Second-order Least Squares Estimation in Nonlinear Models

Liqun Wang
Department of Statistics
University of Manitoba
e-mail: wangl1@cc.umanitoba.ca
web: home.cc.umanitoba.ca/~wangl1

WU Wien, June 26, 2009
Outline

- Second-order least squares (SLS) method
- SLS and ordinary least squares (OLS) method
- SLS and the generalized method of moments (GMM)
- The errors-in-variables problem
- Simulation-based SLSE in errors-in-variables models
- Simulation-based SLSE in mixed effects models
The model $Y = g(X; \beta) + \varepsilon$ with $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) = \sigma^2$.

To estimate $\gamma = (\beta', \sigma^2)'$ with i.i.d. random sample $(Y_i, X_i), i = 1, 2, ..., n$.

The OLSE minimizes the (sum of squared) "first-order" distances

$$S_n(\beta) = \sum_{i=1}^{n} (Y_i - E(Y_i|X_i))^2 = \sum_{i=1}^{n} (Y_i - g(X_i; \beta))^2$$

The OLSE for $\sigma^2$ is defined as $\hat{\sigma}^2_{OLS} = \frac{1}{n} S_n(\hat{\beta}_{OLS})$.

The SLSE minimizes the distances $Y_i - E(Y_i|X_i)$ and $Y_i^2 - E(Y_i^2|X_i)$ simultaneously.

The $\hat{\sigma}^2_{SLS}$ is obtained through optimization.
Second-order Least Squares Method

- The SLSE for $\gamma$ is defined as $\hat{\gamma}_{SLS} = \text{argmin}_{\gamma} Q_n(\gamma)$, where

$$Q_n(\gamma) = \sum_{i=1}^{n} \rho_i'(\gamma) A_i \rho_i(\gamma),$$

$$\rho_i(\gamma) = (Y_i - g(X_i; \beta), Y_i^2 - g^2(X_i; \beta) - \sigma^2)'$$

and $A_i = A(X_i)$ is a $2 \times 2$ n.d. weighting matrix.

- Under conditions 1-4, $\hat{\gamma}_{SLS} \xrightarrow{a.s.} \gamma$, as $n \to \infty$.

- Under conditions 1-6, $\sqrt{n}(\hat{\gamma}_{SLS} - \gamma) \xrightarrow{L} N(0, B^{-1}CB^{-1})$, where

$$B = E \left[ \frac{\partial \rho'(\gamma)}{\partial \gamma} A \frac{\partial \rho(\gamma)}{\partial \gamma'} \right], \quad C = E \left[ \frac{\partial \rho'(\gamma)}{\partial \gamma} A \rho(\gamma) \rho'(\gamma) A \frac{\partial \rho(\gamma)}{\partial \gamma'} \right].$$
Regularity conditions for SLSE

1. $g(x; \beta)$ is a measurable in $x$ and continuous in $\beta$ a.e.
2. $E \|A(X)\| \left( \sup_{\beta} g^4(x; \beta) + 1 \right) < \infty$.
3. The parameter space $\Gamma \subset \mathbb{R}^{p+1}$ is compact.
4. $E[\rho(\gamma) - \rho(\gamma_0)]'A(X)[\rho(\gamma) - \rho(\gamma_0)] = 0$ if and only if $\gamma = \gamma_0$.
5. $g(x; \beta)$ is twice continuously differentiable w.r.t. $\beta$ and
   
   $E \|A(X)\| \sup_{\beta} \left( \left\| \frac{\partial g(X;\beta)}{\partial \beta} \right\|^4 + \left\| \frac{\partial^2 g(X;\beta)}{\partial \beta \partial \beta'} \right\|^4 \right) < \infty$.

6. The matrix $B = E \left[ \frac{\partial \rho'(\gamma)}{\partial \gamma} A(X) \frac{\partial \rho(\gamma)}{\partial \gamma'} \right]$ is nonsingular.
Efficient Choice of Weighing Matrix

- How to choose $A$ to obtain the most efficient estimator $\hat{\gamma}_n$ in the class of all SLSE?

- We can show that $B^{-1}CB^{-1} \geq E^{-1} \left[ \frac{\partial \rho'(\gamma)}{\partial \gamma} A_0 \frac{\partial \rho(\gamma)}{\partial \gamma'} \right]$ and the lower bound is attained for $A = A_0$ in $B$ and $C$, where

$$A_0 = \left[ \sigma^2 (\mu_4 - \sigma^4) - \mu_3^2 \right]^{-1} \times \begin{pmatrix} \mu_4 + 4\mu_3 g(X; \beta) + 4\sigma^2 g^2(X; \beta) - \sigma^4 & -\mu_3 - 2\sigma^2 g(X; \beta) \\ -\mu_3 - 2\sigma^2 g(X; \beta) & \sigma^2 \end{pmatrix},$$

$$\mu_3 = E(\varepsilon^3|X) \text{ and } \mu_4 = E(\varepsilon^4|X).$$

- Since $A_0$ depends on $\gamma$, a two-stage procedure can be used:
  1. minimize $Q_n(\gamma)$ using identity weight $A = I$ to obtain $\tilde{\gamma}_n$ and $\hat{\mu}_3, \hat{\mu}_4$ using residuals $\hat{\varepsilon}_i = Y_i - g(X_i; \tilde{\beta})$;
  2. estimate $A_0$ using $\tilde{\gamma}, \hat{\mu}_3, \hat{\mu}_4$ and minimize $Q_n(\gamma)$ again with $A = \hat{A}_0$. 
The Most Efficient SLS Estimator

The most efficient SLSE has asymptotic covariance matrix

\[
C_0 = \begin{pmatrix}
V(\hat{\beta}_{SLS}) & \frac{\mu_3}{\mu_4 - \sigma^4} V(\hat{\sigma}^2_{SLS}) G_1 G_2^{-1} \\
\frac{\mu_3}{\mu_4 - \sigma^4} V(\hat{\sigma}^2_{SLS}) G_1 G_2^{-1} & V(\hat{\sigma}^2_{SLS})
\end{pmatrix},
\]

where

\[
V(\hat{\beta}_{SLS}) = \left(\sigma^2 - \frac{\mu_3^2}{\mu_4 - \sigma^4}\right) \left(G_2 - \frac{\mu_3^2}{\sigma^2(\mu_4 - \sigma^4)} G_1 G_1'\right)^{-1},
\]

\[
V(\hat{\sigma}^2_{SLS}) = \frac{(\mu_4 - \sigma^4) (\sigma^2(\mu_4 - \sigma^4) - \mu_3^2)}{\sigma^2(\mu_4 - \sigma^4) - \mu_3^2 G_1' G_2^{-1} G_1},
\]

\[
G_1 = E \left[\frac{\partial g(X; \beta)}{\partial \beta}\right], \quad G_2 = E \left[\frac{\partial g(X; \beta)}{\partial \beta} \frac{\partial g(X; \beta)}{\partial \beta'}\right].
\]
Under similar conditions, the OLSE $\hat{\gamma}_{OLS} = (\hat{\beta}'_{OLS}, \hat{\sigma}^2_{OLS})'$ has asymptotic covariance matrix

$$D = \begin{pmatrix} \sigma^2 G_2^{-1} & \mu_3 G_2^{-1} G_1 \\ \mu_3 G_1' G_2^{-1} & \mu_4 - \sigma^4 \end{pmatrix}.$$ 

If $\mu_3 = E(\varepsilon^3) \neq 0$, then

1. $V(\hat{\beta}_{OLS}) - V(\hat{\beta}_{SLS})$ is p.d. when $G_1' G_2^{-1} G_1 \neq 1$, and is n.d. when $G_1' G_2^{-1} G_1 = 1$;
2. $V(\hat{\sigma}^2_{OLS}) \geq V(\hat{\sigma}^2_{SLS})$ with equality holding iff $G_1' G_2^{-1} G_1 = 1$.

If $\mu_3 = 0$, then $\hat{\gamma}_{SLS}$ and $\hat{\gamma}_{OLS}$ have the same asymptotic covariance matrices.
A Simulation Study

- An exponential model $Y = \beta_1 \exp(-\beta_2 X) + \varepsilon$, where $\varepsilon = (\chi^2(3) - 3) / \sqrt{3}$.
- Generate data using $X \sim \text{Uniform}(0, 20)$ and $\beta_1 = 10$, $\beta_2 = 0.6$, $\sigma^2 = 2$.
- Sample size $n = 30, 50, 100, 200$.
- Monte Carlo replications $N = 1000$.
## A Simulation Study

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>VAR</th>
<th>MSE</th>
<th>SLS</th>
<th>VAR</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_1 = 10)</td>
<td>10.0315</td>
<td>2.0245</td>
<td>2.0255</td>
<td>10.2306</td>
<td>1.6380</td>
<td>1.6895</td>
</tr>
<tr>
<td>(\beta_2 = 0.6)</td>
<td>0.6139</td>
<td>0.0189</td>
<td>0.0190</td>
<td>0.6282</td>
<td>0.0141</td>
<td>0.0149</td>
</tr>
<tr>
<td>(\sigma^2 = 2)</td>
<td>2.0027</td>
<td>0.7656</td>
<td>0.7648</td>
<td>1.7026</td>
<td>0.3093</td>
<td>0.3974</td>
</tr>
<tr>
<td>(n = 50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_1 = 10)</td>
<td>10.0238</td>
<td>1.4738</td>
<td>1.4743</td>
<td>10.1880</td>
<td>1.1669</td>
<td>1.2011</td>
</tr>
<tr>
<td>(\beta_2 = 0.6)</td>
<td>0.6109</td>
<td>0.0141</td>
<td>0.0142</td>
<td>0.6241</td>
<td>0.0100</td>
<td>0.0105</td>
</tr>
<tr>
<td>(\sigma^2 = 2)</td>
<td>1.9763</td>
<td>0.5194</td>
<td>0.5194</td>
<td>1.7733</td>
<td>0.2430</td>
<td>0.2941</td>
</tr>
<tr>
<td>(n = 100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_1 = 10)</td>
<td>9.9802</td>
<td>0.9863</td>
<td>0.9867</td>
<td>10.1146</td>
<td>0.6428</td>
<td>0.6553</td>
</tr>
<tr>
<td>(\beta_2 = 0.6)</td>
<td>0.6032</td>
<td>0.0074</td>
<td>0.0074</td>
<td>0.6133</td>
<td>0.0046</td>
<td>0.0048</td>
</tr>
<tr>
<td>(\sigma^2 = 2)</td>
<td>2.0061</td>
<td>0.2693</td>
<td>0.2694</td>
<td>1.8891</td>
<td>0.1573</td>
<td>0.1695</td>
</tr>
<tr>
<td>(n = 200)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_1 = 10)</td>
<td>10.0153</td>
<td>0.5467</td>
<td>0.5469</td>
<td>10.0522</td>
<td>0.3361</td>
<td>0.3384</td>
</tr>
<tr>
<td>(\beta_2 = 0.6)</td>
<td>0.6028</td>
<td>0.0038</td>
<td>0.0038</td>
<td>0.6054</td>
<td>0.0023</td>
<td>0.0024</td>
</tr>
<tr>
<td>(\sigma^2 = 2)</td>
<td>2.0077</td>
<td>0.1129</td>
<td>0.1129</td>
<td>1.9504</td>
<td>0.0774</td>
<td>0.0798</td>
</tr>
</tbody>
</table>
GMM using the first two conditional moments minimizes

\[ Q_n(\gamma) = \left( \sum_{i=1}^{n} \rho_i(\gamma) \right)' A_n \left( \sum_{i=1}^{n} \rho_i(\gamma) \right), \]

where \( \rho_i(\gamma) = (Y_i - g(X_i; \beta), Y_i^2 - g^2(X_i; \beta) - \sigma^2)' \) and \( A_n \) is n.d.

The most efficient GMM estimator has the asymptotic covariance

\[ \left[ E \left( \frac{\partial \rho_i'(\gamma)}{\partial \gamma} \right) A_0 E \left( \frac{\partial \rho_i(\gamma)}{\partial \gamma'} \right) \right]^{-1}, \]

where \( A_0 = E^{-1}[\rho_i(\gamma)\rho_i'(\gamma)] \) is the optimal weighting matrix.

We have \( V(\hat{\beta}_{GMM}) \geq V(\hat{\beta}_{SLS}) \) and \( V(\hat{\sigma}_{GMM}^2) \geq V(\hat{\sigma}_{SLS}^2). \)
The relationship of interest: \( Y = \beta_0 + \beta_x X + \varepsilon \), where
\( Y \): response variable, \( X \): explanatory variable,
\( \varepsilon \): is uncorrelated with \( X \) and \( E(\varepsilon) = 0 \).

Given an \( i.i.d. \) random sample \( (X_i, Y_i), i = 1, 2, \ldots, n \)

The ordinary least squares estimator (MLE under normality) is unbiased and consistent: as \( n \to \infty \),

\[
\hat{\beta}_x = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \xrightarrow{P} \frac{Cov(X, Y)}{Var(X)} = \beta_x.
\]

Implicit assumption: \( X \) is directly and precisely measured.
Example of Measurement Error

- Coronary heart disease in relation to systolic blood pressure:

  \[ E(Y|X) = g(\beta_0 + \beta_x X, \ldots), \]

  \(Y\): CHD indicator or severity, \(X\): long-term average SBP, and \(g\) is a known function, e.g., logistic.

- The observed SBP variable is
  
  \(Z\): blood pressure measured during a clinic visit on a given day

- Therefore \(Z = X + e\), where \(e\) is a random ME
Example of Measurement Error

- Individual lung cancer risk and exposure to certain air pollutants:

  \[ E(Y|X) = g(\beta_0 + \beta'_x X, \ldots), \]

  \( Y \): lung cancer incidence, \( X \): individual exposure to the pollutants, and \( g \) is a known function, e.g. logistic.

- The observed exposure variable is

  \( Z \): level of pollutants measured at certain monitoring stations, or calculated group average

- Therefore \( X = Z + e \), where \( e \) is a random ME
A pharmacokinetic study of the efficacy of a drug:

\[ E(Y|X) = g(X, \beta, ...), \]

where \( Y \): effect of the drug; \( X \): actual absorption of the medical substance in bloodstream

The observed predictor is \( Z \): predetermined dosage of the drug

Therefore \( X = Z + e \), where \( e \) is a random ME.

Yield of a crop and the amount of fertilizer used:

\[ Y = g(X, \beta, ...), \]

where \( Y \): yield; \( X \): actual absorption of the fertilizer in the crop

The actual observed predictor is \( Z \): predetermined dose of the fertilizer

Therefore \( X = Z + e \), where \( e \) is a random ME.
Examples of Measurement Error

- Capital asset pricing model (CAPM): \( R_a = \beta_0 + \beta_1 R_m + u \), where \( R_a, R_m \) are the excess returns of an asset and true market portfolio respectively.

- \( R_m \) is unobserved and estimated by regressing on market portfolio.

- A more general factor model (Fama and French (1993); Carhart (1997)):

\[
R_a = \beta_0 + \beta_1 F_m + \beta_2 F_{smb} + \beta_3 F_{hml} + \beta_4 F_{umd} + u
\]

where the unobserved true factors
- \( F_m = R_m \): market effect
- \( F_{smb} \): portfolio size effect (small minus big)
- \( F_{hml} \): book-to-market effect (high minus low)
- \( F_{umd} \): momentum effect (up minus down)

- The constructed factors: \( \hat{F} = F + e \)
Examples of Measurement Error

- Index option price volatilities:

\[ V_t^r = \beta_0 + \beta_1 V_t^i + \beta_2 V_{t-1}^h + \varepsilon_t \]

where \( V_t^r, V_t^i, V_t^h \) are the realized, implied, historical volatility respectively.

- The implied volatility \( V_t^i \) is estimated using some option pricing model: \( V_t^i = \bar{V}_t^i + e \).

- Income function in labor market:
  \( Y \): personal income (wage)
  \( X \): education, experience, job-related ability, etc.
  \( Z \): schooling, working history, etc.

- Consumption function of Friedman (1957):
  \( Y \): permanent consumption
  \( X \): permanent income
  \( Z \): annual income or tax data
Examples of Measurement Error

- Environmental variables:
  \( X \): biomass, greenness of vegetation, etc.
  \( Z \): satellite image or spatial average

- Long-term nutrition (fat, energy) intake, alcohol (smoke) consumption, etc.
  \( X \): actual intake or consumption
  \( Z \): report on food questionnaire or 24 hour recall interview

- Some demographic variables
  \( X \): education, experience, family wealth, poverty, etc.
  \( Z \): schooling, working history, tax report income, etc.
Generate independent $X_i \sim UNIF(-2, 2), i = 1, 2, \ldots, n = 20$

Generate independent $\varepsilon_i \sim N(0, 0.1)$ and let
$Y_i = \beta_0 + \beta_x X_i + \varepsilon_i$, where $\beta_0 = 0, \beta_x = 1$

Fit the least squares line to $(Y_i, X_i)$

Generate independent $e_i \sim N(0, 0.5)$ and let $Z_i = X_i + e_i$

Fit the least squares line to $(Y_i, Z_i)$

Repeat using $\sigma^2_e = 1, 2$ respectively
Impact of Measurement Error: A simulation study

\[
\sigma_e^2 = 0 \\
b = 1, R^2 = 0.93
\]

\[
\sigma_e^2 = 0.5 \\
b = 0.71, R^2 = 0.8
\]

\[
\sigma_e^2 = 1 \\
b = 0.53, R^2 = 0.61
\]

\[
\sigma_e^2 = 2 \\
b = 0.36, R^2 = 0.23
\]
Generate independent $X_i \sim \text{UNIF}(0, 1), i = 1, 2, \ldots, n = 40$

Generate $Y_i = \sin(2\pi X_i) + \varepsilon_i$, where $\varepsilon_i \sim N(0, 0.2^2)$

Generate $Z_i = X_i + e_i$, where $e_i \sim N(0, 0.2^2)$

Plot $(X_i, Y_i)$ and $(Z_i, Y_i)$
Impact of Measurement Error: A simulation study
The relationship of interest: \( Y = \beta_0 + \beta_x X + \varepsilon, \varepsilon|X \sim (0, \sigma^2) \)

Actual data: \( Y, Z = X + e, \) where \( e \) is independent of \( X. \)

The naive model ignoring ME: \( Y = \beta_0^* + \beta_z Z + \varepsilon^* \)

The naive least squares estimator

\[
\hat{\beta}_z \xrightarrow{P} \beta_z = \frac{\sigma_x^2 \beta_x}{\sigma_z^2} = \lambda \beta_x
\]

The attenuation factor

\[
\lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2} \leq 1 \quad \text{and} \quad \lambda = 1 \quad \text{if and only if} \quad \sigma_e^2 = 0.
\]

The LSE of the intercept: \( \hat{\beta}_0^* \xrightarrow{P} \beta_0^* = \beta_0 + (1 - \lambda) \beta_x \mu_x \)

The LSE of the error variance: \( \hat{\sigma}^2^* \xrightarrow{P} \sigma^2^* = \sigma^2 + \lambda \beta_x^2 \sigma_e^2 \)
The simple linear model: $Y = \beta_0 + \beta_x X + \varepsilon$, $Z = X + e$

Suppose $X \sim N(\mu_x, \sigma_x^2)$, $e \sim N(0, \sigma_e^2)$, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, independent.

The joint distribution of the observed variables $(Y, Z)$ is normal.

Therefore all observable information is contained in the first two moments

\[
E(Y) = \beta_0 + \beta_x \mu_x, \quad E(Z) = \mu_x
\]

\[
\text{Var}(Y) = \beta_x^2 \sigma_x^2 + \sigma_\varepsilon^2, \quad \text{Cov}(Y, Z) = \beta_x \sigma_x^2
\]

\[
\text{Var}(Z) = \sigma_x^2 + \sigma_e^2
\]

There are 5 moment equations but 6 unknown parameters.

In practice usually \textit{ad hoc} restrictions are imposed to ensure unique solution (e.g. $\sigma_e^2$, $\sigma_e^2/\sigma_\varepsilon^2$ or $\sigma_e^2/\sigma_x^2$ is known or can be estimated using extra data).
The response model: \( Y = g(X, \beta) + \varepsilon \), where
\( Y \): the response variable; \( X \): unobserved predictor (vector);
\( \varepsilon \): random error independent of \( X \); and
\( g \) is nonlinear in general, e.g., generalized linear models.

The observed predictor is \( Z \) (vector)

Classical ME: \( Z = X + e \), \( e \) independent of \( X \) and therefore
\( \text{Var}(Z) > \text{Var}(X) \). E.g. blood pressure.

Berkson ME: \( X = Z + e \), \( e \) independent of \( Z \) and therefore
\( \text{Var}(X) > \text{Var}(Z) \). E.g. pollutants exposure.

The two types of ME lead to different statistical structures of the full model and therefore require different treatments.
Identifiability of Nonlinear EIV Models

- The identifiability of nonlinear EIV models is a long-standing and challenging problem.
- Nonlinear models with Berkson ME are generally identifiable without extra data:
  - Rudemo, Ruppert and Streibig (1989): logistic model
  - Huwang and Huang (2000): univariate polynomial models
- Nonlinear classical ME models are identifiable with replicate data: Li (2002), Schennach (2004).
- Identifiability with instrumental variables (IV):
  - Hausman et al. (1991): univariate polynomial models
  - Wang and Hsiao (1995, 2007): regression function $g \in L_1(R^k)$
  - Schennach (2007): $|g|$ is univariate and bounded by polynomials in $R$. 
Maximum Likelihood Estimation

- Likelihood analysis in nonlinear EIV models is difficult, because of intractability of the likelihood function.
- Example: Suppose $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and $e \sim N(0, \sigma_e^2)$.
- The likelihood function is a product of the conditional density

$$f(y|z) = \int f(y|x)f(x|z)dx$$

$$= \frac{1}{2\pi\sigma_\varepsilon\sigma_e} \int \exp \left[ -\frac{(y - g(x; \beta))^2}{2\sigma_\varepsilon^2} - \frac{(x - z)^2}{2\sigma_e^2} \right] dx.$$  

- The closed form is not available for general $g$.
- Numerical approximations such as quadrature methods result in inconsistent estimators.

Assume the conditional density $f(x|z)$ has known parametric form: Hsiao (1989, 1992), Li and Hsiao (2004)


Nonlinear model with replicate data: Li (2002), Schennach (2004)

Approximate estimation when ME are small:
- regression calibration: Carroll and Stefanski (1990), Gleser (1990), Rosner, Willett and Spielgelman (1990)

Estimation in Berkson ME models:
- logistic model: Rudemo, Ruppert and Streibig (1989)
- univariate polynomial model: Huwang and Huang (2000)
A quadratic model with Berkson ME

\[ Y = \beta_0 + \beta_1 X^2 + \varepsilon, \varepsilon \sim N(0, \sigma^2) \]
\[ X = Z + e, e \sim N(0, \sigma_e^2) \]

The first two conditional moments

\[ E(Y|Z) = \beta_0 + \beta_1 \sigma_e^2 + \beta_1 Z^2 \]
\[ E(Y^2|Z) = \sigma^2 + (\beta_0 + \beta_1 \sigma_e^2)^2 + 2\beta_1^2 \sigma_e^4 \]
\[ + 2\beta_1 (\beta_0 + 3\beta_1 \sigma_e^2) Z^2 + \beta_1^2 Z^4 \]

All unknown parameters are identifiable by these two equations and the nonlinear least square method.
Estimation in Berkson ME models

- A Berkson ME model: $Y = g(X; \beta) + \varepsilon, X = Z + e$, where $e$ is independent of $Z, \varepsilon$ and $e \sim f_e(u, \psi)$.

- The goal is to estimate $\gamma = (\beta', \psi', \sigma^2)'$ given random sample $(Y_i, Z_i), i = 1, 2, ..., n$.

- The SLSE is $\hat{\gamma}_n = \arg\min_{\gamma} Q_n(\gamma)$, where $Q_n(\gamma) = \sum_{i=1}^{n} \rho_i'(\gamma)A_i\rho_i(\gamma)$,

$$\rho_i(\gamma) = (Y_i - E(Y_i|Z_i, \gamma), Y_i^2 - E(Y_i^2|Z_i, \gamma))'$$

and $A_i = W(Z_i)$ is a $2 \times 2$ weighting matrix.

- Under some regularity conditions, as $n \to \infty$, we have $\hat{\gamma}_n \overset{a.s.}{\to} \gamma$ and $\sqrt{n}(\hat{\gamma}_n - \gamma) \overset{L}{\to} N(0, B^{-1}CB^{-1})$, where

$$B = E \left[ \frac{\partial \rho'(\gamma)}{\partial \gamma} A \frac{\partial \rho(\gamma)}{\partial \gamma'} \right], \quad C = E \left[ \frac{\partial \rho'(\gamma)}{\partial \gamma} A \rho(\gamma) \rho'(\gamma) A \frac{\partial \rho(\gamma)}{\partial \gamma'} \right]$$
A quadratic model

\[ Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i, \]
\[ X_i = Z_i + e_i, \]

where \( \varepsilon_i \sim N(0, \sigma^2) \), \( e_i \sim N(0, \sigma_e^2) \) independent.

Generate data using \( Z_i \sim N(2, 1) \) and
\( \beta_0 = 3, \beta_1 = 2, \beta_2 = 1, \sigma^2 = 1, \sigma_e^2 = 2. \)

Sample size \( n = 100 \)

Monte Carlo replications \( N = 1000 \)
Example: Quadratic Model

\[ \beta_0 = 3 \quad \beta_1 = 2 \quad \beta_2 = 1 \quad \sigma^2 = 1 \quad \sigma^2_e = 2 \]

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\sigma^2$</th>
<th>$\sigma^2_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS1</td>
<td>3.046</td>
<td>2.052</td>
<td>0.995</td>
<td>0.983</td>
<td>2.028</td>
</tr>
<tr>
<td>(Std.)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.010)</td>
<td>(0.019)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>SLS2</td>
<td>3.024</td>
<td>2.048</td>
<td>0.975</td>
<td>1.073</td>
<td>2.026</td>
</tr>
<tr>
<td>(Std.)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.010)</td>
<td>(0.020)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>NLS</td>
<td>5.025</td>
<td>1.929</td>
<td>1.024</td>
<td>88.356</td>
<td>NA</td>
</tr>
<tr>
<td>(Std.)</td>
<td>(0.064)</td>
<td>(0.087)</td>
<td>(0.026)</td>
<td>(0.622)</td>
<td>NA</td>
</tr>
</tbody>
</table>

SLS1: SLSE using identity weight
SLS2: SLSE using optimal weight
NLS: Naive nonlinear least squares estimates ignoring ME
In general the first two moments are

\[ E(Y_i | Z_i, \gamma) = \int g(Z + u, \beta) f_e(u; \psi) du \]
\[ E(Y_i^2 | Z_i, \gamma) = \int g^2(Z + u, \beta) f_e(u; \psi) du + \sigma^2 \]

If the integrals have no closed forms and the dimension is higher than two or three, then numerical minimization of \( Q_n(\gamma) \) is difficult.

In this case, they can be replaced by Monte Carlo simulators:

\[ \frac{1}{S} \sum_{j=1}^{S} \frac{g(Z_i + u_{ij}, \beta)f_e(u_{ij}; \psi)}{h(u_{ij})} \], \quad \frac{1}{S} \sum_{j=1}^{S} \frac{g^2(Z_i + u_{ij}, \beta)f_e(u_{ij}; \psi)}{h(u_{ij})} + \sigma^2 \]

where \( u_{ij} \) are generated from a known density \( h(u) \).
Choose a known density $h(u)$ such that $\text{Supp}(h) \supseteq \text{Supp}(f_e(u; \psi))$.

Generate random points $u_{ij} \sim h(u)$, $i = 1, \ldots, n$, $j = 1, \ldots, 2S$ and calculate $\rho_{i,1}(\gamma)$ using $u_{ij}, j = 1, 2, \ldots, S$ and $\rho_{i,2}(\gamma)$ using $u_{ij}, j = S + 1, S + 2, \ldots, 2S$.

Then $\rho_{i,1}(\gamma)$ and $\rho_{i,2}(\gamma)$ are conditionally independent given data and therefore

$$Q_{n,S}(\gamma) = \sum_{i=1}^{n} \rho'_{i,1}(\gamma) A_i \rho_{i,2}(\gamma),$$

is an unbiased simulator for $Q_n(\gamma)$.

The simulation-based SLS estimator is $\hat{\gamma}_{n,S} = \arg\min_{\gamma} Q_{n,S}(\gamma)$. 

The simulation-based SLS Estimator
Simulation-based SLS Estimator

- Under the same regularity conditions for the SLSE, for any fixed $S$, $\hat{\gamma}_{n,S} \xrightarrow{a.s.} \gamma$ as $n \to \infty$ and $\sqrt{n}(\hat{\gamma}_{n,S} - \gamma) \xrightarrow{L} N(0, B^{-1}C_S B^{-1})$, where

$$2C_S = E \left[ \frac{\partial \rho_1'(\gamma)}{\partial \gamma} W \rho_2(\gamma) \rho_2'(\gamma) W \frac{\partial \rho_1(\gamma)}{\partial \gamma'} \right]$$

$$+ E \left[ \frac{\partial \rho_1'(\gamma)}{\partial \gamma} W \rho_2(\gamma) \rho_1'(\gamma) W \frac{\partial \rho_2(\gamma)}{\partial \gamma'} \right]$$

- How much efficiency is lost due to simulation?
- We can show that

$$C_S = C + \frac{1}{S} M_1 + \frac{1}{S^2} M_2,$$

where $M_1$ and $M_2$ are two constant matrices.
- Therefore the efficiency loss is of order $O(1/S)$. 
A linear-exponential model

\[ Y_i = \beta_1 X_{1i} + \beta_2 \exp(-\beta_3 X_{2i}) + \varepsilon_i, \]
\[ X_{1i} = Z_{1i} + e_{1i}, \]
\[ X_{2i} = Z_{2i} + e_{2i}, \]

where \( \varepsilon_i \sim N(0, \sigma^2) \), \( e_{1i} \sim N(0, \sigma_1^2) \), \( e_{2i} \sim N(0, \sigma_2^2) \) independent.

Generate data using \( Z_i \sim N(1, 1) \) and \( \beta_0 = 3, \beta_1 = 2, \beta_2 = 1, \sigma^2 = 1, \sigma_1^2 = 1, \sigma_2^2 = 1.5 \).

Choose \( h(u) \) to be the density of \( N_2(0, \text{diag}(5, 5)) \) and \( S = 1000 \).

Sample size \( n = 100 \), and Monte Carlo replications \( N = 1000 \).
### Example: Linear-Exponential Model

\[
\begin{align*}
\beta_1 &= 3 \\
\beta_2 &= 2 \\
\beta_3 &= 1 \\
\sigma^2 &= 1 \\
\sigma_1^2 &= 1 \\
\sigma_2^2 &= 1.5
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\sigma^2$</th>
<th>$\sigma_1^2$</th>
<th>$\sigma_2^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS1</td>
<td>3.000</td>
<td>2.009</td>
<td>0.878</td>
<td>1.023</td>
<td>1.073</td>
<td>1.356</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.009)</td>
<td>(0.011)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>SLS2</td>
<td>2.987</td>
<td>2.066</td>
<td>0.869</td>
<td>1.026</td>
<td>1.039</td>
<td>1.275</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.003)</td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>SbSLS</td>
<td>3.002</td>
<td>1.898</td>
<td>0.947</td>
<td>1.000</td>
<td>1.003</td>
<td>1.319</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>NLS</td>
<td>3.215</td>
<td>2.391</td>
<td>1.017</td>
<td>45.557</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(3.365)</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

- **SLS1**: SLSE using identity weight
- **SLS2**: SLSE using optimal weight
- **SbSLS**: Simulation-based SLSE using identity weight
- **NLS**: Naive nonlinear least squares estimates ignoring ME
Estimation in Classical ME Model

- A semiparametric model with classical ME and IV:

\[ Y = g(X, \beta) + \varepsilon \]
\[ Z = X + e \]
\[ X = \Gamma W + U \]

- \( Y \in \mathbb{R}, Z \in \mathbb{R}^k, W \in \mathbb{R}^\ell \) are observed;
- \( X \in \mathbb{R}^k, \beta \in \mathbb{R}^p, \Gamma \in \mathbb{R}^{k \times \ell} \) are unobserved;
- \( E(\varepsilon | X, Z, W) = 0 \) and \( E(e | X, W) = 0 \);
- \( U \) and \( W \) independent and \( E(U) = 0 \);
- Suppose \( U \sim f_U(u; \phi) \) which is known up to \( \phi \in \mathbb{R}^q \).
- \( X, \varepsilon \) and \( e \) have nonparametric distributions.
SLS-IV Estimation for Classical ME Models

- Under model assumptions:

\[
E(Z \mid W) = \Gamma W \\
E(Y \mid W) = \int g(x; \beta)f_U(x - \Gamma W; \phi)dx \\
E(YZ \mid W) = \int xg(x; \beta)f_U(x - \Gamma W; \phi)dx \\
E(Y^2 \mid W) = \int g^2(x; \beta)f_U(x - \Gamma W; \phi)dx + \sigma_\varepsilon^2
\]

- \( \Gamma \) can be estimated by the LSE \( \hat{\Gamma} = (\sum Z_j W_j')(\sum W_j W_j')^{-1} \).

- Given \( \hat{\Gamma} \), to estimate \( \gamma = (\beta, \phi, \sigma^2) \) using (2)-(4).

- The SLS-IV estimator is \( \hat{\gamma}_n = \arg\min_\psi \sum_{i=1}^n \rho_i'(\gamma)A_i\rho_i(\gamma) \), where

\[
\rho_i(\gamma) = (Y_i - E(Y_i \mid W_i, \gamma), Y_i^2 - E(Y_i^2 \mid W_i, \gamma), Y_i Z_i - E(Y_i Z_i \mid W_i, \gamma))'.
\]
Orange Tree data (Draper and Smith 1981, p.524): Trunk circumference (in mm) of 5 orange trees measured on 7 occasions over a period of 1600 days from December 31, 1968.

<table>
<thead>
<tr>
<th>Day</th>
<th>Tree 1</th>
<th>Tree 2</th>
<th>Tree 3</th>
<th>Tree 4</th>
<th>Tree 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>118</td>
<td>30</td>
<td>33</td>
<td>30</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>484</td>
<td>58</td>
<td>69</td>
<td>51</td>
<td>62</td>
<td>49</td>
</tr>
<tr>
<td>664</td>
<td>87</td>
<td>111</td>
<td>75</td>
<td>112</td>
<td>81</td>
</tr>
<tr>
<td>1004</td>
<td>115</td>
<td>156</td>
<td>108</td>
<td>167</td>
<td>125</td>
</tr>
<tr>
<td>1231</td>
<td>120</td>
<td>172</td>
<td>115</td>
<td>179</td>
<td>142</td>
</tr>
<tr>
<td>1372</td>
<td>142</td>
<td>203</td>
<td>139</td>
<td>209</td>
<td>174</td>
</tr>
<tr>
<td>1582</td>
<td>145</td>
<td>203</td>
<td>140</td>
<td>214</td>
<td>177</td>
</tr>
</tbody>
</table>
Orange Tree Data

- All growth curves have a similar shape
- However, the growth rate of each curve is different
Orange Tree Data

- A logistic growth model

\[ y_{it} = \frac{\xi_i}{1 + \exp[-(x_{it} - \beta_1)/\beta_2]} + \epsilon_{it}, \]

where

- \( y_{it} = \) circumference, \( i = 1, \ldots, 5, t = 1, \ldots, 7 \)
- \( x_{it} = \) days, \( i = 1, \ldots, 5, t = 1, \ldots, 7 \)
- \( \xi_i \) is a random parameter: \( \xi_i = \varphi + \delta_i \)
- \( \varphi \) is the fixed effect
- \( \delta_i \) is random effect, usually assumed \( \delta_i \sim N(0, \sigma_\delta^2) \)
- \( \epsilon_{it} \sim N(0, \sigma_\epsilon^2) \) are i.i.d. random errors
Pharmacokinetics of cefamandole (Davidian and Giltinan 1995): A dose of 15 mg/kg body weight is administered by ten-minute intravenous infusion to six healthy male volunteers, and plasma concentration is measured at 14 time points.

<table>
<thead>
<tr>
<th>Time</th>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Subject 3</th>
<th>Subject 4</th>
<th>Subject 5</th>
<th>Subject 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>127.00</td>
<td>120.00</td>
<td>154.00</td>
<td>181.00</td>
<td>253.00</td>
<td>140.00</td>
</tr>
<tr>
<td>15</td>
<td>80.00</td>
<td>90.10</td>
<td>94.00</td>
<td>119.00</td>
<td>176.00</td>
<td>120.00</td>
</tr>
<tr>
<td>20</td>
<td>47.40</td>
<td>70.00</td>
<td>84.00</td>
<td>84.30</td>
<td>150.00</td>
<td>106.00</td>
</tr>
<tr>
<td>30</td>
<td>39.90</td>
<td>40.10</td>
<td>56.00</td>
<td>56.10</td>
<td>90.30</td>
<td>60.40</td>
</tr>
<tr>
<td>45</td>
<td>24.80</td>
<td>24.00</td>
<td>37.10</td>
<td>39.80</td>
<td>69.60</td>
<td>60.90</td>
</tr>
<tr>
<td>60</td>
<td>17.90</td>
<td>16.10</td>
<td>28.90</td>
<td>23.30</td>
<td>42.50</td>
<td>42.20</td>
</tr>
<tr>
<td>75</td>
<td>11.70</td>
<td>11.60</td>
<td>25.50</td>
<td>22.70</td>
<td>30.60</td>
<td>26.80</td>
</tr>
<tr>
<td>90</td>
<td>10.90</td>
<td>9.20</td>
<td>20.00</td>
<td>13.00</td>
<td>19.60</td>
<td>22.00</td>
</tr>
<tr>
<td>120</td>
<td>5.70</td>
<td>5.20</td>
<td>12.40</td>
<td>8.00</td>
<td>13.80</td>
<td>14.50</td>
</tr>
<tr>
<td>150</td>
<td>2.55</td>
<td>3.00</td>
<td>8.30</td>
<td>2.40</td>
<td>11.40</td>
<td>8.80</td>
</tr>
<tr>
<td>180</td>
<td>1.84</td>
<td>1.54</td>
<td>4.50</td>
<td>1.60</td>
<td>6.30</td>
<td>6.00</td>
</tr>
<tr>
<td>240</td>
<td>1.50</td>
<td>0.73</td>
<td>3.40</td>
<td>1.10</td>
<td>3.80</td>
<td>3.00</td>
</tr>
<tr>
<td>300</td>
<td>0.70</td>
<td>0.37</td>
<td>1.70</td>
<td>0.48</td>
<td>1.55</td>
<td>1.30</td>
</tr>
<tr>
<td>360</td>
<td>0.34</td>
<td>0.19</td>
<td>1.19</td>
<td>0.29</td>
<td>1.22</td>
<td>1.03</td>
</tr>
</tbody>
</table>
An exponential model with two random effects

\[ y_{it} = \xi_{1i} \exp(-\xi_{2i} x_{it}) + \varepsilon_{it} \]

\[ \xi_{1i} = \varphi_1 + \delta_{1i}, \quad \xi_{2i} = \varphi_2 + \delta_{2i} \]
A General Nonlinear Mixed Effects Model

- The model

\[ y_{it} = g(x_{it}, \xi_i, \beta) + \varepsilon_{it}, \quad t = 1, 2, \ldots, T_i \]

\[ \xi_i = Z_i \varphi + \delta_i, \quad i = 1, 2, \ldots, n, \]

where

\[ y_{it} \in \mathbb{R}, \quad x_{it} \in \mathbb{R}^k, \quad \xi_i \in \mathbb{R}^\ell, \quad \beta \in \mathbb{R}^p, \quad \varphi \in \mathbb{R}^q \]

\[ \delta_i \sim f_\delta(u; \psi), \quad \psi \in \mathbb{R}^r, \text{ independent of } Z_i \text{ and } X_i = (x_{i1}, x_{i2}, \ldots, x_{iT_i})' \]

\[ \varepsilon_{it} \text{ are i.i.d. and } E(\varepsilon_{it}|X_i, Z_i, \delta_i) = 0, \quad E(\varepsilon_{it}^2|X_i, Z_i, \delta_i) = \sigma_{\varepsilon}^2 \]

- The goal is to estimate \( \gamma = (\beta, \varphi, \psi, \sigma_{\varepsilon}^2) \)
Maximum likelihood estimation:

Generalized method of moments (GMM) estimation for linear (and some nonlinear) dynamic models:

In general, the maximum likelihood estimators are difficult to compute and existing approximation methods rely on normality assumption (Hartford and Davidian (2000)).
**Exponential model**

\[
y_{it} = \xi_1 i \exp(-\xi_2 i x_{it}) + \varepsilon_{it}
\]

\[
\xi_i = \varphi + \delta_i, \delta_i \sim N_2 [(0,0), \text{diag}(\psi_1, \psi_2)]
\]

The first two moments of \(y_{it}\) given \(X_i\) are

\[
E(y_{it}|X_i) = \varphi_1 \exp(-\varphi_2 x_{it} + \psi_2 x_{it}^2/2)
\]

\[
E(y_{it}y_{is}|X_i) = (\varphi_1^2 + \psi_1) \exp \left[-\varphi_2(x_{it} + x_{is}) + \psi_2(x_{it} + x_{is})^2/2 \right]
\]

\[+\sigma_{its}\]

where \(\sigma_{its} = \sigma^2_{\varepsilon}\) if \(t = s\), and zero otherwise.

\(\varphi_1, \varphi_2\) and \(\psi_2\) are identified by the first equation and the nonlinear least squares method, while \(\psi_1\) and \(\sigma^2_{\varepsilon}\) are identified by the second equation.
The first two conditional moments:

\[
\mu_{it}(\gamma) = E_{\gamma}(y_{it}|X_i, Z_i) = \int g(x_{it}, u, \beta) f_\delta(u - Z_i\varphi; \psi) du,
\]

\[
\nu_{its}(\gamma) = E_{\gamma}(y_{it}y_{is}|X_i, Z_i)
\]

\[
= \int g(x_{it}, u, \beta)g(x_{is}, u, \beta)f_\delta(u - Z_i\varphi; \psi)du + \sigma_{its},
\]

where \(\sigma_{its} = \sigma_\varepsilon^2\) if \(t = s\), and zero otherwise.

The SLSE for \(\gamma\) is \(\hat{\gamma}_N = \arg\min_{\gamma} Q_N(\gamma)\), where

\[
Q_n(\gamma) = \sum_{i=1}^n \rho_i'(\gamma)A_i\rho_i(\gamma),
\]

\[
\rho_i(\gamma) = (y_{it} - \mu_{it}(\gamma), y_{it}y_{is} - \nu_{its}(\gamma), 1 \leq t \leq s \leq T_i)' \]

and \(A_i\) is n.d. and may depend on \(X_i, Z_i\).
Example: Exponential Model

- The model

\[ y_{it} = \xi_1 i \exp(-\xi_2 i x_{it}) + \varepsilon_{it}, \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2) \]
\[ \xi_i = \varphi + \delta_i, \delta_i \sim N_2 [(0, 0), (\psi_1, \psi_2, \psi_{12})] \]
\[ x_{it} = x_t \sim \text{Unif}(0, 5) \]

- First two moments

\[ \mu_{it}(\gamma) = (\varphi_1 - \psi_{12} x_{it}) \exp(-\varphi_2 x_{it} + \psi_2 x_{it}^2/2) \]
\[ \nu_{its}(\gamma) = \left[ \psi_1 + (\varphi_1 - \psi_{12} (x_{it} + x_{is}))^2 \right] \times \exp \left[ -\varphi_2 (x_{it} + x_{is}) + \psi_2 (x_{it} + x_{is})^2/2 \right] + \sigma_{its}, \]

where \( \sigma_{its} = \sigma_\varepsilon^2 \) if \( t = s \), and zero otherwise.

- Also compute quasilikelihood estimates for \( \varphi_1, \varphi_2, \psi_2, \psi_{12} \) assuming \( \psi_1, \sigma_\varepsilon^2 \) are known.

- Monte Carlo replications: 1000
Simulation 1: Exponential model with \( n = 20, \ T = 5 \)

<table>
<thead>
<tr>
<th>( \varphi_1 = 10 )</th>
<th>( \varphi_2 = 5 )</th>
<th>( \psi_1 = 1 )</th>
<th>( \psi_2 = 0.7 )</th>
<th>( \psi_{12} = 0.5 )</th>
<th>( \sigma^2 = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS1</td>
<td>9.9024</td>
<td>4.9369</td>
<td>1.0032</td>
<td>0.6803</td>
<td>0.5003</td>
</tr>
<tr>
<td>SSE</td>
<td>0.0499</td>
<td>0.0229</td>
<td>0.0092</td>
<td>0.0055</td>
<td>0.0055</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.5816</td>
<td>0.7264</td>
<td>0.2915</td>
<td>0.1749</td>
<td>0.1733</td>
</tr>
<tr>
<td>SLS2</td>
<td>9.8597</td>
<td>4.9365</td>
<td>0.9940</td>
<td>0.6913</td>
<td>0.5012</td>
</tr>
<tr>
<td>SSE</td>
<td>0.0442</td>
<td>0.0214</td>
<td>0.0092</td>
<td>0.0056</td>
<td>0.0055</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.4030</td>
<td>0.6785</td>
<td>0.2919</td>
<td>0.1768</td>
<td>0.1734</td>
</tr>
<tr>
<td>QLE</td>
<td>11.2574</td>
<td>5.4979</td>
<td>-</td>
<td>0.6056</td>
<td>0.4935</td>
</tr>
<tr>
<td>SSE</td>
<td>0.0333</td>
<td>0.0186</td>
<td>-</td>
<td>0.0051</td>
<td>0.0055</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.6392</td>
<td>0.7707</td>
<td>-</td>
<td>0.1868</td>
<td>0.1743</td>
</tr>
</tbody>
</table>

**SLS1:** SLSE using identity weight  
**SLS2:** SLSE using optimal weight  
**QLE:** Quasilikelihood estimates  
**SSE:** Monte Carlo simulation standard error  
**RMSE:** Root mean squared error
Simulation 2: Exponential model with $n = 40, T = 7$

<table>
<thead>
<tr>
<th></th>
<th>$\varphi_1 = 10$</th>
<th>$\varphi_2 = 5$</th>
<th>$\psi_1 = 1$</th>
<th>$\psi_2 = 0.7$</th>
<th>$\psi_{12} = 0.5$</th>
<th>$\sigma^2 = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS1</td>
<td>9.9178</td>
<td>4.8742</td>
<td>0.9959</td>
<td>0.6454</td>
<td>0.5104</td>
<td>0.9915</td>
</tr>
<tr>
<td>SSE</td>
<td>0.0475</td>
<td>0.0310</td>
<td>0.0089</td>
<td>0.0048</td>
<td>0.0055</td>
<td>0.0034</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.5029</td>
<td>0.9888</td>
<td>0.2804</td>
<td>0.1614</td>
<td>0.1732</td>
<td>0.1073</td>
</tr>
<tr>
<td>SLS2</td>
<td>9.9049</td>
<td>4.8969</td>
<td>0.9971</td>
<td>0.6572</td>
<td>0.5055</td>
<td>0.9332</td>
</tr>
<tr>
<td>SSE</td>
<td>0.0391</td>
<td>0.0264</td>
<td>0.0091</td>
<td>0.0052</td>
<td>0.0054</td>
<td>0.0034</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.2404</td>
<td>0.8406</td>
<td>0.2870</td>
<td>0.1691</td>
<td>0.1709</td>
<td>0.1269</td>
</tr>
<tr>
<td>QLE</td>
<td>11.4357</td>
<td>5.8306</td>
<td>-</td>
<td>0.6335</td>
<td>0.4920</td>
<td>-</td>
</tr>
<tr>
<td>SSE</td>
<td>0.0184</td>
<td>0.0129</td>
<td>-</td>
<td>0.0052</td>
<td>0.0055</td>
<td>-</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.5491</td>
<td>0.9246</td>
<td>-</td>
<td>0.1759</td>
<td>0.1739</td>
<td>-</td>
</tr>
</tbody>
</table>

SLS1:  SLSE using identity weight
SLS2:  SLSE using optimal weight
QLE:   Quasilikelihood estimates
SSE:   Monte Carlo simulation standard error
RMSE:  Root mean squared error
Example: Logistic model

- The model

\[
y_{it} = \frac{\xi_i}{1 + \exp[-(x_{it} - \beta_1)/\beta_2]} + \epsilon_{it}, \epsilon_{it} \sim N(0, \sigma_{\epsilon}^2)
\]

\[
\xi_i = \varphi + \delta_i, \delta_i \sim N(0, \psi)
\]

\[
x_{it} = x_t = (20, 40, ..., 20T)
\]

- First two moments

\[
\mu_{it}(\gamma) = \frac{\varphi}{1 + \exp[(\beta_1 - x_{it})/\beta_2]}
\]

\[
\nu_{its}(\gamma) = \frac{\varphi^2 + \psi}{(1 + \exp[(\beta_1 - x_{it})/\beta_2])(1 + \exp[(\beta_1 - x_{is})/\beta_2])} + \sigma_{its}
\]

- Compute SLS using identity weight.

- Monte Carlo replications: 500
Choose $h(u)$ to be the density of $N(0, \sigma_0^2)$ with $\sigma_0^2 = 5$

Generate $u_{ij} \sim h(u)$ with $S = 1000$.

Compute

$$
\mu_{it,1}(\gamma) = \frac{1}{S} \sum_{j=1}^{S} u_{ij} \sqrt{\sigma_0^2/\psi} \exp \left[ -(u_{ij} - \varphi)^2/2\psi + u_{ij}^2/2\sigma_0^2 \right]
\quad \frac{1 + \exp[(\beta_1 - x_{it})/\beta_2]}{},
$$

$$
\nu_{its,1}(\gamma) = \frac{1}{S} \sum_{j=1}^{S} u_{ij}^2 \sqrt{\sigma_0^2/\psi} \exp \left[ -(u_{ij} - \varphi)^2/2\psi + u_{ij}^2/2\sigma_0^2 \right]
\quad \frac{1 + \exp[(\beta_1 - x_{it})/\beta_2]}{(1 + \exp[(\beta_1 - x_{it})/\beta_2])(1 + \exp[(\beta_1 - x_{is})/\beta_2])}
\quad + \sigma_{its}
$$

and $\mu_{it,2}(\gamma), \nu_{its,2}(\gamma)$ similarly using $u_{ij}, j = S + 1, \ldots, 2S$.

Compute SBE using identity weight.
Simulation 3: Logistic model with $n = 7, T = 5$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1 = 70$</th>
<th>$\beta_2 = 34$</th>
<th>$\varphi = 20$</th>
<th>$\psi = 9$</th>
<th>$\sigma^2_\varepsilon = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS</td>
<td>69.9058</td>
<td>34.0463</td>
<td>19.8818</td>
<td>9.0167</td>
<td>1.0140</td>
</tr>
<tr>
<td></td>
<td>(0.0720)</td>
<td>(0.0592)</td>
<td>(0.0510)</td>
<td>(0.0142)</td>
<td>(0.0215)</td>
</tr>
<tr>
<td>SBE</td>
<td>69.9746</td>
<td>34.1314</td>
<td>18.9744</td>
<td>10.7648</td>
<td>0.9921</td>
</tr>
<tr>
<td></td>
<td>(0.1143)</td>
<td>(0.1159)</td>
<td>(0.1137)</td>
<td>(0.0216)</td>
<td>(0.0607)</td>
</tr>
</tbody>
</table>

SLS:     SLSE using identity weight  
SBE:     Simulation-based estimates using identity weight  
(  ):     Simulation standard errors

Simulation 4: Logistic model with $n = 30, T = 10$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1 = 70$</th>
<th>$\beta_2 = 34$</th>
<th>$\varphi = 20$</th>
<th>$\psi = 9$</th>
<th>$\sigma^2_\varepsilon = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS</td>
<td>70.0203</td>
<td>34.0303</td>
<td>20.0319</td>
<td>8.9625</td>
<td>1.0016</td>
</tr>
<tr>
<td></td>
<td>(0.0398)</td>
<td>(0.0395)</td>
<td>(0.0258)</td>
<td>(0.0128)</td>
<td>(0.0249)</td>
</tr>
<tr>
<td>SBE</td>
<td>69.9754</td>
<td>34.2096</td>
<td>19.1365</td>
<td>10.8034</td>
<td>0.8936</td>
</tr>
<tr>
<td></td>
<td>(0.1183)</td>
<td>(0.1146)</td>
<td>(0.1094)</td>
<td>(0.0180)</td>
<td>(0.0537)</td>
</tr>
</tbody>
</table>
Example: Logistic model with 2 random effects

The model

\[
y_{it} = \frac{\xi_{1i}}{1 + \exp[-(x_{it} - \xi_{2i})/\beta]} + \varepsilon_{it}, \varepsilon_{it} \sim N(0, \sigma_{\varepsilon}^2)
\]

\[
\xi_i = \varphi + \delta_i, \delta_i \sim N_2[(0, 0), \text{diag}(\psi_1, \psi_2)]
\]

The closed forms of the moments are not available.

Generate \( S = 500 \) points \( \{u_{ij}\} \sim N_2[(200, 700), \text{diag}(81, 81)] \).

Monte Carlo replications: 500

Simulation 5: Sample sizes \( n = 7, T = 5 \).

<table>
<thead>
<tr>
<th>True</th>
<th>( \beta = 350 )</th>
<th>( \varphi_1 = 200 )</th>
<th>( \varphi_2 = 700 )</th>
<th>( \psi_1 = 100 )</th>
<th>( \psi_2 = 625 )</th>
<th>( \sigma_{\varepsilon}^2 = 25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBE</td>
<td>349.8222</td>
<td>199.3850</td>
<td>699.3057</td>
<td>104.8866</td>
<td>634.3594</td>
<td>25.3303</td>
</tr>
<tr>
<td></td>
<td>(0.5896)</td>
<td>(0.5984)</td>
<td>(0.5620)</td>
<td>(0.0088)</td>
<td>(0.0533)</td>
<td>(0.2605)</td>
</tr>
</tbody>
</table>

SBE: Simulation-based estimates using identity weight

(·): Simulation standard errors