Optimal Designs for Nonlinear Regression with Dependent Errors using Matrix Norms

Pauline Baur, Prof. Dr. Kirsten Schorning, Prof. Dr. Holger Dette

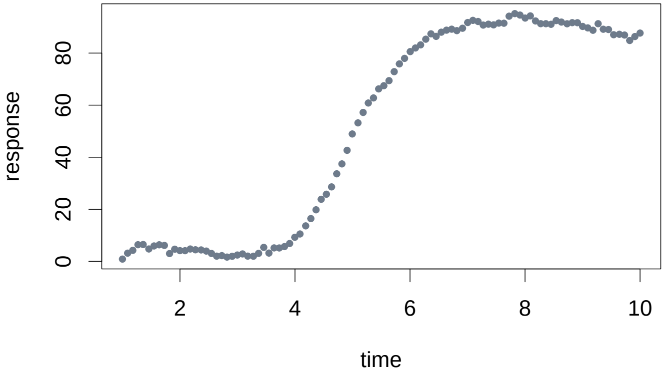
mODa14, June 15th, 2026

TU Dortmund University

technische universität dortmund

Transregio 391

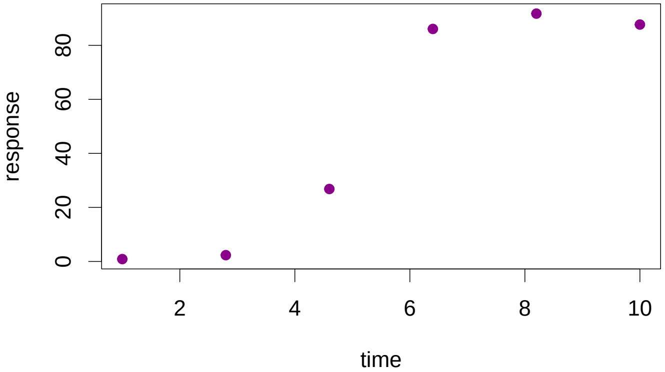
Motivation



We observe a stochastic process at multiple timepoints

Goal: Find optimal observation points in time following a model-based approach

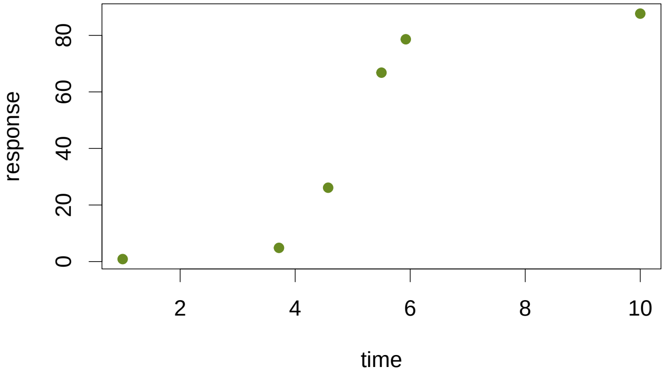
Motivation



We observe a stochastic process at multiple timepoints

Goal: Find optimal observation points in time following a model-based approach

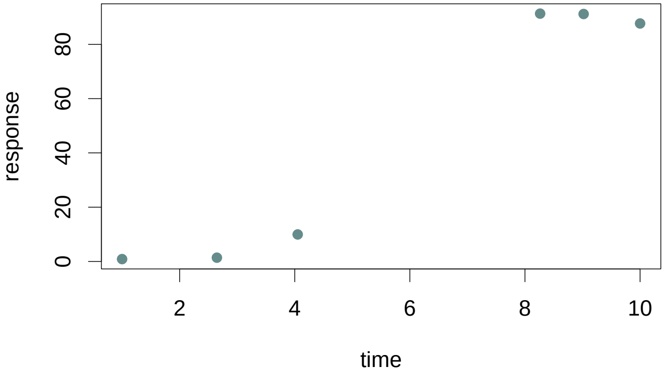
Motivation



We observe a stochastic process at multiple timepoints

Goal: Find optimal observation points in time following a model-based approach

Motivation



We observe a stochastic process at multiple timepoints

Goal: Find optimal observation points in time following a model-based approach

Model Assumptions

We assume the nonlinear regression model

$$Y_{t_i} = f(t_i, \theta) + \int_0^{t_i} \sigma(s) d\varepsilon_s, \quad i = 1, \ldots, n, \quad 0 < a = t_1 < \ldots < t_n = b$$

- Y Observable process
- f Regression function
- θ Unknown regression parameter, $\theta \in \mathbb{R}^p$
- σ Square-integrable variance function
- ε Brownian Motion
- $\mathcal{X}_n$ Design space $\{(t_1, \ldots, t_n) | a = t_1 < \ldots < t_n = b\}$

Goal: Choose observation times to optimally estimate θ

Challenge: Handle dependent errors

Optimal Design Theory for Independent Errors

Determine Fisher information matrix of design $d = (t_1, \dots, t_n) \in \mathcal{X}_n$

$$M_n(\theta; d) = \sum_{i=1}^{n} \mathbb{E}_\theta \left[\left(\frac{d}{d\theta} \ell(Y_{t_i}, t_i, \theta) \right) \left(\frac{d}{d\theta} \ell(Y_{t_i}, t_i, \theta) \right)^\top \right]$$

with ℓ denoting the log-likelihood function of $Y = (Y_{t_1}, \dots, Y_{t_n})^\top$

Optimal design theory: A design $d^* = (t_1^*, \dots, t_n^*)$ is ϕ-optimal if it minimizes function $\phi(M_n(\theta; d))$ for all $d = (t_1, \dots, t_n) \in \mathcal{X}_n$, i.e.

$$\phi(M_n(\theta; d^*)) \leq \phi(M_n(\theta; d)) \qquad \forall d \in \mathcal{X}_n$$

Optimal Design for Dependent Errors

- Independent errors: $d \mapsto \phi(M_n(\theta; d))$ is convex
- Dependent errors: $\phi(M_n(\theta; d))$ is not convex

We look for a different approach for the dependent case

Our plan: Construct a new optimality criterion ϕ, still depending on Fisher information matrices

Continuous-Time Model

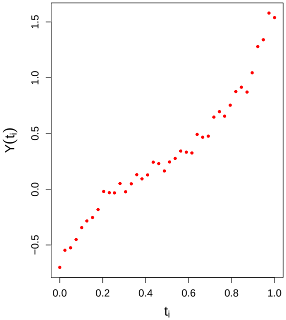


Figure 1: Discrete-time regression model

$$Y_{t_i} = f(t_i, \theta) + \int_0^{t_i} \sigma(s) d\varepsilon_s, \quad i = 1, \ldots n, \quad (t_1, ..., t_n) \in \mathcal{X}_n$$

Continuous-Time Model

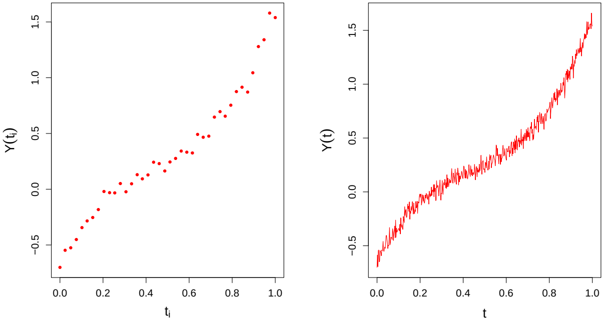


Figure 2: Discrete-time vs. continuous-time regression model

$$Y_t = f(t, \theta) + \int_0^t \sigma(s) d\varepsilon_s, \quad t \in [a, b]$$

Our Approach: Approximate Continuous-Time Setting

Minimize the distance between the inverse Fisher information of the discrete model ($M_n(d,\theta)$) and continuous model ($M(\theta)$)

$$d^* = (t_1^*, \dots, t_n^*) = \arg\min_{d \in \mathcal{X}_n} \| M_n^{-1}(d,\theta) - M^{-1}(\theta) \|_F^2 \tag{1}$$

with Frobenius norm

$$\|M\|_F = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} |m_{ij}|^2} = \sqrt{\operatorname{trace}\left(M^\top M\right)} \text{ for } M \in \mathbb{R}^{p \times p}$$

To Do:

1. Derive likelihood and Fisher information for both frameworks
2. Implement optimality criterion $\phi(d) = \| M_n^{-1}(d,\theta) - M^{-1}(\theta) \|_F^2$ in (1) in R

Discrete-Time Model: Likelihood

Re-write the model using differences:

$$Z_1 =: Y_{t_1} = f(a, \theta) + \int_0^a \sigma(s) d\varepsilon_s,$$

$$Z_i =: Y_{t_i} - Y_{t_{i-1}} = \underbrace{f(t_i, \theta) - f(t_{i-1}, \theta)}_{=:g(t_i,\theta)} + \underbrace{\int_{t_{i-1}}^{t_i} \sigma(s) d\varepsilon_s}_{=:\eta_i}, \quad i = 2, \ldots, n.$$

Discrete-Time Model: Likelihood

Re-write the model using differences:

$$Z_1 =: Y_{t_1} = f(a,\theta) + \int_0^a \sigma(s)d\varepsilon_s,$$

$$Z_i =: Y_{t_i} - Y_{t_{i-1}} = \underbrace{f(t_i,\theta) - f(t_{i-1},\theta)}_{=:g(t_i,\theta)} + \underbrace{\int_{t_{i-1}}^{t_i} \sigma(s)d\varepsilon_s}_{=:\eta_i}, \quad i = 2,\dots,n.$$

Apply properties of Brownian Motion:

- $\eta_i \sim \mathcal{N}\left(0, \int_{t_{i-1}}^{t_i} \sigma(s)^2 ds\right)$ are independent increments
- $Z_i \sim \mathcal{N}\left(g(t_i,\theta), \int_{t_{i-1}}^{t_i} \sigma(s)^2 ds\right)$
- $Z_1 \sim \mathcal{N}\left(f(a,\theta), \int_0^a \sigma(s)^2 ds\right)$
- $Z_1, Z_2, \ldots$ are independent

$\Rightarrow$ Calculate the log-likelihood function of $Z = (Z_1, Z_2, \dots, Z_n)$

Discrete-Time Model: Likelihood

Re-write the model using differences:

$$Z_1 =: Y_{t_1} = f(a,\theta) + \int_0^a \sigma(s) d\varepsilon_s,$$

$$Z_i =: Y_{t_i} - Y_{t_{i-1}} = \underbrace{f(t_i,\theta) - f(t_{i-1},\theta)}_{=:g(t_i,\theta)} + \underbrace{\int_{t_{i-1}}^{t_i} \sigma(s) d\varepsilon_s}_{=:\eta_i}, \quad i = 2,\dots,n.$$

Apply properties of Brownian Motion:

- $\eta_i \sim \mathcal{N}\left(0, \int_{t_{i-1}}^{t_i} \sigma(s)^2 ds\right)$ are independent increments
- $Z_i \sim \mathcal{N}\left(g(t_i,\theta), \int_{t_{i-1}}^{t_i} \sigma(s)^2 ds\right)$
- $Z_1 \sim \mathcal{N}\left(f(a,\theta), \int_0^a \sigma(s)^2 ds\right)$
- $Z_1, Z_2, \dots$ are independent

$\Rightarrow$ Calculate the log-likelihood function of $Z = (Z_1, Z_2, \dots, Z_n)$

Fisher Information Matrices

Discrete-time: with some straightforward calculations:

$$M_n(d,\theta) = \sum_{i=2}^{n} \frac{1}{\int_{t_{i-1}}^{t_i} \sigma(s)^2 ds} \left(\frac{d}{d\theta} g(t_i,\theta)\right) \left(\frac{d}{d\theta} g(t_i,\theta)\right)^\top + \frac{\left(\frac{d}{d\theta} f(a,\theta)\right)\left(\frac{d}{d\theta} f(a,\theta)\right)^\top}{\int_0^a \sigma(s)^2 ds}$$

Continuous-time: Identify differences with stochastic differential equations

Calculate with Itô integrals:

$$M(\theta) = \int_a^b \left(\frac{\frac{d}{d\theta}\frac{d}{dt} f(t,\theta)}{\sigma(t)}\right) \left(\frac{\frac{d}{d\theta}\frac{d}{dt} f(t,\theta)}{\sigma(t)}\right)^\top dt + \frac{\left(\frac{d}{d\theta} f(a,\theta)\right)\left(\frac{d}{d\theta} f(a,\theta)\right)^\top}{\int_0^a \sigma(s)^2 ds}$$

Fisher Information Matrices

Discrete-time: with some straightforward calculations:

$$M_n(d,\theta) \overset{\sigma(t)=1}{=} \sum_{i=2}^{n}(t_i - t_{i-1})\left(\frac{\frac{\mathrm{d}}{\mathrm{d}\theta}g(t_i,\theta)}{t_i - t_{i-1}}\right)\left(\frac{\frac{\mathrm{d}}{\mathrm{d}\theta}g(t_i,\theta)}{t_i - t_{i-1}}\right)^{\top} + \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\theta}f(a,\theta)\right)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}f(a,\theta)\right)^{\top}}{a}$$

Continuous-time: Identify differences with stochastic differential equations

Calculate with Itô integrals:

$$M(\theta) \overset{\sigma(t)=1}{=} \int_a^b \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\mathrm{d}}{\mathrm{d}t}f(t,\theta)\right)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\mathrm{d}}{\mathrm{d}t}f(t,\theta)\right)^{\top} dt + \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\theta}f(a,\theta)\right)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}f(a,\theta)\right)^{\top}}{a}$$

Optimality Criterion

We showed with the Cauchy–Schwarz inequality that

$$M_n(d,\theta) \leq_L M(\theta) \qquad \forall d \in \mathcal{X}_n$$

Furthermore, with Riemann sums:

$$M_n(d,\theta) \underset{n\to\infty}{\rightarrow} M(\theta)$$

Therefore, the function ϕ is a reasonable optimality criterion:

$$\phi(d) = \phi(d,\theta) = \|M_n^{-1}(d,\theta) - M^{-1}(\theta)\|_F^2$$

A design $d^* = (t_1, \dots, t_n) \in \mathcal{X}_n$ is ϕ-optimal if and only if

$$d^* = \underset{d\in\mathcal{X}_n}{\arg\min}\, \phi(d) = \underset{d\in\mathcal{X}_n}{\arg\min} \|M_n^{-1}(d,\theta) - M^{-1}(\theta)\|_F^2$$

Optimality Criterion

Problem: The function $\phi(d)$ is non-convex and the corresponding optimal design difficult to compute analytically

Approach:

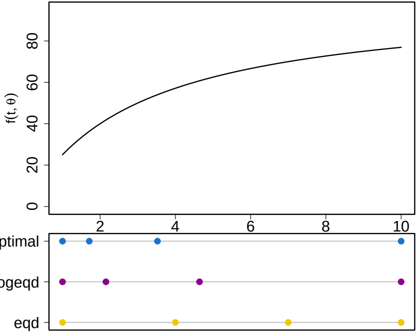
1. Use metaheuristic algorithm (GenSA) to compute candidate d^*.
2. Determine partial derivative of $\phi(d)$:

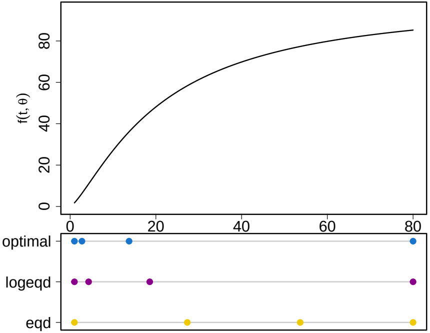
$$
\begin{aligned}
\frac{d}{dt_i}\phi(d) = \frac{d}{dt_i}\phi(t_1,\dots,t_n) &= \frac{d}{dt_i}\|M_n^{-1}(d,\theta) - M^{-1}(\theta)\|_F^2 \\
&= \operatorname{trace}\left\{\frac{d}{dt_i}\left(M_n^{-1}(d,\theta) - M^{-1}(\theta)\right)^2\right\} \\
&= \operatorname{trace}\left\{2M_n^{-1}(d,\theta)\left(M^{-1}(\theta) - M_n^{-1}(d,\theta)\right)M_n^{-1}(d,\theta)\frac{dM_n(d,\theta)}{dt_i}\right\}
\end{aligned}
$$

3. Derive necessary condition to check optimality of d^*: If $d^* = (t_1^*,\dots,t_n^*) \in \mathcal{X}_n$ is optimal, it holds $\frac{d}{dt_i}\phi(d^*) = 0$ for all $i \in \{2,\dots,n-1\}$

Simulation Study: Setup and Designs for $n = 4$

Emax-Model	Logistic-Model
$f(t,\theta) = E_0 + E_{max}\frac{t}{EC_{50}+t}$	$f(t,\theta) = E_0 + E_{max}\frac{\exp(\alpha+\beta\log(t))}{1+\exp(\alpha+\beta\log(t))}$
$\theta = (E_0, E_{max}, EC_{50}) = (0, 100, 3)$	$\theta = (E_0, E_{max}, \alpha, \beta) = (0, 100, -4.03, 1.32)$
$\sigma(s)^2 = 1.5$	$\sigma(s)^2 = \log(0.1s + 2)$
$[a, b] = [1, 10]$	$[a, b] = [1, 80]$





Simulation Study: Setup and Designs for $n = 4$

Emax-Model	Logistic-Model
$f(t,\theta) = E_0 + E_{\max} \frac{t}{EC_{50}+t}$	$f(t,\theta) = E_0 + E_{\max} \frac{\exp(\alpha+\beta\log(t))}{1+\exp(\alpha+\beta\log(t))}$
$\theta = (E_0, E_{\max}, EC_{50}) = (0, 100, 3)$	$\theta = (E_0, E_{\max}, \alpha, \beta) = (0, 100, -4.03, 1.32)$
$\sigma(s)^2 = 1.5$	$\sigma(s)^2 = \log(0.1s + 2)$
$[a, b] = [1, 10]$	$[a, b] = [1, 80]$

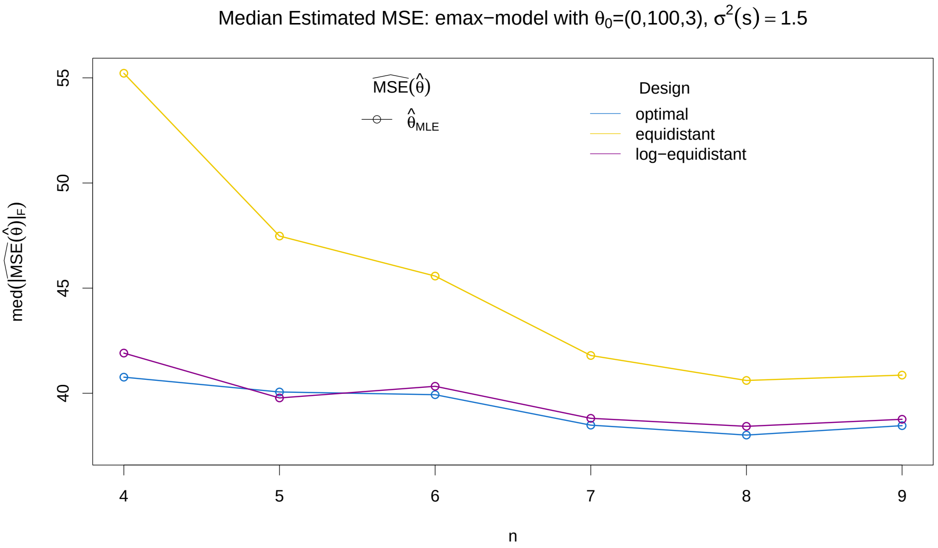
Optimal designs fulfill $\frac{d}{dt_i}\phi(d^*) = 0$ for all $i \in \{2, \dots, n-1\}$

Estimator: MLE $\hat{\theta}_{\text{MLE}}$ calculated using Generative Simulated Annealing

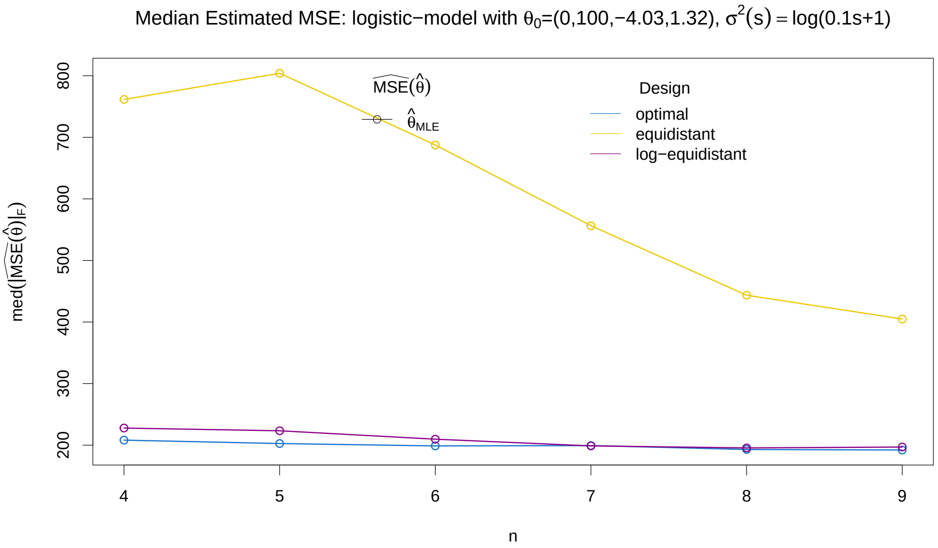
Main question to be answered on 15.000 simulated datasets for $n \in \{4, \dots, 9\}$:

- Does our design improve the estimator's MSE?

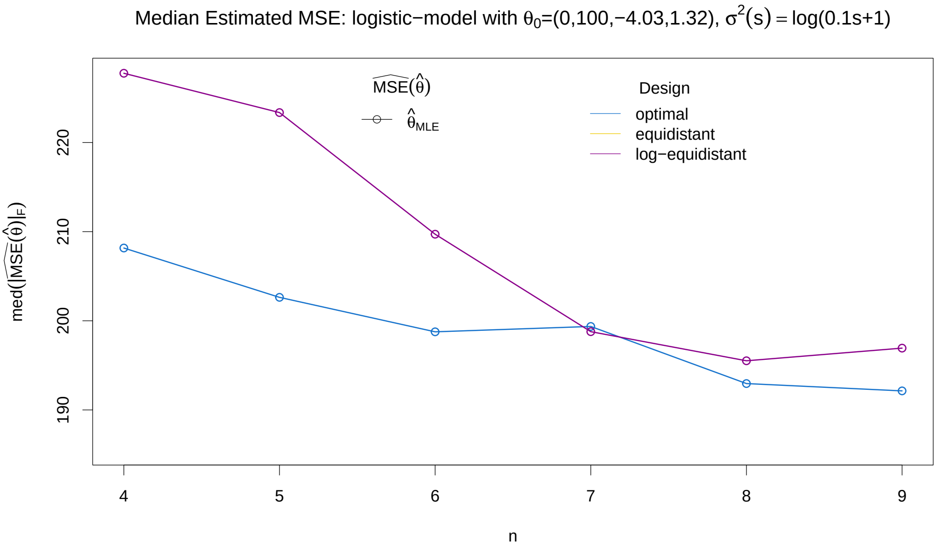
Simulation Study: Emax Model



Simulation Study: Logistic Model



Simulation Study: Logistic Model



Summary and Outlook

The (locally) optimal design (based on ϕ) improves the parameter estimator's MSE

Further results are available for other matrix norms (square root distance and Procrustes distance)

Future research:

Parameter estimation in models with a parametric variance function

$$Y_{t_i} = f(t_i, \theta) + \int_0^{t_i} \sigma(s, \gamma) \varepsilon_s$$

Designs that are less sensitive towards parameter misspecification

- Study robustness of locally optimal designs in simulations
- Bayesian designs, adaptive designs, ...

Extend theory to optimal designs in space and time

References

Bishwal, J. P. N. (2008). Parameter Estimation in Stochastic Differential Equations. Springer.

Dette, H., Konstantinou, M., & Zhigljavsky, A. (2017). A new approach to optimal designs for correlated observations. The Annals of Statistics, 45(4), 1579 – 1608.

Iacus, S. M. (2008). Simulation and Inference for Stochastic Differential Equations: With R Examples. Springer.

Ibragimov, I. A., & Has'minskii, R. Z. (1981). Statistical Estimation: Asymptotic Theory. Springer.

Seber, G., & Wild, C. (1989). Nonlinear regression. Wiley.

Silvey, S. D. (1980). Optimal design. Springer Netherlands.

Øksendal, B. (2003). Stochastic Differential Equations. Springer.

THANK YOU FOR YOUR ATTENTION!

Parameter Estimation

MLE	MLE on linearized model
E.g. with Generative Simulated Annealing: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \ell(Z, d, \theta)$	

Parameter Estimation

MLE	MLE on linearized model
	Linearize with Taylor approximation: $Z_i = g(t_i, \theta) + \eta_i$
E.g. with Generative Simulated Annealing: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \ell(Z, d, \theta)$	

Parameter Estimation

MLE	MLE on linearized model
	Linearize with Taylor approximation: $Z_i \approx g(t_i, \theta_0) + \left(\frac{d}{d\theta} g(t_i, \theta_0)\right)(\theta - \theta_0) + \eta_i$
E.g. with Generative Simulated Annealing: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \ell(Z, d, \theta)$	

Parameter Estimation

MLE	MLE on linearized model
	Linearize with Taylor approximation: $Z_i \approx g(t_i, \theta_0) + \left(\frac{d}{d\theta} g(t_i, \theta_0)\right)(\theta - \theta_0) + \eta_i$
E.g. with Generative Simulated Annealing: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \ell(Z, d, \theta)$	Analytically derive linearized MLE: $\hat{\theta}_{\text{LIN}} = \theta_0 + M_n^{-1}(\theta_0; d) \frac{d}{d\theta} \ell(Z, d, \theta_0)$

Parameter Estimation

MLE	MLE on linearized model
	Linearize with Taylor approximation: $Z_i \approx g(t_i, \theta_0) + \left(\frac{d}{d\theta} g(t_i, \theta_0)\right)(\theta - \theta_0) + \eta_i$
E.g. with Generative Simulated Annealing: $\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \ell(Z, d, \theta)$	Analytically derive linearized MLE: $\hat{\theta}_{\text{LIN}} = \theta_0 + M_n^{-1}(\theta_0; d) \frac{d}{d\theta} \ell(Z, d, \theta_0)$
+ Feasible in practice – Numerically challenging – Time-consuming	+ Unbiased and efficient – Infeasible in practice $\Rightarrow$ Use as benchmark

Simulation Study: Emax Model