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

Second-order Least Squares Method

- 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 σ^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 γ is defined as $\hat{\gamma}_{SLS} = \mathrm{argmin}_{\gamma} Q_n(\gamma)$, where

$$\begin{aligned} Q_n(\gamma) &= \sum_{i=1}^{n} \rho_i'(\gamma) A_i \rho_i(\gamma), \\ \rho_i(\gamma) &= (Y_i - g(X_i;\beta), Y_i^2 - g^2(X_i;\beta) - \sigma^2)' \end{aligned}$$

  and $A_i = A(X_i)$ is a 2×2 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 β a.e.
2. $E\left\|A(X)\right\|\left(\sup_\beta g^4(X;\beta)+1\right)<\infty$.
3. The parameter space $\Gamma \subset \boldsymbol{R}^{p+1}$ 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. β and $E\left\|A(X)\right\|\sup_\beta\left(\left\|\frac{\partial g(X;\beta)}{\partial\beta}\right\|^4+\left\|\frac{\partial^2 g(X;\beta)}{\partial\beta\partial\beta'}\right\|^4\right)<\infty$.
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

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

  $\mu_3 = E(\varepsilon^3|X)$ and $\mu_4 = E(\varepsilon^4|X)$.
- Since A_0 depends on γ, 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

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

where

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

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

$$G_1 = E\left[\frac{\partial g(X;\beta)}{\partial \beta}\right], \quad G_2 = E\left[\frac{\partial g(X;\beta)}{\partial \beta}\frac{\partial g(X;\beta)}{\partial \beta'}\right].$$

SLS and OLS Estimators

- 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 $X \sim \mathit{Uniform}(0, 20)$ 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

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

SLS and Generalized Method of Moments Estimator

- 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 $V(\hat{\sigma}^2_{GMM}) \geq V(\hat{\sigma}^2_{SLS})$.

Simple Linear Regression

- The relationship of interest: $Y = \beta_0 + \beta_x X + \varepsilon$, where Y: response variable, X: explanatory variable, ε: is uncorrelated with X and $E(\varepsilon) = 0$.
- Given an $i.i.d.$ random sample $(X_i, Y_i), i = 1, 2, ..., n$
- 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:

$$E(Y|X) = g(\beta_0 + \beta_x X, ...),$$

  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:

$$E(Y|X) = g(\beta_0 + \beta_x' X, ...),$$

  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

Example of Measurement Error

- 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

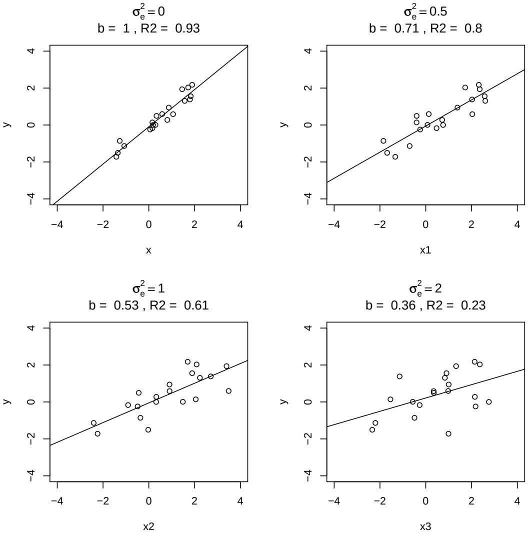
Examples of Measurement Error

- 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.

Impact of Measurement Error: A simulation study

- Generate independent $X_i \sim UNIF(-2, 2), i = 1, 2, ..., n = 20$
- 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 $\sigma_e^2 = 1, 2$ respectively

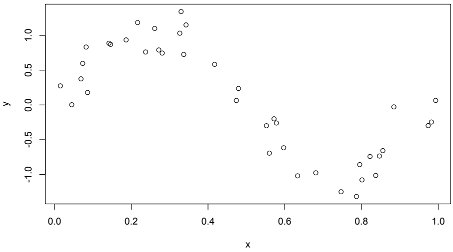
Impact of Measurement Error: A simulation study

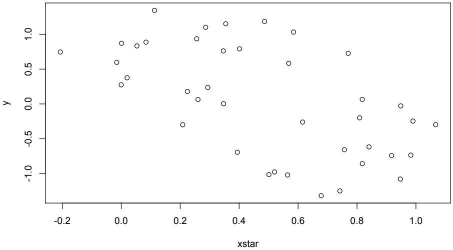


Impact of Measurement Error: A simulation study

- Generate independent $X_i \sim UNIF(0,1), i = 1, 2, ..., n = 40$
- 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





Impact of Measurement Error: Theory

- 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

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

- 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: $\hat{\sigma}^{2*} \xrightarrow{P} \sigma^{2*} = \sigma^2 + \lambda \beta_x^2 \sigma_e^2$

Identifiability of Normal Linear Model

- 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

$$
\begin{aligned}
E(Y) &= \beta_0 + \beta_x \mu_x, \quad E(Z) = \mu_x \\
Var(Y) &= \beta_x^2 \sigma_x^2 + \sigma_\varepsilon^2, \quad Cov(Y, Z) = \beta_x \sigma_x^2 \\
Var(Z) &= \sigma_x^2 + \sigma_e^2
\end{aligned}
$$

- There are 5 moment equations but 6 unknown parameters.
- In practice usually *ad hoc* restrictions are imposed to ensure unique solution (e.g. σ_e^2, $\sigma_e^2/\sigma_\varepsilon^2$ or σ_e^2/σ_x^2 is known or can be estimated using extra data).

Nonlinear Measurement Error Model

- The response model: $Y = g(X, \beta) + \varepsilon$, where
  Y: the response variable; X: unobserved predictor (vector);
  ε: 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 $Var(Z) > Var(X)$. E.g. blood pressure.
- Berkson ME: $X = Z + e$, e independent of Z and therefore $Var(X) > Var(Z)$. 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
  - Wang (2003, 2004): general nonlinear 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 $g \in L_1(\boldsymbol{R}^k)$
  - Schennach (2007): $|g|$ is univariate and bounded by polynomials in $\boldsymbol{R}$.

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

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

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

Estimation in Nonlinear ME Models

- The IV method assuming that ME variance shrinks to zero as sample size tends to infinity: Wolter and Fuller (1982), Amemiya (1985, 1990), Stefanski and Carroll (1985), Amemiya and Fuller (1988)
- Assume the conditional density $f(x|z)$ has known parametric form: Hsiao (1989, 1992), Li and Hsiao (2004)
- Univariate polynomial model with IV: Hausman et al (1991), Hausman, Newey and Powell (1995), Cheng and Schneeweiss (1998), Huang and Huwang (2001)
- Nonlinear model with replicate data: Li (2002), Schennach (2004)
- Nonlinear semiparametric model with IV: Wang and Hsiao (1995, 2007), Schennach (2007)

Estimation in Nonlinear EIV Models

- Approximate estimation when ME are small:
  - regression calibration: Carroll and Stefanski (1990), Gleser (1990), Rosner, Willett and Spielgelman (1990)
  - simulation-extrapolation (SIMEX): Cook and Stefanski (1994), Stefanski and Cook (1995), Carroll et al (1996)
- Estimation in Berkson ME models:
  - logistic model: Rudemo, Ruppert and Streibig (1989)
  - univariate polynomial model: Huwang and Huang (2000)
  - general nonlinear models: Wang (2003, 2004)

Identifiability of Berkson ME Model: an Example

- A quadratic model with Berkson ME

$$
\begin{aligned}
Y &= \beta_0 + \beta_1 X^2 + \varepsilon, \varepsilon \sim N(0, \sigma^2) \\
X &= Z + e, e \sim N(0, \sigma_e^2)
\end{aligned}
$$

- The first two conditional moments

$$
\begin{aligned}
E(Y|Z) &= \beta_0 + \beta_1 \sigma_e^2 + \beta_1 Z^2 \\
E(Y^2|Z) &= \sigma^2 + \left(\beta_0 + \beta_1 \sigma_e^2\right)^2 + 2\beta_1^2 \sigma_e^4 \\
&\quad + 2\beta_1 \left(\beta_0 + 3\beta_1 \sigma_e^2\right) Z^2 + \beta_1^2 Z^4
\end{aligned}
$$

- 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, ε 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 $\hat{\gamma}_n = \mathrm{argmin}_\gamma\, Q_n(\gamma)$, where $Q_n(\gamma) = \sum_{i=1}^n \rho_i'(\gamma) A_i \rho_i(\gamma)$,

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

  and $A_i = W(Z_i)$ is a 2×2 weighting matrix.
- Under some regularity conditions, as $n \to \infty$, we have $\hat{\gamma}_n \xrightarrow{a.s.} \gamma$ and $\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{L} N(0, B^{-1} C B^{-1})$, where

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

SLS Estimation in Berkson ME Models: an Example

- A quadratic model

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i, \\
X_i &= Z_i + e_i,
\end{aligned}
$$

  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

	$\beta_0 = 3$	$\beta_1 = 2$	$\beta_2 = 1$	$\sigma^2 = 1$	$\sigma_e^2 = 2$
SLS1	3.046	2.052	0.995	0.983	2.028
(Std.)	(0.013)	(0.014)	(0.010)	(0.019)	(0.013)
SLS2	3.024	2.048	0.975	1.073	2.026
(Std.)	(0.013)	(0.013)	(0.010)	(0.020)	(0.012)
NLS	5.025	1.929	1.024	88.356	NA
(Std.)	(0.064)	(0.087)	(0.026)	(0.622)	NA

SLS1: SLSE using identity weight

SLS2: SLSE using optimal weight

NLS: Naive nonlinear least squares estimates ignoring ME

Simulation-based SLS Estimator

- In general the first two moments are

$$
\begin{aligned}
E\left(Y_i|Z_i,\gamma\right) &= \int g\left(Z+u,\beta\right) f_e(u;\psi)du \\
E\left(Y_i^2|Z_i,\gamma\right) &= \int g^2\left(Z+u,\beta\right) f_e(u;\psi)du + \sigma^2
\end{aligned}
$$

- 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 $h(u)$.

Simulation-based SLS Estimator

- Choose a known density $h(u)$ such that $Supp(h) \supseteq Supp(f_e(u;\psi))$.
- Generate random points $u_{ij} \sim h(u)$, $i = 1, ..., n, j = 1, ..., 2S$ and calculate $\rho_{i,1}(\gamma)$ using $u_{ij}, j = 1, 2, ..., S$ and $\rho_{i,2}(\gamma)$ using $u_{ij}, j = S+1, S+2, ..., 2S$
- 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 $\hat{\gamma}_{n,S} = \mathrm{argmin}_\gamma\, Q_{n,S}(\gamma)$.

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

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

- 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 $O(1/S)$.

Simulation-based SLSE: an Example

- A linear-exponential model

$$\begin{aligned} Y_i &= \beta_1 X_{1i} + \beta_2 \exp(-\beta_3 X_{2i}) + \varepsilon_i, \\ X_{1i} &= Z_{1i} + e_{1i}, X_{2i} = Z_{2i} + e_{2i}, \end{aligned}$$

  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 $N_2(0, diag(5, 5))$ and $S = 1000$.
- Sample size $n = 100$, and Monte Carlo replications $N = 1000$.

Example: Linear-Exponential Model

	$\beta_1 = 3$	$\beta_2 = 2$	$\beta_3 = 1$	$\sigma^2 = 1$	$\sigma_1^2 = 1$	$\sigma_2^2 = 1.5$
SLS1	3.000	2.009	0.878	1.023	1.073	1.356
(Std.)	(0.011)	(0.008)	(0.004)	(0.009)	(0.011)	(0.007)
SLS2	2.987	2.066	0.869	1.026	1.039	1.275
(Std.)	(0.009)	(0.009)	(0.003)	(0.009)	(0.010)	(0.005)
SbSLS	3.002	1.898	0.947	1.000	1.003	1.319
(Std.)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.008)
NLS	3.215	2.391	1.017	45.557	NA	NA
(Std.)	(0.008)	(0.007)	(0.006)	(3.365)	NA	NA

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:

$$
\begin{aligned}
Y &= g(X, \beta) + \varepsilon \\
Z &= X + e \\
X &= \Gamma W + U
\end{aligned}
$$

- $Y \in \boldsymbol{R}, Z \in \boldsymbol{R}^k, W \in \boldsymbol{R}^\ell$ are observed;
- $X \in \boldsymbol{R}^k, \beta \in \boldsymbol{R}^p, \Gamma \in \boldsymbol{R}^{k \times \ell}$ are unobserved;
- $E(\varepsilon \mid X, Z, W) = 0$ and $E(e \mid X, W) = 0$;
- U and W independent and $E(U) = 0$;
- Suppose $U \sim f_U(u; \phi)$ which is known up to $\phi \in \boldsymbol{R}^q$.
- X, ε and e have nonparametric distributions.

SLS-IV Estimation for Classical ME Models

- Under model assumptions:

$$\begin{aligned} E(Z \mid W) &= \Gamma W & (1) \\ E(Y \mid W) &= \int g(x;\beta) f_U(x - \Gamma W;\phi)dx & (2) \\ E(YZ \mid W) &= \int x g(x;\beta) f_U(x - \Gamma W;\phi)dx & (3) \\ E(Y^2 \mid W) &= \int g^2(x;\beta) f_U(x - \Gamma W;\phi)dx + \sigma_\varepsilon^2 & (4) \end{aligned}$$

- Γ 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 $\hat{\gamma}_n = \mathrm{argmin}_\psi \sum_{i=1}^n \rho_i'(\gamma) A_i \rho_i(\gamma)$, where

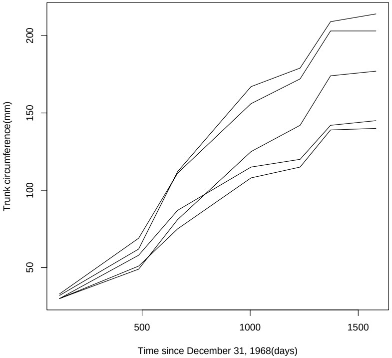
$$\rho_i(\gamma) = (Y_i - E(Y_i|W_i,\gamma), Y_i^2 - E(Y_i^2|W_i,\gamma), Y_i Z_i - E(Y_i Z_i|W_i,\gamma))'.$$

Example of Longitudinal Data

- 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.

	Trunk Circumference				
Day	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5
118	30	33	30	32	30
484	58	69	51	62	49
664	87	111	75	112	81
1004	115	156	108	167	125
1231	120	172	115	179	142
1372	142	203	139	209	174
1582	145	203	140	214	177

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

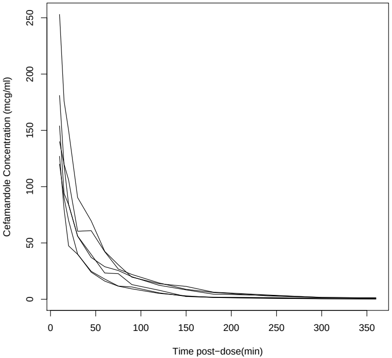
- $y_{it} =$ circumference $, i = 1, ..., 5, t = 1, ..., 7$
- $x_{it} =$ days $, i = 1, ..., 5, t = 1, ..., 7$
- ξ_i is a random parameter: $\xi_i = \varphi + \delta_i$
- φ is the fixed effect
- δ_i is random effect, usually assumed $\delta_i \sim N(0, \sigma_\delta^2)$
- $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ are $i.i.d.$ random errors

Example of Longitudinal Data

- 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.

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

Cefamandole Plasma Concentration



- An exponential model with two random effects

$$
\begin{aligned}
y_{it} &= \xi_{1i} \exp\left(-\xi_{2i} x_{it}\right) + \varepsilon_{it} \\
\xi_{1i} &= \varphi_1 + \delta_{1i}, \quad \xi_{2i} = \varphi_2 + \delta_{2i}
\end{aligned}
$$

A General Nonlinear Mixed Effects Model

- The model

$$\begin{aligned} y_{it} &= g(x_{it}, \xi_i, \beta) + \varepsilon_{it}, t = 1, 2, ..., T_i \\ \xi_i &= Z_i \varphi + \delta_i, i = 1, 2, ..., n, \end{aligned}$$

  where

- $y_{it} \in \boldsymbol{R}$, $x_{it} \in \boldsymbol{R}^k$, $\xi_i \in \boldsymbol{R}^\ell$, $\beta \in \boldsymbol{R}^p$, $\varphi \in \boldsymbol{R}^q$
- $\delta_i \sim f_\delta(u; \psi), \psi \in \boldsymbol{R}^r$, independent of Z_i and $X_i = (x_{i1}, x_{i2}, ..., x_{iT_i})'$
- ε_{it} are $i.i.d.$ and $E(\varepsilon_{it}|X_i, Z_i, \delta_i) = 0$, $E(\varepsilon_{it}^2|X_i, Z_i, \delta_i) = \sigma_\varepsilon^2$
- The goal is to estimate $\gamma = (\beta, \varphi, \psi, \sigma_\varepsilon^2)$

Estimation in Mixed Effects Models

- Maximum likelihood estimation:
  Lindstrom and Bates (1990), Davidian and Gallant (1993), Ke and Wang (2001), Vonesh *et al.* (2002), Wu (2002), Daimon and Goto (2003), Lai and Shih (2003)
- Generalized method of moments (GMM) estimation for linear (and some nonlinear) dynamic models:
  Wooldridge (1999), Arellano and Honoré (2001), Hsiao, Pesaran and Tahmiscioglu (2002), Arellano and Carrasco (2003), Honoré and Hu (2004)
- In general, the maximum likelihood estimators are difficult to compute and existing approximation methods rely on normality assumption (Hartford and Davidian (2000)).

Identification Using First Two Moments: An Example

- Exponential model

$$
\begin{aligned}
y_{it} &= \xi_{1i} \exp\left(-\xi_{2i} x_{it}\right) + \varepsilon_{it} \\
\xi_i &= \varphi + \delta_i, \delta_i \sim N_2\left[(0,0), \mathrm{diag}(\psi_1, \psi_2)\right]
\end{aligned}
$$

- The first two moments of y_{it} given X_i are

$$
\begin{aligned}
E(y_{it}|X_i) &= \varphi_1 \exp\left(-\varphi_2 x_{it} + \psi_2 x_{it}^2/2\right) \\
E(y_{it} y_{is}|X_i) &= (\varphi_1^2 + \psi_1) \exp\left[-\varphi_2(x_{it} + x_{is}) + \psi_2(x_{it} + x_{is})^2/2\right] \\
&\quad + \sigma_{its}
\end{aligned}
$$

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

- φ_1, φ_2 and ψ_2 are identified by the first equation and the nonlinear least squares method, while ψ_1 and σ_ε^2 are identified by the second equation.

Second-order Least Squares Estimator

- The first two conditional moments:

$$\begin{aligned}
\mu_{it}(\gamma) &= E_\gamma(y_{it}|X_i, Z_i) = \int g\left(x_{it}, u, \beta\right) f_\delta(u - Z_i\varphi; \psi)du, \\
\nu_{its}(\gamma) &= E_\gamma(y_{it}y_{is}|X_i, Z_i) \\
&= \int g(x_{it}, u, \beta)g(x_{is}, u, \beta)f_\delta(u - Z_i\varphi; \psi)du + \sigma_{its},
\end{aligned}$$

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

- The SLSE for γ is $\hat{\gamma}_N = \mathrm{argmin}_\gamma\, Q_N(\gamma)$, where $Q_n(\gamma) = \sum_{i=1}^n \rho_i'(\gamma)A_i\rho_i(\gamma)$,

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

  and A_i is n.d. and may depend on X_i, Z_i.

Example: Exponential Model

- The model

$$
\begin{aligned}
y_{it} &= \xi_{1i} \exp\left(-\xi_{2i} x_{it}\right) + \varepsilon_{it}, \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2) \\
\xi_i &= \varphi + \delta_i, \delta_i \sim N_2\left[(0,0), (\psi_1, \psi_2, \psi_{12})\right] \\
x_{it} &= x_t \sim \text{Unif}(0,5)
\end{aligned}
$$

- First two moments

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

  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 $n = 20$, $T = 5$

	$\varphi_1 = 10$	$\varphi_2 = 5$	$\psi_1 = 1$	$\psi_2 = 0.7$	$\psi_{12} = 0.5$	$\sigma^2 = 1$
SLS1	9.9024	4.9369	1.0032	0.6803	0.5003	0.9827
SSE	0.0499	0.0229	0.0092	0.0055	0.0055	0.0051
RMSE	1.5816	0.7264	0.2915	0.1749	0.1733	0.1612
SLS2	9.8597	4.9365	0.9940	0.6913	0.5012	0.9395
SSE	0.0442	0.0214	0.0092	0.0056	0.0055	0.0051
RMSE	1.4030	0.6785	0.2919	0.1768	0.1734	0.1722
QLE	11.2574	5.4979	-	0.6056	0.4935	-
SSE	0.0333	0.0186	-	0.0051	0.0055	-
RMSE	1.6392	0.7707	-	0.1868	0.1743	-

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 $n = 40$, $T = 7$

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

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

$$
\begin{aligned}
y_{it} &= \frac{\xi_i}{1+\exp[-(x_{it}-\beta_1)/\beta_2]} + \varepsilon_{it}, \varepsilon_{it} \sim N(0,\sigma_\varepsilon^2) \\
\xi_i &= \varphi + \delta_i, \delta_i \sim N(0,\psi) \\
x_{it} &= x_t = (20, 40, ..., 20T)
\end{aligned}
$$

- First two moments

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

- Compute SLS using identity weight.
- Monte Carlo replications: 500

Example: Logistic model (Cont'd)

- 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

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

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

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

Simulation 3: Logistic model with $n = 7$, $T = 5$.

	$\beta_1 = 70$	$\beta_2 = 34$	$\varphi = 20$	$\psi = 9$	$\sigma_\varepsilon^2 = 1$
SLS	69.9058	34.0463	19.8818	9.0167	1.0140
	(0.0720)	(0.0592)	(0.0510)	(0.0142)	(0.0215)
SBE	69.9746	34.1314	18.9744	10.7648	0.9921
	(0.1143)	(0.1159)	(0.1137)	(0.0216)	(0.0607)

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

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

	$\beta_1 = 70$	$\beta_2 = 34$	$\varphi = 20$	$\psi = 9$	$\sigma_\varepsilon^2 = 1$
SLS	70.0203	34.0303	20.0319	8.9625	1.0016
	(0.0398)	(0.0395)	(0.0258)	(0.0128)	(0.0249)
SBE	69.9754	34.2096	19.1365	10.8034	0.8936
	(0.1183)	(0.1146)	(0.1094)	(0.0180)	(0.0537)

Example: Logistic model with 2 random effects

- The model

$$\begin{aligned} y_{it} &= \frac{\xi_{1i}}{1+\exp[-(x_{it}-\xi_{2i})/\beta]} + \varepsilon_{it}, \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2) \\ \xi_i &= \varphi + \delta_i, \delta_i \sim N_2\left[(0,0), \text{diag}(\psi_1, \psi_2)\right] \end{aligned}$$

- 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 $n = 7$, $T = 5$.

True	$\beta = 350$	$\varphi_1 = 200$	$\varphi_2 = 700$	$\psi_1 = 100$	$\psi_2 = 625$	$\sigma_\varepsilon^2 = 25$
SBE	349.8222	199.3850	699.3057	104.8866	634.3594	25.3303
	(0.5896)	(0.5984)	(0.5620)	(0.0088)	(0.0533)	(0.2605)

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