Alternative Statistical Models

Surajit Ray
Outline

- Recap of (ordinary) least-squares
- Violations of statistical assumptions
- Weighted least-squares
- Generalized least-squares
- Concluding comments
Recap of (ordinary) least-squares

Model for data $(t_j, y_j), j = 1, \ldots, n$:

$$y_j = y(t_j; q) + \varepsilon_j$$

- $y(t_j; q)$ is deterministic model, with parameters $q$.
- $\varepsilon_j$ are random errors.

Goal of the inverse problem: estimate $q$.

Standard statistical assumptions for the model:

1. $y(t_j; q)$ is correct model $\Rightarrow$ mean of $\varepsilon_j$ is 0 for all $j$.
2. Variance of $\varepsilon_j$ is constant for all $j$, equal to $\sigma^2$.
3. Error terms $\varepsilon_j, \varepsilon_k$ are independent for $j \neq k$. 
OLS estimation

- Minimize

\[ J(q) = \sum_{i=1}^{n} |y_i - y(t_i; q)|^2 \]  (1)

in \( q \), to give \( \hat{q}_{ols} \).

- Estimate \( \sigma^2 \) by

\[ \hat{\sigma}_{ols}^2 = \frac{1}{n - p} J(\hat{q}_{ols}) \]

where \( p = \text{dim}(q) \).
OLS Estimation

- converges to $q$ as $n$ increases
- makes **efficient** use of the data, i.e., has small standard error
- approximate s.e.($\hat{q}_{ols,k}$) = square root of $(k,k)$ element in

$$\text{Cov}(\hat{q}) = \hat{\sigma}_{ols}^2 [X^TX]^{-1}$$

where $X_{r,c} = \left( \frac{\partial y(t_r; q)}{\partial q_c} \right)$ evaluated at $\hat{q}_{ols}$.

For example, if $q = (C, K)^T$, then

$$X = \begin{bmatrix}
\frac{\partial y(t_1; q)}{\partial C} & \frac{\partial y(t_1; q)}{\partial K} \\
\frac{\partial y(t_2; q)}{\partial C} & \frac{\partial y(t_2; q)}{\partial K} \\
\vdots & \vdots \\
\frac{\partial y(t_n; q)}{\partial C} & \frac{\partial y(t_n; q)}{\partial K}
\end{bmatrix}.$$
Violations of statistical assumptions

Compute residuals, \( r_j = y_j - y(t_j; \widehat{q}_{ols}) \), plot against \( t_j \):

![Residual plot for the (damped) spring-mass-dashpot model, fitted using OLS.](image)

Figure 1: Residual plot for the (damped) spring-mass-dashpot model, fitted using OLS.
Analysis of Residual plot

1. Do we have the correct deterministic model?
2. Is variance of $\varepsilon_j$ constant across time range? No!
3. Are errors independent? No!

Implications: $\hat{q}_{ols}$ is no longer a good estimator for $q$.

Assuming that answer to #1 is “Yes”, how can we change our statistical model assumptions to better model reality?
- Transform the data (e.g., log transform)?
- Explicitly model nonconstant variance, and correlations between measurements.
- Incorporate this into the estimation method.
Weighted least-squares

(a) Deal with nonconstant variance: Assume

\[ \text{Var}(\varepsilon_j) = \frac{\sigma^2}{w_j}, \quad j = 1, \ldots, n \]

for known \( w_j \).

\( w_j \) large \( \Leftrightarrow \) observation \( (t_j, y_j) \) is of high quality.

Instead of OLS, minimize

\[ \tilde{J}(q) = \sum_{i=1}^{n} w_i |y_i - y(t_i; q)|^2 \]

in \( q \).
Weighted Least Squares

In practice, don’t know \( w_j \):

- Estimate \( \text{Var}(\varepsilon_j) \) from repeated measurements at time \( t_j \):
  \[
  w_j = \frac{\sigma^2}{\hat{\sigma}_j^2}.
  \]
- If error is larger for larger \(|y_j|\), let \( w_j^{-1} = y_j^2 \).
- Alternatively, assume that \( w_j^{-1} = y^2(t_j; q) \).
- Assume some other model for \( w_j \), e.g.,
  \[
  w_j^{-1} = y^{2\theta}(t_j; q)
  \]
  where \( \theta \) is to be estimated from the data.
Weighted Least Squares

Deal with correlated observations and nonconstant variance:
Let $\mathbf{e} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ and assume

$$\text{Cov}(\varepsilon) = \sigma^2 V,$$

for known matrix $V$.

Let $a = (a_1, a_2, \ldots, a_n)^T$, $y(q) = (y(t_1; q), \ldots, y(t_n; q))^T$, $W = V^{-1}$. 
The weighted least-squares (WLS) estimator minimizes

\[ J_{\text{wls}}(q) = \{ a - y(q) \}^T W \{ a - y(q) \} \]  \hspace{1cm} (2)

in \( q \), to give \( \hat{q}_{\text{wls}} \).

If above covariance model holds (together with assumption 1) then \( \hat{q}_{\text{wls}} \) has good properties.
Generalized least-squares (GLS)

Estimate $V$ from the data too!

Model $\text{Cov}(\varepsilon) = \sigma^2 V$ in two stages, based on additional parameters $\theta$, $\alpha$ to be estimated from the data.

(a) Model $\text{Var}(\varepsilon_j)$. For example,

$$\text{Var}(\varepsilon_j) = \sigma^2 y^{2\theta}(t_j; q), \quad \text{where } \theta = \theta \text{ (i.e., scalar).}$$

Define diagonal matrices:

$$G(q, \theta) = \text{diag}\{y^{2\theta}(t_1; q), y^{2\theta}(t_2; q), \ldots, y^{2\theta}(t_n; q)\}, \quad \text{and}$$

$$\{G(q, \theta)\}^{1/2} = \text{diag}\{y^{\theta}(t_1; q), y^{\theta}(t_2; q), \ldots, y^{\theta}(t_n; q)\}.$$
(b) Model $\text{Corr}(\varepsilon_j, \varepsilon_k)$ for $j \neq k$. For example,

$$\text{Corr}(\varepsilon_j, \varepsilon_k) = \alpha^{|j-k|},$$

where $\alpha = \alpha$ (i.e., scalar) and $|\alpha| < 1$.

Organize into a matrix:

$$\text{Corr}(\varepsilon) = \Gamma(\alpha) = \begin{pmatrix}
1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\
\alpha & 1 & \alpha & \cdots & \alpha^{n-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \cdots & 1
\end{pmatrix}.$$

Put pieces together, write

$$\text{Cov}(\varepsilon) = \sigma^2 \{V(q, \theta, \alpha)\} = \sigma^2 \{G(q, \theta)\}^{1/2} \Gamma(\alpha) \{G(q, \theta)\}^{1/2}.$$
Typical GLS algorithm follows three steps:

(i) Estimate $q$ with OLS. Set $\hat{q}_{gls} = \hat{q}_{ols}$.

(ii) Estimate (somehow) $(\hat{\theta}, \hat{\alpha})$. Plug into the model for $V$ to form estimated weight matrix,

$$\hat{W} = \{V(\hat{q}_{gls}, \hat{\theta}, \hat{\alpha})\}^{-1}.$$ 

(iii) Minimize (approximate) WLS objective function (2) in $q$, namely,

$$\hat{J}_{wls}(q) = \{a - y(q)\}^T \hat{W} \{a - y(q)\}.$$ 

Set the minimizing value equal to $\hat{q}_{gls}$.

Return to step (ii).

Iterate until “convergence”.
Finally, compute estimate of $\sigma^2$:

$$\hat{\sigma}^2_{gls} = \frac{1}{n - p} \{a - y(\hat{q}_{gls})\}^T \{V(\hat{q}_{gls}, \hat{\theta}, \hat{\alpha})\}^{-1} \{a - y(\hat{q}_{gls})\}.$$ 

Methods for estimating $(\hat{\theta}, \hat{\alpha})$ [step (ii)] are beyond the scope of this talk. Commonly used methods include

- psuedolikelihood (PL)
- restricted maximum likelihood (REML).
Concluding comments

- Does this approach work better? Are our new statistical assumptions met?
- Check! Calculate the GLS weighted residuals,
\[ r_{gls} = \{V(\hat{q}_{gls}, \hat{\theta}, \hat{\alpha})\}^{-1/2} \{a - y(\hat{q}_{gls})\} \]
and plot against \(t_j\).
- What about standard errors for \(\hat{q}_{gls}\)?
- Approximate s.e.(\(\hat{q}_{gls,k}\)) = square root of \((k, k)\) element in
\[ \hat{\sigma}^2_{gls} \left[ X^T \{V(\hat{q}_{gls}, \hat{\theta}, \hat{\alpha})\}^{-1} X \right]^{-1} \]
where \(X_{r,c} = \left( \frac{\partial y(t_r;q)}{\partial q_c} \right)\) evaluated at \(\hat{q}_{gls}\).