$y_i^* = \sqrt{w_i} y_i, x_i^* = \sqrt{w_i} x_i,$$

Solve Weighted Least Square in Linear Case

Consider linear context,

$$y_i = x_i^T \beta + \varepsilon_i.$$

Denote

$$y_i^* = \sqrt{w_i} y_i, x_i^* = \sqrt{w_i} x_i,$$

Then

$$y_i^* = {x_i^*}^T \beta + \sqrt{w_i} \varepsilon_i,$$

where $Var(\sqrt{w_i}\varepsilon_i) = w_i Var(\varepsilon_i) = \sigma^2$

Solve Weighted Least Square in Linear Case (Cont.)

Then minimizing the weighted (least) sum squares of error

$$S = \sum_{i=1}^{n} w_i (y_i - x_i^T \beta)^2,$$

is the same as minimizing the ordinary (least) sum squares of error

$$S = \sum_{i=1}^{n} (y_i^* - {x_i^*}^T \beta)^2.$$

Solve Weighted Least Square in Linear Case (Cont.)

Then minimizing the weighted (least) sum squares of error

$$S = \sum_{i=1}^{n} w_i (y_i - x_i^T \beta)^2,$$

is the same as minimizing the ordinary (least) sum squares of error

$$S = \sum_{i=1}^{n} (y_i^* - x_i^{*T} \beta)^2.$$

In matrix notation, the weighted least squares estimator of β is

$$\hat{\beta} = (X^{*T} X^*)^{-1} X^{*T} Y^* = (X^T W X)^{-1} X^T W Y, W = diag\{w_1, \ldots, w_n\}.$$

Estimation

- Instead of minimizing $\sum_{i=1}^{m}(y_i - y(t_i; C, K))^2$ in OLS, here minimize

  $$\tilde{L}(C, K) = \sum_{i=1}^{m} w_i (y_i - y(t_i; C, K))^2,$$

  to give $\hat{C}$ and $\hat{K}$

Value of w_i

- In practice, we don't know w_i. Several ways to estimate w_i:

Value of w_i

- In practice, we don't know w_i. Several ways to estimate w_i:
    1. Estimate $Var(\varepsilon_i)$ as $\hat{\sigma}_i^2$ from repeated measurment at time t_i:

$$w_i = \frac{\sigma^2}{\hat{\sigma}_i^2}.$$

Value of w_i

- In practice, we don't know w_i. Several ways to estimate w_i:
  1. Estimate $Var(\varepsilon_i)$ as $\hat{\sigma}_i^2$ from repeated measurment at time t_i:
  $$w_i = \frac{\sigma^2}{\hat{\sigma}_i^2}.$$
  2. If error is larger for larger $|y_i|$, simply let $w_i = \frac{1}{y_i^2}$

Value of w_i

- In practice, we don't know w_i. Several ways to estimate w_i:
    1. Estimate $Var(\varepsilon_i)$ as $\hat{\sigma}_i^2$ from repeated measurment at time t_i:
    $$w_i = \frac{\sigma^2}{\hat{\sigma}_i^2}.$$
    2. If error is larger for larger $|y_i|$, simply let $w_i = \frac{1}{y_i^2}$
    3. Or, alternatively, assume that $w_i = \frac{1}{y^2(t_i;C,K)}$

Assumption

More general, now deal with correlated observations and nonconstant variance (weighted least square only deals with nonconstant variance):

Assumption

More general, now deal with correlated observations and nonconstant variance (weighted least square only deals with nonconstant variance):

- Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)^T$, and assume

$$Cov(\varepsilon) = \sigma^2 V, \text{ for known matrix V.}$$

Assumption

More general, now deal with correlated observations and nonconstant variance (weighted least square only deals with nonconstant variance):

- Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)^T$, and assume

$$Cov(\varepsilon) = \sigma^2 V, \text{ for known matrix V.}$$

- Let $W = V^{-1}$. Remember that if V is diagonal matrix, then this is the case in weighted least square, and $W = diag\{w_1, w_2, \ldots, w_m\}$.

Estimation

- The generalized Least Square estimator minimizes

  $$\tilde{L}(C,K) = \{y - y(t;C,K)\}^T W \{y - y(t;C,K)\},$$

  to given $\hat{C}$ and $\hat{K}$.

Estimation

- The generalized Least Square estimator minimizes

$$\tilde{L}(C,K) = \{y - y(t;C,K)\}^T W \{y - y(t;C,K)\},$$

  to given $\hat{C}$ and $\hat{K}$.
- If the proposed covariance model holds, then the estimators have good properties.

Anything More?

- Does this improved model work better?

Anything More?

- Does this improved model work better?
- If not, what might be the main problem?

Anything More?

- Does this improved model work better?
- If not, what might be the main problem?
- Let Jim take over.