

**CORRELATION OF PARAMETER ESTIMATORS FOR
MODELS ADMITTING MULTIPLE PARAMETRIZATIONS**

Kaska Adoteye¹, Robert Baraldi², Kevin Flores³,
John Nardini⁴, H.T. Banks⁵ §, W. Clayton Thompson⁶

Center for Research in Scientific Computation
Department of Mathematics
North Carolina State University
Raleigh, NC 27695-8212, USA

⁵Cardiovascular and Metabolic Diseases Research Unit
Pfizer Inc.
Cambridge, MA 02139, USA

Abstract: When estimating parameters using noisy data, uncertainty quantification methods provide a way to investigate the confidence one has in the parameter estimates, as well as to obtain information on the possible dependence of parametric estimators on one another. In this note, we consider uncertainty quantification techniques that allow visualization of the distributions of these parameter estimators for evidence of possible correlation. We consider three mathematical models (the logistic curve, the Richards curve, and the spring equation), which permit multiple parametrizations, and compare the corresponding parameter estimators for possible dependence/independence. The uncertainty quantification techniques we employ include the correlation coefficients, asymptotic as well as exact confidence regions or ellipsoids, and Monte Carlo plots generated by the DRAM algorithm.

Key Words: inverse problems, effects of parameterizations, independent vs. dependent estimators, exact vs. asymptotic confidence ellipsoids, Monte Carlo simulations, Bayesian analysis, DRAM

Received: October 20, 2015

© 2015 Academic Publications, Ltd.
url: www.acadpubl.eu

§Correspondence author

1. Introduction

In the scientific study of physical and biological systems, mathematical modeling provides a conceptual framework for the quantitative investigation of the processes being considered. The development of this conceptual framework is an iterative process in which an abstracted and mathematized representation of the system is used to make predictions which are then tested experimentally; these tests provide greater insight into the operation of the system and its mathematical representation [10]. The link between a mathematical model and experimental data is described by a statistical model that encodes any uncertainty as a result of model misspecification, measurement error, etc. This leads to the problem of uncertainty quantification [7, 23] to assess the extent to which model based conclusions are robust to modeling and data errors.

In this report, we consider one component of uncertainty quantification: the computation of confidence regions for parameters estimated from noisy data. When multiple model parameters are estimated, it is almost always the case that the estimation of all parameters must be computed simultaneously (e.g., by minimizing the sum of squared errors between the model and the data). As an estimator is formally a random variable [10, 22], uncertainty quantification can be viewed as the study of the properties of the joint distribution of these random variables. Ideally, for any two parameters of the mathematical model, their corresponding estimators will be independent (their joint distribution can be factored) so that uncertainty in the estimation of one parameter will have no effect on the uncertainty in the estimation of the other parameter. In practice, however, this is rarely the case. While various techniques and algorithms exist to identify a subset of parameters that can be estimated with the greatest degree of certainty [3, 11], an alternative problem of interest and utility is how the structure of the model itself has an impact on the joint distribution of the estimators.

In particular, we examine two aspects of this problem. First, we consider how alternative model parameterizations (transformations of the parameter space such that the model solution is unchanged) have an impact on the precision with which the parameters can be estimated. Second, we consider how alternative computational approaches for parameter estimation and/or the construction of confidence regions can improve understanding of the estimator properties. To address these issues, we consider three simple example systems: a logistic growth model, a generalized logistic growth (Richards' curve) model and a damped spring-mass model. In each case, it is shown that the model can be represented by at least two different but ultimately equivalent

parameterizations, (e.g., the logistic equation can be equivalently written as $\dot{P} = rP(1 - P/K)$ or $\dot{P} = AP - BP^2$ with $A = r$ and $B = r/K$). For each model, synthetic data is generated by adding pseudo-random measurement error at fixed measurement times. These data are then used to estimate the model parameters, which are compared to the true values used to create the data. For comparison, we consider both frequentist (nonlinear least squares) and Bayesian (delayed rejection adaptive Metropolis – DRAM) methods for parameter estimation. The asymptotic properties of nonlinear least squares estimators are well-characterized [5, 13, 22] and approximate confidence regions are constructed using either an estimate of the covariance matrix or by level sets of the least squares cost function. Confidence regions for the DRAM algorithm are visualized by direct sampling from the estimated posterior distribution.

Even though only scalar examples (with simple forward-solve solutions) are used, these simulation studies provide some interesting insight into the interaction between the parameterization of a mathematical model, the statistical properties of estimators, and the approximation and visualization of confidence regions.

We begin in Section 2 with a brief outline of the mathematical and statistical framework and associated methodology we employ. This is followed in Section 3 by the formulation of the mathematical models we use in our investigations. Finally a summary of our findings is followed by brief concluding remarks.

2. Methodology

2.1. Statistical Analysis and Inverse Problem Methods

In order to estimate parameters using asymptotic theory, we consider mathematical models of the form

$$\frac{dz}{dt} = g(t, z(t), q) \quad (1)$$

$$z(t_0) = z_0 \quad (2)$$

with observation process

$$f(t, \theta) = \mathcal{C}z(t, \theta)$$

where $\theta = (q, \tilde{z}_0) \in \mathbb{R}^{p+\tilde{m}}$ is the vector of unknown parameters, q is a vector of model parameters, \tilde{z}_0 is the portion of initial conditions that is unknown (if any), and \mathcal{C} maps the model solution $z(t, \theta)$ in \mathbb{R}^m to the observed states $f(t, \theta)$. In this investigation, for simplicity, the initial condition is always assumed to be

completely known, and thus $\theta = q$. Also, in this investigation the observation operator will always produce a scalar, and thus \mathcal{C} maps \mathbb{R}^m to \mathbb{R} .

As experimental data are typically available at discrete times, we will assume that the observations occur at n discrete times t_j , and thus the observations will be

$$f(t_j, \theta) = \mathcal{C}z(t_j, \theta), \quad j = 1, \dots, n.$$

To account for measurement error, we use a statistical model

$$Y_j = f(t_j, \theta_0) + \mathcal{E}_j, \quad j = 1, \dots, n \quad (3)$$

for our observations, where \mathcal{E}_j is a random variable that represents the random noise that causes our observed data to deviate from our model solution, and θ_0 is the hypothesized “true” parameter vector that generates the observations $\{Y_j\}_{j=1}^n$. The existence of this parameter vector is a standard assumption in frequentist statistical formulations. We will also assume $\mathbb{E}(\mathcal{E}_j) = 0$ for each j , which results from our implicit assumption that our model in Equations (1)-(2) correctly describes the underlying phenomenon. We assume that $\mathcal{E}_j, j = 1, \dots, n$, are independent and identically distributed. In this study, we will assume $\mathcal{E}_j, j = 1, \dots, n$ are normally distributed with variance σ_0^2 and thus $Y_j \sim \mathcal{N}(f(t_j, \theta_0), \sigma_0^2)$.

Since \mathcal{E}_j is a random variable, Y_j is a random variable with corresponding realization (i.e., data) y_j , which for this investigation will be simulated in a manner described in Section 3.4. Our goal is to then estimate θ_0 (which will be known to us throughout this investigation) by creating a random variable estimator Θ whose realizations for a given data set y_j will be estimates $\hat{\theta}$ of θ_0 . Our estimates $\hat{\theta}$ will approximate our “true” parameters θ_0 , and are obtained by minimizing the least squares cost functional

$$J(\theta) = \sum_{j=1}^n [y_j - f(t_j, \theta)]^2, \quad (4)$$

and thus, with Q being the space of admissible parameters,

$$\theta_0 \approx \hat{\theta} = \arg \min_{\theta \in Q} J(\theta).$$

This process of estimating parameters from data is known as an inverse problem, and we will compute all inverse problems using `fminsearch` in MATLAB.

Once we have an estimate $\hat{\theta}$, we wish to ascertain the statistical properties of the estimator Θ . Although we do not know the distribution of the estimator Θ ,

we can approximate it under *asymptotic theory* (as $n \rightarrow \infty$) by the multivariate Gaussian distribution [7, 10, 22]

$$\Theta \sim \mathcal{N}(\theta_0, \Sigma_0^n)$$

where, based on the previous assumptions, the covariance matrix Σ_0^n is approximated by

$$\Sigma_0^n \approx \hat{\Sigma}^n = \hat{\sigma}_0^2 \left[\chi^{nT}(\hat{\theta}) \chi^n(\hat{\theta}) \right]^{-1}. \quad (5)$$

Here

$$\hat{\sigma}_0^2 = \frac{1}{n-p} \sum_{j=1}^n [y_j - f(t_j, \hat{\theta})]^2 = \frac{J(\hat{\theta})}{n-p}$$

is an unbiased estimate for σ_0^2 , p is the number of parameters being estimated, and χ^n is the sensitivity matrix

$$\chi_{jk}^n(\theta) = \frac{\partial f(t_j, \theta)}{\partial \theta_k}, \quad j = 1, \dots, n, \quad k = 1, \dots, p, \quad (6)$$

where θ_k is the k^{th} component of the vector θ . This sensitivity matrix can be computed using several different methods including differencing techniques or exact sensitivity equations [7, 10]. Here, we use the “complex step method” [18].

In this manuscript we will look at ways to visualize and understand the estimated distribution $\mathcal{N}(\hat{\theta}, \hat{\Sigma}^n)$. The first such method we will employ is related to the approximate covariance matrix $\hat{\Sigma}^n$ of the estimator. We will compute the correlation coefficients

$$\rho_{ij} = \frac{\hat{\Sigma}_{ij}^n}{\sqrt{\hat{\Sigma}_{ii}^n} \sqrt{\hat{\Sigma}_{jj}^n}},$$

where $\rho_{ij} \in [-1, 1]$ increases in absolute value as the correlation between the estimator components increases. In order to determine the confidence we have in the parameter estimates, we also compute the asymptotic theory based standard error $SE(\hat{\theta}_k) = \sqrt{\hat{\Sigma}_{kk}^n}$ for the k^{th} parameter (we are more confident in the parameter estimate as the standard error decreases).

2.2. Confidence “Ellipsoids” or Regions

2.2.1. Asymptotic Ellipsoids

In order to further understand the distribution of our estimator Θ , we will construct ellipsoids based on the covariance matrix. We define the quantity

$$Q = (\Theta - \theta_0)^T (\Sigma_0^n)^{-1} (\Theta - \theta_0).$$

and note that, if Σ_0^n is positive definite (see below), then $Q \sim \chi_p^2$, a Chi-square distribution with p degrees of freedom [21, Thm. 2.9]. Recall that, given our statistical assumptions, our covariance matrix Σ_0^n for our estimator Θ can be approximated as in Equation (5). Therefore, here we have approximately

$$Q = (\Theta - \theta_0)^T \chi^T \chi (\Theta - \theta_0) / \sigma^2 \sim \chi_p^2.$$

Also, since $J(\Theta)/(n - p)$ is an unbiased estimator for σ^2 , we have $J(\Theta)/\sigma^2 \sim \chi_{n-p}^2$, a Chi-square distribution with $n - p$ degrees of freedom [13].

Letting $s^2 = J(\Theta)/(n - p)$, we can define the quantity

$$F(\Theta) = \frac{(\Theta - \theta_0)^T \chi^T \chi (\Theta - \theta_0)}{ps^2}$$

and see that

$$\begin{aligned} F(\Theta) &= \frac{(\Theta - \theta_0)^T \chi^T \chi (\Theta - \theta_0)}{ps^2} \\ &= \frac{(\Theta - \theta_0)^T \chi^T \chi (\Theta - \theta_0) / p}{J(\Theta) / (n - p)} \\ &= \left(\frac{(\Theta - \theta_0)^T \chi^T \chi (\Theta - \theta_0)}{J(\Theta)} \right) \left(\frac{n - p}{p} \right) \\ &= \left(\frac{(\Theta - \theta_0)^T \chi^T \chi (\Theta - \theta_0)}{J(\Theta)} \right) \left(\frac{n - p}{p} \right) \left(\frac{\sigma^2}{\sigma^2} \right) \\ &= \frac{[(\Theta - \theta_0)^T \chi^T \chi (\Theta - \theta_0) / \sigma^2] / p}{[J(\Theta) / \sigma^2] / (n - p)}. \end{aligned}$$

In both the numerator and denominator we have a Chi-squared distribution scaled by its number of degrees of freedom, which is by definition the F-distribution [13]. Thus,

$$F(\Theta) = \frac{(\Theta - \theta_0)^T \chi^T \chi (\Theta - \theta_0)}{ps^2} = \frac{(\Theta - \theta_0)^T (\Sigma_0^n)^{-1} (\Theta - \theta_0)}{p} \sim F_{p, n-p}.$$

It follows that an approximate $100(1 - \alpha)\%$ confidence ellipsoid [5, 13, 22] is given by

$$\left\{ \theta : (\theta - \hat{\theta})^T \left(\hat{\Sigma}^n \right)^{-1} (\theta - \hat{\theta}) \leq p F_{p, n-p}^\alpha \right\} \quad (7)$$

where $F_{p, n-p}^\alpha$ is the upper- α critical value of the $F_{p, n-p}$ distribution. This is the asymptotic confidence ellipsoid for the OLS estimate $\hat{\theta}$, where we see that this confidence ellipsoid is completely determined by the covariance matrix $\hat{\Sigma}^n$ and the confidence level α , and thus this provides another method of visualizing and understanding the distribution for our estimator Θ . In this work we plotted the level curves of the ellipsoid from Equation (7) using the command `contour` in MATLAB for $\alpha = .01, .05$, and $.10$ for the 99%, 95%, and 90% confidence ellipsoids, respectively.

An important note is that we began this section with the assumption that our covariance matrix Σ_0^n is positive definite. It is known that, in general, a covariance matrix is positive semi-definite, and thus it is positive definite if it is also full rank. Here, we are numerically estimating the covariance matrix, and thus we also require the matrix to have a reasonable condition number in order to allow for numerical inversion of the matrix. In this presentation, since we work with basic systems with a low number of parameters, this was not an issue, but in general this matrix may be nearly singular (i.e., ill-conditioned) and this may lead to numerical issues.

2.2.2. Exact Confidence Regions or “Exact Ellipsoids”

The asymptotic ellipsoids in the previous section are based on the covariance matrix, which by its very definition considers linear relationships. In order to obtain a more general representation of the distribution of our estimator Θ , we calculated the exact confidence regions or “exact ellipsoids” as detailed in [22]. Summarizing [22], we will base these “ellipsoids” on the cost functional in Equation (4), and thus they will take the form of the exact confidence region

$$\{ \theta : J(\theta) \leq cJ(\hat{\theta}) \} \quad (8)$$

where $c > 1$. We only need to find a suitable value for c . According to [22], we have

$$J(\theta) - J(\hat{\theta}) \approx (\theta - \hat{\theta})^T \chi^T \chi (\theta - \hat{\theta}).$$

Therefore by substituting $s^2 = J(\hat{\theta})/(n - p)$ into equation (7), we can establish

$$J(\theta) - J(\hat{\theta}) \leq J(\hat{\theta}) \frac{p}{n - p} F_{p, n-p}^\alpha.$$

Rearranging this, we obtain the exact confidence region

$$\left\{ \theta : J(\theta) \leq J(\hat{\theta}) \left(1 + \frac{p}{n-p} F_{p, n-p}^\alpha \right) \right\}, \quad (9)$$

which is the region described in Equation (8) with $c = 1 + \frac{p}{n-p} F_{p, n-p}^\alpha$. This will approximately have the required asymptotic confidence level of $100(1-\alpha)\%$, but the computation itself avoids the linearization seen in the asymptotic ellipsoids.

2.3. DRAM (Bayesian Analysis)

The above analysis tools assume that there is a true parameter value θ_0 , while in a Bayesian framework we assume that the parameters are random variables with associated densities. In order to compute these densities, we begin with an initial density $\pi_0(\theta)$ for the parameters, known as a *prior* density. We then use data realizations y in order to create a posterior density $\pi(\theta|y)$ through the use of a likelihood function $\pi(y|\theta)$. In solving the inverse problem under this framework, we use Bayes' theorem for inverse problems [16] given in Definition 1.

Definition 1. (Bayes' theorem for inverse problems) We assume that the parameter vector θ is a random variable which has a known prior density $\pi_0(\theta)$ (possibly non-informative), and corresponding realizations y of the random variable Y associated with the measurement process. The posterior density of θ , given measurements y , is then

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi_0(\theta)}{\pi(y)} = \frac{\pi(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} \pi(y|\theta)\pi_0(\theta)d\theta} \quad (10)$$

where we have assumed that the marginal density $\pi(y) = \int_{\mathbb{R}^p} \pi(\theta, y)d\theta = \int_{\mathbb{R}^p} \pi(y|\theta)\pi_0(\theta)d\theta \neq 0$ (a normalizing factor) is the integral over all possible joint densities $\pi(\theta, y)$. Note here that $\pi(y|\theta)$ is a likelihood function which describes how likely a data set y is when given a model solution at the parameter value θ .

There have been many techniques developed to compute the integral in Definition 1. For a small number of parameters, cubature or Monte Carlo techniques can be used. When the number of parameters increases, a popular method is to construct a Markov chain whose stationary distribution is the posterior density $\pi(\theta|y)$ of Equation (10). With this method we sample the parameter space, accepting the parameters based on closeness of the model solution to the data. It is known that a Markov chain defined by the random

walk Metropolis algorithm will converge (e.g., in probability or measure) if the algorithm is allowed to sample the parameter space a large number of times [24]. Here we used the delayed rejection adaptive Metropolis (DRAM) algorithm, which is described in detail in [2, 14, 15, 24] and available as a package for MATLAB at <http://helios.fmi.fi/~lainema/mcmc/>. Our simulations assumed a non-informative prior given by a uniform distribution over a space that served as large bounds on our parameters. We simulated $M=50,000$ chain iterations to assure mixing. We also assumed the measurement errors were normally distributed.

One benefit to this approach is that the adaptive DRAM algorithm samples directly from the posterior distribution, and one can visualize this posterior distribution by plotting such samples. One downside to using DRAM is the long computational times, which have been explored in comparison to the asymptotic approach in [17]. It is important to note that while there has been a successful parallel implementation of DRAM [25], here we used a serial implementation in order to simplify implementation since computational times or function evaluations, etc., were not a focus of our efforts.

3. Mathematical Models

Here we consider three models (the logistic growth model, the Richards curve and a damped spring-mass system model) and equivalent parametrizations of each model. We wish to explore the effect of different parametrizations on the distributions for the estimators Θ using the methods described above (asymptotic and exact ellipsoids, correlation coefficients, and DRAM Monte Carlo plots).

3.1. Logistic

The first model we consider is the widely used Verhulst-Pearl or logistic model for a bounded, dynamically changing population $P(t)$, given by the differential equation

$$\frac{dP}{dt}(t) = rP(t) \left(1 - \frac{P(t)}{K} \right), \quad (11)$$

where r is the intrinsic growth rate, and K is the carrying capacity for the population under consideration. We call this the (r, K) parametrization for the logistic curve. Alternatively, we may modify Equation (11) to obtain the (A, B) parametrization for the logistic curve given by

$$\frac{dP}{dt}(t) = AP(t) - BP(t)^2. \quad (12)$$

Note that if we set $A = r$ and $B = \frac{r}{K}$ we see that the models are equivalent. Similarly, we can modify Equation (11) to obtain the equivalent (C, D) parametrization for the logistic curve as

$$\frac{dP}{dt}(t) = CP(t)(D - P(t)), \quad (13)$$

which contains the parameters $C = \frac{r}{K}$ and $D = K$. Throughout this presentation we will examine the system using the parameter set $(r, K) = (50, 10)$ and initial condition $P(0) = 1$. Table 1 contains plots of the logistic equation with these parameter values.

3.2. Richards

The second model that we consider is the Richards Curve, which also describes the growth of a bounded, dynamically changing population, here denoted by $f(t)$. Letting κ be the population growth rate, α be the carrying capacity, and δ a free parameter, the (κ, δ) parametrization of the Richards curve is given by

$$\frac{df}{dt} = \frac{\kappa}{1 - \delta} f \left[\left(\frac{f}{\alpha} \right)^{\delta - 1} - 1 \right], \quad \delta \neq 1. \quad (14)$$

We note that for $\delta = 2$ this equation is exactly the logistic equation with $r = \kappa$ and $K = \alpha$, and that the Richards curve is also a generalization of other population models, such as the Gompertz equation. We also note that $\kappa = \eta(1 - \delta)\alpha^{\delta - 1}$ where η is a growth parameter.

Now, if we let $A = \frac{\kappa}{1 - \delta}$ and $B = \delta$, then we can rewrite Equation (14) as

$$\frac{df}{dt} = Af \left[\left(\frac{f}{\alpha} \right)^{B - 1} - 1 \right], \quad B \neq 1. \quad (15)$$

We call this the (A, B) parametrization for the Richards curve.

Throughout this work, we let $f(0) = 1$, fix $\alpha = 8$, and only try to estimate (κ, δ) for Equation (14) or (A, B) for Equation (15). We will examine the system using the parameter set $(\kappa, \delta) = (2, 2)$, and thus the true dynamics are similar to those of the logistic equation with $K = 2$ and $r = 2$, but now instead of trying to estimate the carrying capacity we're trying to estimate the inflection of the growth. Table 1 contains plots of the Richards curve with these parameter values.

3.3. Spring Equation

The third model we consider is the spring-mass-dashpot system with mass m , damping coefficient c , and spring constant k . If $C = c/m$ and $K = k/m$, then the oscillating spring (in a standard engineering and mathematical formulation) may be modeled as

$$\begin{aligned} \frac{d^2x(t)}{dt^2} + C \frac{dx(t)}{dt} + Kx(t) &= 0, \\ x(t_0) = x_0, \dot{x}(t_0) &= v_0, \end{aligned} \quad (16)$$

which we will call the (C, K) parametrization for the spring equation. In this project, we take $x(t_0) = 1$ and $\dot{x}(t_0) = 0$, with $t_0 = 0$.

Because Equation (16) is a homogeneous second order differential equation, we can solve it analytically. This analytic solution will have the form

$$x(t) = e^{-\frac{C}{2}t} (A \cos(\omega t) + B \sin(\omega t)), \quad (17)$$

which can be further modified [19] to obtain

$$x(t) = e^{-\frac{C}{2}t} (R \sin(\omega t + \delta)), \quad (18)$$

where δ (or A and B) can be determined from the initial conditions.

By solving the characteristic polynomial of Equation (16), we find that

$$\omega = \frac{\sqrt{4K - C^2}}{2}, \quad (19)$$

or that, equivalently,

$$K = \omega^2 + \frac{C^2}{4}. \quad (20)$$

And thus we can reformulate Equation (16) to

$$\frac{d^2x(t)}{dt^2} + C \frac{dx(t)}{dt} + \left(\omega^2 + \frac{C^2}{4} \right) x(t) = 0, \quad (21)$$

which we will call the (C, ω) parametrization for the spring equation. Throughout this presentation, we will examine this system using the parameter set $(C, K) = (1/4, 1)$. Table 1 contains plots of the spring equation with these parameter values.

3.4. Adding Noise

For each model above, we simulated the model with the chosen parameter values and solved for the model solution at evenly spaced times as data points in a given chosen interval $[0, T]$ for each model (13 data points for the logistic curve, 21 for the Richards and spring, respectively, as depicted in Table 1). These data points are equivalent to $f(t_j, \theta_0)$ in Equation (3). In order to obtain our “observed” data points y_j , we then took the points $f(t_j, \theta_0)$ and perturbed them by realizations of the noise \mathcal{E}_j from a normal distribution with mean 0 and variance $\sigma_0^2 = nl^2$ where $nl = 0.01, 0.05$, and 0.2 . The initial condition was not perturbed by noise, and thus $y_0 = f(0, \theta_0)$. Table 1 depicts the forward solution $f(t_j, \theta_0)$ for each of the models expressed in Section 3, as well as the data points y_j collected given the various noise levels.

4. Results

4.1. Covariance Results

Table 2 contains the estimated covariance matrices (Equation (5)) for each of the models considered for each noise level. We first inspect these matrices to understand what information might be available in them. Recalling the parameter values used to generate the data (for the logistic curve $r = 50, K = 10$, for the Richards curve $\kappa = \delta = 2$ and for the spring equation $C = 1/4, K = 1$), we see that the variances of the estimators are small relative to the actual parameter values. This suggests that we are able to estimate the parameters with reasonable to high confidence. More precisely, Table 3 presents the parameter estimates for each model and the corresponding standard errors $SE(\hat{\theta}_k)$, which we see are very small compared to the parameter estimates.

We next turn to the correlation coefficients given in Table 4, and study these for the degree, if any, of correlation in the estimators. For some of the model parametrizations it is clear that there is correlation in the estimators as represented by the magnitude of the correlation coefficient (close to ± 1). These are the (A, B) parametrizations for the logistic and Richards curves, as well as the (κ, δ) parametrization for the Richards curve. There are also some parametrizations (e.g., both parametrizations of the spring equation) where it is clear that there is very little, if any, correlation in the estimators as represented by the magnitude of the correlation coefficient (close to 0). At the extremes (near ± 1) correlation coefficients are easy to read, but for intermediate values

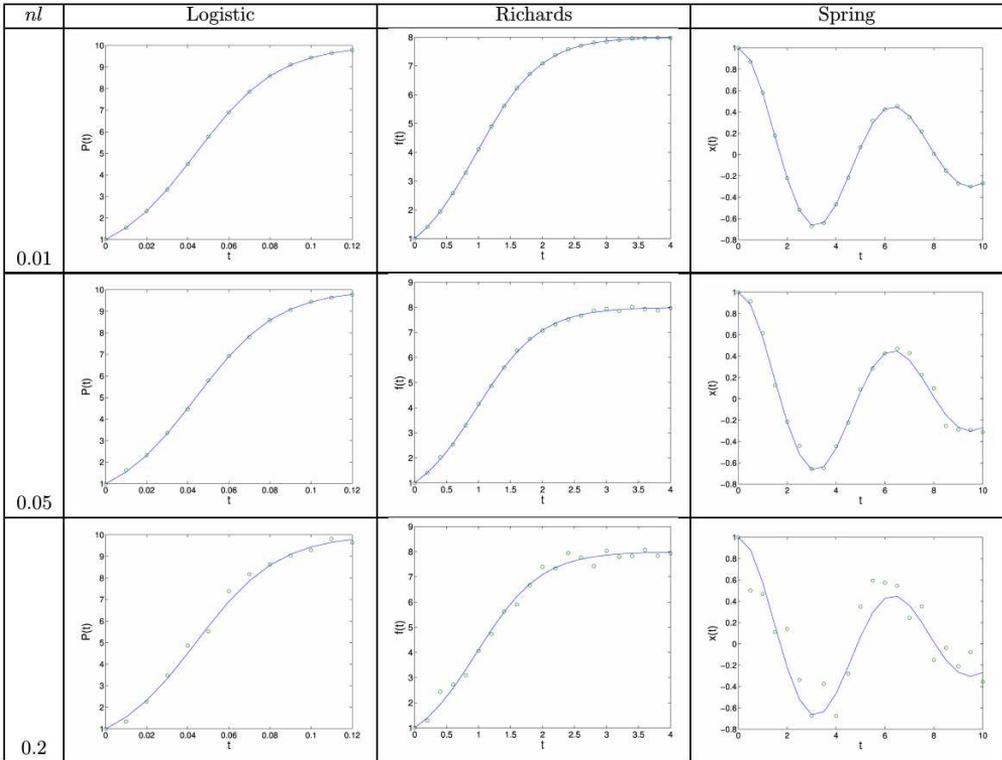


Table 1: The forward solutions $f(t, \theta_0)$ of the models using the parameters specified in Section 3, along with data y_j collected at evenly spaced time points t_j , and perturbed by noise \mathcal{E}_j from a normal distribution with standard deviation nl . Each forward solution is plotted as a solid line, while the data (with noise added) are plotted as open circles. Each column corresponds to a different model, while each row corresponds to a different noise level.

it is more difficult. Therefore, for the (r, K) and (C, D) parametrizations for the logistic curve, it is difficult to ascertain whether or not the values represent strong correlation in the estimators. Typically, accepted values of the correlation coefficient that determine significant correlation differ from discipline to discipline, and we will avoid a discussion on that here. To further interpret our correlation coefficient results, we will compare these to parameter correlation results from alternative methods in the next section.

	Logistic (r, K)		Logistic (A, B)		Logistic (C, D)	
nl = 0.01	0.0023	-0.0002	0.0023	0.0003	0.5606	-0.4351
	-0.0002	$0.4230 \cdot 10^{-4}$	0.0003	0.0001	-0.4351	0.4230
					$\cdot 10^{-4}$	
nl = 0.05	0.0362	-0.0035	0.0362	0.0054	0.8769	-0.6752
	-0.0035	0.0007	0.0054	0.0009	-0.6752	0.6522
					$\cdot 10^{-3}$	
nl = 0.2	1.1892	-0.1004	1.1892	0.1736	0.0279	-0.0195
	-0.1004	0.0176	0.1736	0.0279	-0.0195	0.0176
	Richards (κ, δ)		Richards (A, B)			
nl = 0.01	0.0859	0.0979	0.1490	0.1301	0.1301	0.1161
	0.0979	0.1161	0.1301	0.1161		
					$\cdot 10^{-3}$	
nl = 0.05	0.0020	0.0023	0.0041	0.0033		
	0.0023	0.0027	0.0033	0.0027		
nl = 0.2	0.0389	0.0443	0.0565	0.0538		
	0.0443	0.0525	0.0538	0.0525		
	Spring (C, K)		Spring (C, ω)			
nl = 0.01	0.4700	0.1623	0.4700	0.0524	0.0524	0.1287
	0.1623	0.5401	0.0524	0.1287		
					$\cdot 10^{-5}$	
nl = 0.05	0.1039	0.0358	0.1039	0.0119	0.0119	0.0287
	0.0358	0.1187	0.0119	0.0287		
					$\cdot 10^{-3}$	
nl = 0.2	0.0027	0.0009	0.0027	0.0003		
	0.0009	0.0032	0.0003	0.0007		

Table 2: These are the covariance matrices for each parametrization for each model, and for each noise level. The first parameter corresponds to the first diagonal, and the second parameter to the second diagonal element. So, for example, the logistic (r, K) covariance matrix has the variance in the r estimate as the top left entry, and the variance in the K estimate in the bottom right entry.

4.2. Ellipsoidal and DRAM Results

We compared the results for the asymptotic ellipsoids, exact ellipsoids, and DRAM Monte Carlo simulations side by side in Tables 5 - 11. This is done for each model (each parametrization of the logistic, Richards, and spring curves) and for each noise level ($nl = 0.01, 0.05$, and 0.2). In order to better compare results, the plots for each model have the same axes across noise levels. Thus, for example, the logistic equation using the parameters (r, K) for noise level $nl = 0.01$ plots all have the same axes, but those axes will be different than those seen for the noise level $nl = 0.05$, or for the (A, B) parametrization. The axes used are always $\hat{\theta} \pm 4 \cdot \max SE(\hat{\theta}_k)$, where the standard errors $SE(\hat{\theta}_k)$ can

	nl=0.01		nl=0.05		nl=0.2	
Logistic	Estimate	SE	Estimate	SE	Estimate	SE
r	49.9802	0.0481	50.0581	0.1902	51.9379	1.0905
K	10.0026	0.0065	9.9830	0.0255	9.8877	0.1327
A	49.9802	0.0481	50.0581	0.1902	51.9379	1.0905
B	4.9967	0.0075	5.0144	0.0296	5.2528	0.1671
C	4.9967	0.0075	5.0144	0.0296	5.2528	0.1671
D	10.0026	0.0065	9.9830	0.0255	9.8877	0.1327
Richards	Estimate	SE	Estimate	SE	Estimate	SE
κ	2.0155	0.0093	1.9828	0.0447	2.0324	0.1972
δ	2.0167	0.0108	1.9798	0.0524	2.0553	0.2291
A	-1.9823	0.0122	-2.0238	0.0643	-1.9258	0.2376
B	2.0167	0.0108	1.9798	0.0524	2.0553	0.2291
Spring	Estimate	SE	Estimate	SE	Estimate	SE
c	0.2482	0.0022	0.2365	0.0102	0.2582	0.0523
K	1.0005	0.0023	0.9889	0.0109	1.0162	0.0564
c	0.2482	0.0022	0.2365	0.0102	0.2582	0.0523
ω	0.9925	0.0011	0.9874	0.0054	0.9998	0.0273

Table 3: Here we show the parameter estimates $\hat{\theta}_k$ and standard errors $SE(\hat{\theta}_k)$ for each parameterization for each model at each noise level considered in this manuscript. Each row corresponds to a different parameter. For example, the second to last row corresponds to the parameter c in the (c, ω) parametrization for the spring equation, with the first two columns corresponding to the noise level $nl = 0.01$, the next two columns corresponding to the noise level $nl = 0.05$, and the last two columns corresponding to the noise level $nl = 0.2$.

be found in Table 3.

Immediately we see that the exact ellipsoids and DRAM Monte Carlo plots contain the same information for each model and each noise level, while the asymptotic ellipsoids always appear as precise 2-D ellipses. This occurs because the asymptotic ellipsoids are based on the covariance matrices, which by definition embody linear relationships. As expected from the definitions, the sign of the slope of the major axis in the asymptotic ellipsoids is given by the sign of the correlation coefficients in Table 4.

We compared results from the asymptotic ellipsoids to the correlation coefficient results. The asymptotic ellipsoids confirm what was seen with the

	nl=0.01	nl=0.05	nl=0.2
Logistic (r, K)	-0.7152	-0.7143	-0.6939
Logistic (A, B)	0.9529	0.9529	0.9527
Logistic (C, D)	-0.8935	-0.8928	-0.8798
Richards (κ, δ)	0.9803	0.9798	0.9808
Richards (A, B)	0.9891	0.9899	0.9882
Spring (c, K)	0.3222	0.3222	0.3182
Spring (c, ω)	0.2131	0.2181	0.2047

Table 4: Here we show the correlation coefficients ρ_{ij} for each parametrization of each model at each noise level considered in this paper. Each column corresponds to a different noise level, while each row corresponds to a different model parametrization.

correlation coefficients for the Richards curve and the spring equation. That is, the asymptotic ellipsoids for the Richards curve are diagonal, which suggests correlation in the estimators, while the asymptotic ellipsoids for the spring curve are not, which suggests no or little correlation in the estimators. For the logistic curve (r, K) parametrization, the asymptotic ellipsoid is vertical with an incredibly mild tilt, which would suggest little or no correlation in the estimators. When looking at the correlation coefficients we weren't able to accurately determine whether there was strong correlation or not. For the logistic curve (C, D) parametrization, we see that the asymptotic ellipsoid has a pronounced diagonal tilt, which strongly suggests correlation in the estimators. The (A, B) parametrization of the logistic curve has a slightly sloped diagonal in the asymptotic ellipsoids, which suggests a slight correlation, while the correlation coefficients showed clear correlation. This confirms that the asymptotic ellipsoids provide a quick visualization of the information contained in the covariance matrices.

While the asymptotic ellipsoids and covariance matrices give information on the linear relationship of the estimators, the exact ellipsoids and DRAM Monte Carlo plots visually reveal any non-linear relationship of the estimators. In the results depicted here, a non-linear relationship only appears in the results for the (A, B) parametrization of the Richards curve, and it shows up very clearly in the exact ellipsoids and DRAM Monte Carlo plots as a banana shape (Table 9). It can be very difficult to determine a priori any non-linear relationship in the estimators, but the exact ellipsoids and DRAM methods are able to accurately reveal when there is.

In addition to the qualitative/quantitative information contained in each plot,

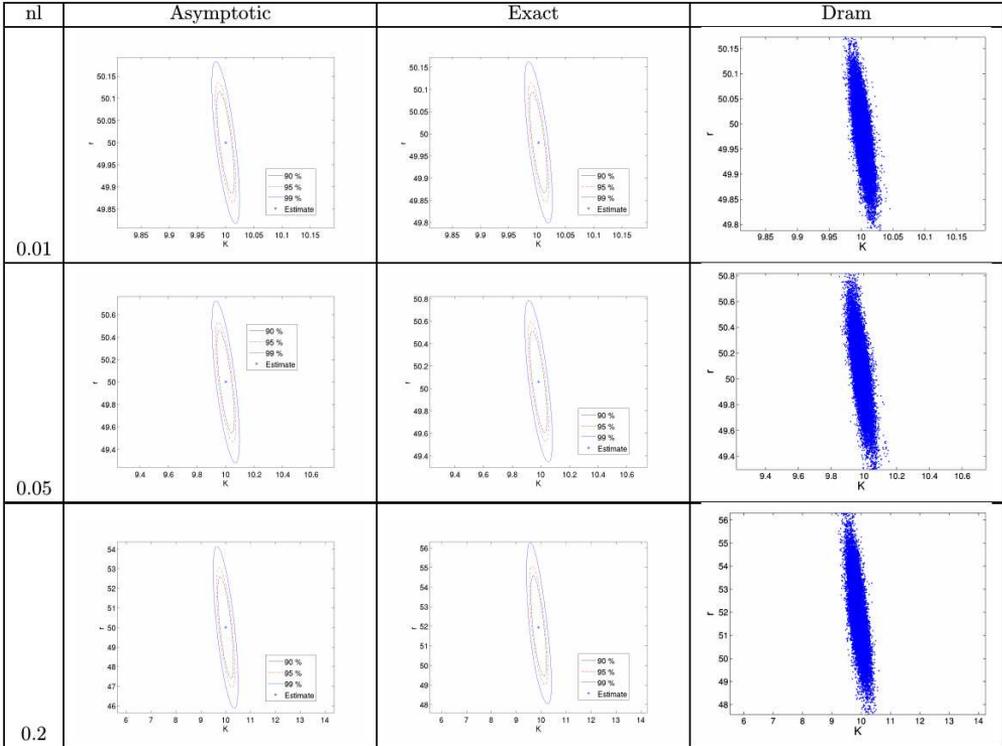


Table 5: We plot the results for the logistic curve with the (r, K) parametrization. The left column of plots are the asymptotic ellipsoids, the middle column of plots are the exact ellipsoids, and the right column of plots are the DRAM Monte Carlo plots. The top row of plots correspond to noise level $nl = 0.01$, the middle row to noise level $nl = 0.05$, and the bottom row to noise level $nl = 0.2$.

it might be of interest to consider the computational times and/or ease of implementation required for each method. However, since we made no efforts to optimize our codes to minimize computational times, number of functional evaluations, etc., such comparisons might be relatively meaningless. Suffice it to say that the findings reported for either the asymptotic or exact confidence regions required less effort than did the DRAM based results.

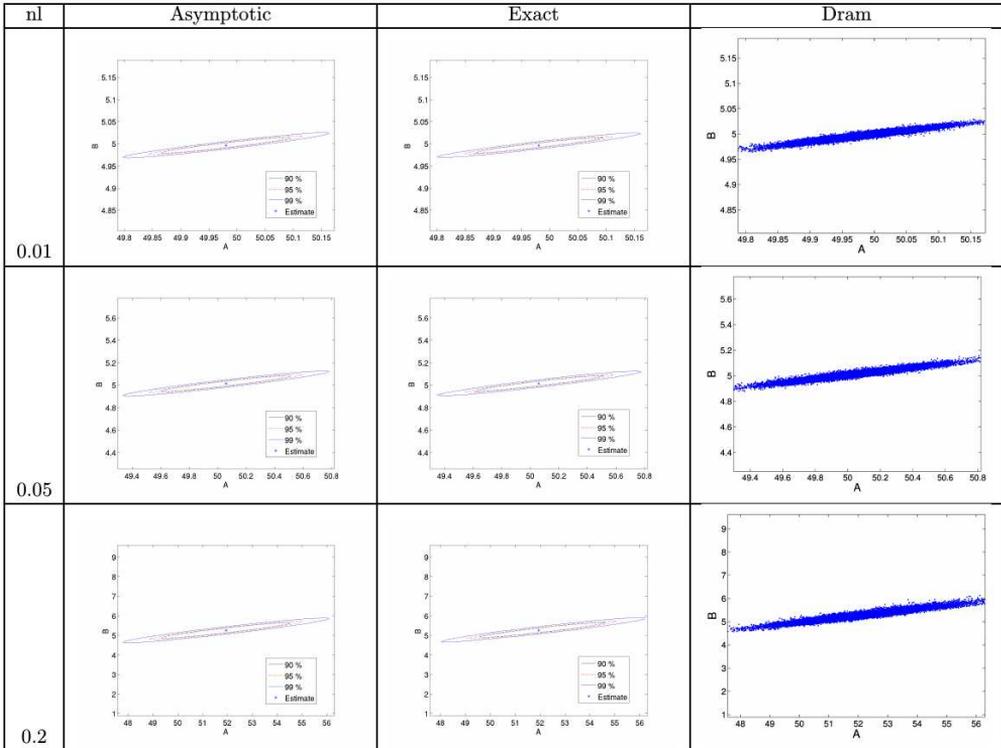


Table 6: We present the results for the logistic curve with the (A, B) parametrization. The left column of plots are the asymptotic ellipsoids, the middle column of plots are the exact ellipsoids, and the right column of plots are the DRAM Monte Carlo plots. The top row of plots correspond to noise level $nl = 0.01$, the middle row to noise level $nl = 0.05$, and the bottom row to noise level $nl = 0.2$.

5. Discussion and Concluding Remarks

The considerations presented here arose from the simple question of what, if anything, we can say about dependence/independence of parameter estimators arising in inverse problems using simple computational methods related to estimator covariances. Initial findings with several simple one dimensional systems help to clarify the issues involved.

From the asymptotic theory-based covariances alone (see Table 2) which employ the linearized sensitivities in their construction, little definitive information can be obtained. If one computes the corresponding correlation coeffi-

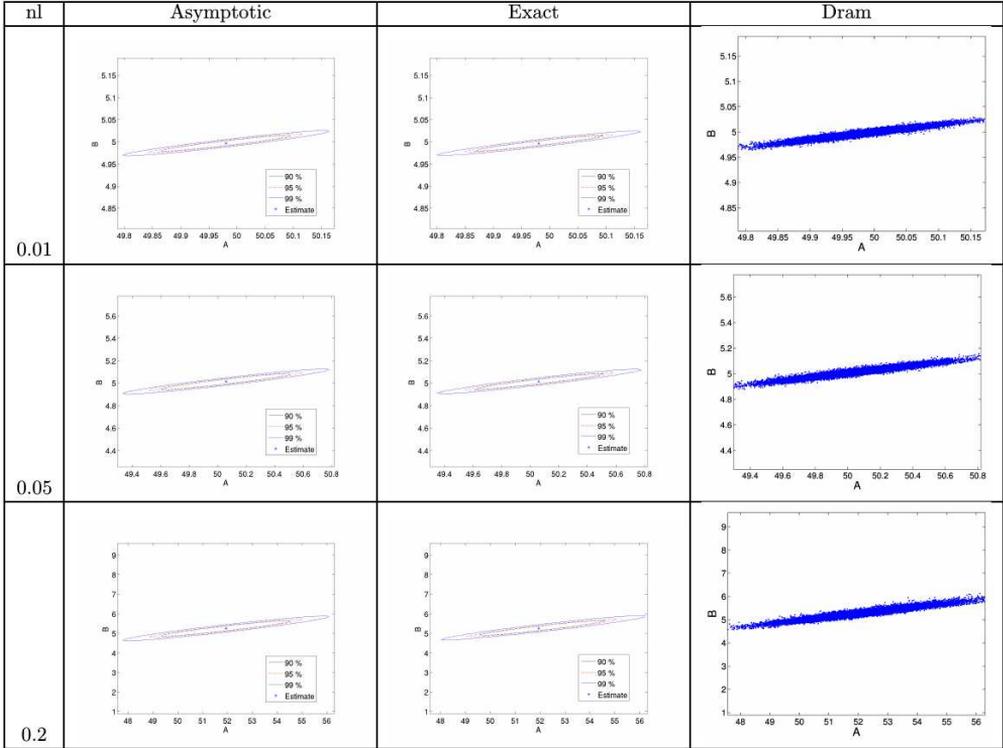


Table 7: We plot the results for the logistic curve with the (C, D) parametrization. The left column of plots are the asymptotic ellipsoids, the middle column of plots are the exact ellipsoids, and the right column of plots are the DRAM Monte Carlo plots. The top row of plots correspond to noise level $nl = 0.01$, the middle row to noise level $nl = 0.05$, and the bottom row to noise level $nl = 0.2$.

cients, then in the extreme cases of ρ near zero (suggesting independence) or near ± 1 (dependence) some useful information can be obtained about *linear* dependence. If one uses this along with the asymptotic based ellipsoids, information content can be enhanced in the non-extreme cases. If one turns to the exact ellipsoids (which are not computed using the linearized sensitivities), one can see not only linear dependence but a nonlinear dependence can be detected if present.

When we compared these methods to the Bayesian DRAM algorithm, we obtain the above mentioned information in addition to other information on the prediction intervals, parameter distributions, etc. [14, 15]. The downside is

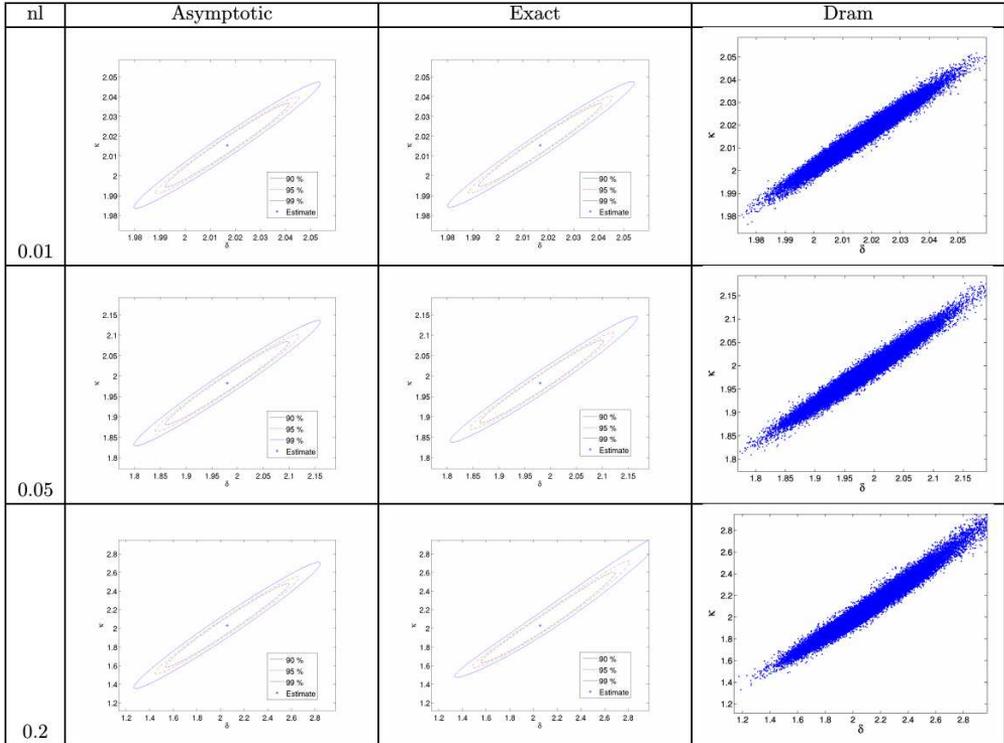


Table 8: We plot the results for the Richards curve with the (κ, δ) parametrization. The left column of plots are the asymptotic ellipsoids, the middle column of plots are the exact ellipsoids, and the right column of plots are the DRAM Monte Carlo plots. The top row of plots correspond to noise level $nl = 0.01$, the middle row to noise level $nl = 0.05$, and the bottom row to noise level $nl = 0.2$.

this takes by far the longest to compute (at least for the examples considered here). Therefore, if all one needs is information about the correlation (or lack thereof) in parameter estimators (with an eye toward reparametrization), then we would not recommend using the DRAM algorithm in a first effort.

One should recognize that effort required in obtaining correlation information can be important factors in larger problems and that run times, and more meaningful measures such as function evaluations, ease in implementation, etc., might be of interest. In that regard, it is important to note that there is a parallel implementation of DRAM [25], and also that the computation of the exact ellipsoids can be easily made parallel (in this note we only used serial imple-

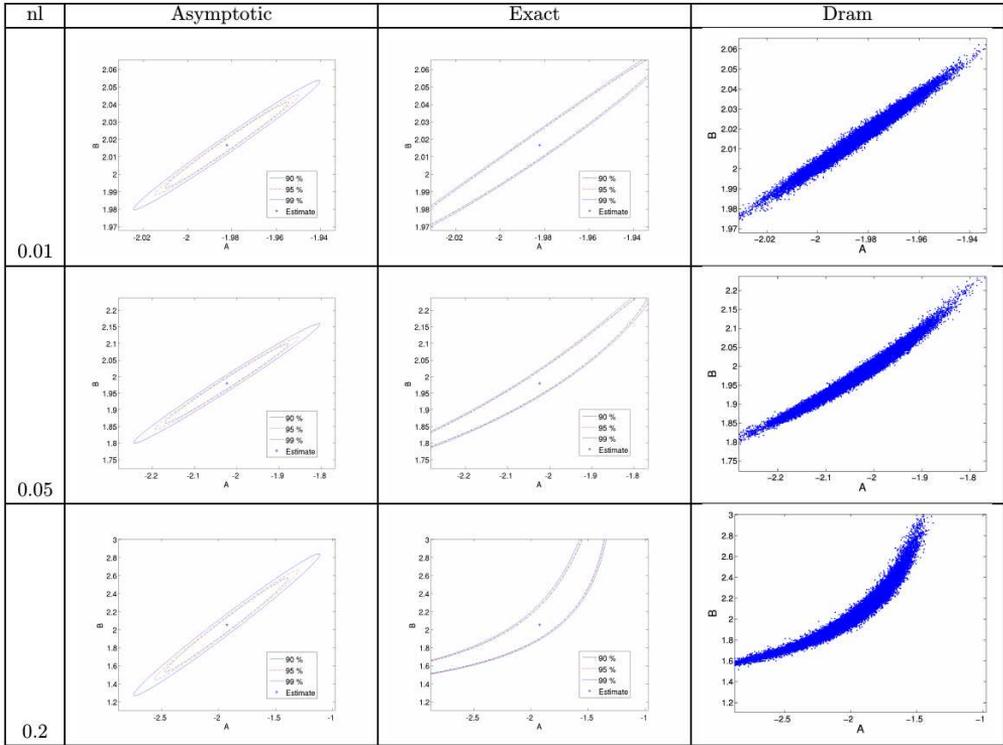


Table 9: We plot the results for the Richards curve with the (A, B) parametrization. The left column of plots are the asymptotic ellipsoids, the middle column of plots are the exact ellipsoids, and the right column of plots are the DRAM Monte Carlo plots. The top row of plots correspond to noise level $nl = 0.01$, the middle row to noise level $nl = 0.05$, and the bottom row to noise level $nl = 0.2$.

mentations). If one considers more elaborate and computationally challenging mathematical models (e.g., such as the nonlinear models for HIV progression [4, 10], models for cell proliferation [7], climate models [23], nonlinear viscoelasticity [17], immune response in transplant patients [6]), then computational costs for the Bayesian approach [17, 23] could be an important consideration.

There are many limitations in our preliminary investigations: while the present study concerns itself exclusively with longitudinal/repeated measurements [12] data, these preliminary insights are not limited to that situation; one only needs a system in which the assumptions of the asymptotic theory and Bayesian based DRAM apply. Moreover, we have not tried to apply op-

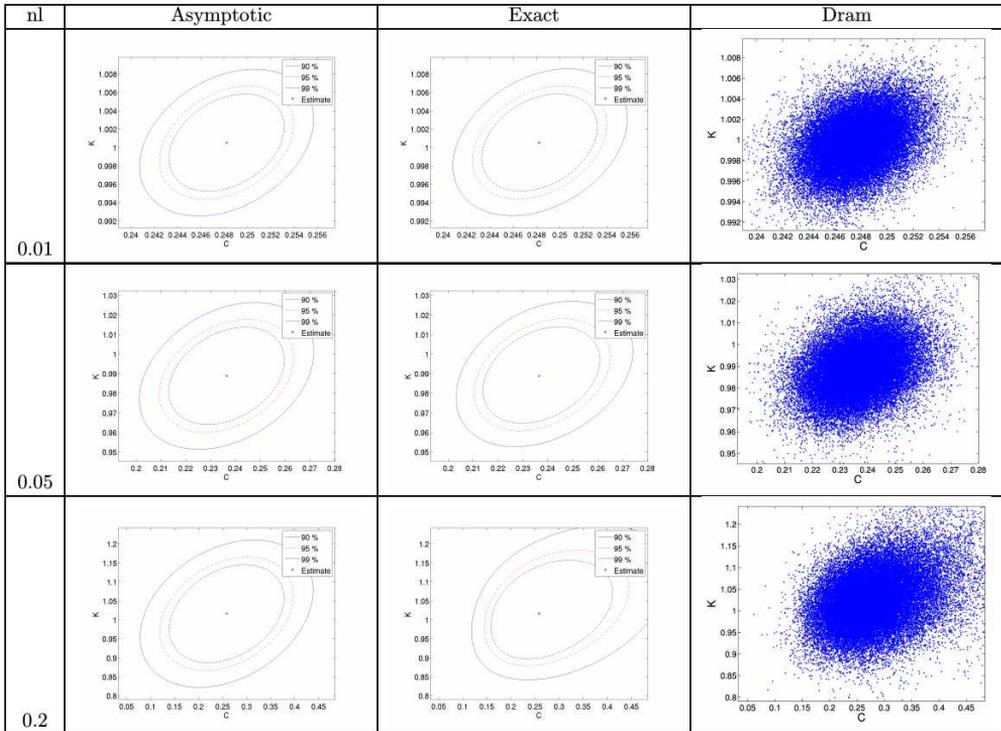


Table 10: We plot the results for the spring equation with the (C, K) parametrization. The left column of plots are the asymptotic ellipsoids, the middle column of plots are the exact ellipsoids, and the right column of plots are the DRAM Monte Carlo plots. The top row of plots correspond to noise level $nl = 0.01$, the middle row to noise level $nl = 0.05$, and the bottom row to noise level $nl = 0.2$.

timal design techniques [1, 7, 8, 9] to obtain really efficient computational schemes for the associated inverse problems in our comparisons. Finally we have not addressed at all the motivating issues [22, Chapter 3] of identifiability, ill-conditioning, and reparameterization for systems in which the parameteric estimators are found to exhibit dependence.

Acknowledgements

This research was supported in part by the Air Force Office of Scientific Research under grant number AFOSR FA9550-12-1-0188, in part by the National

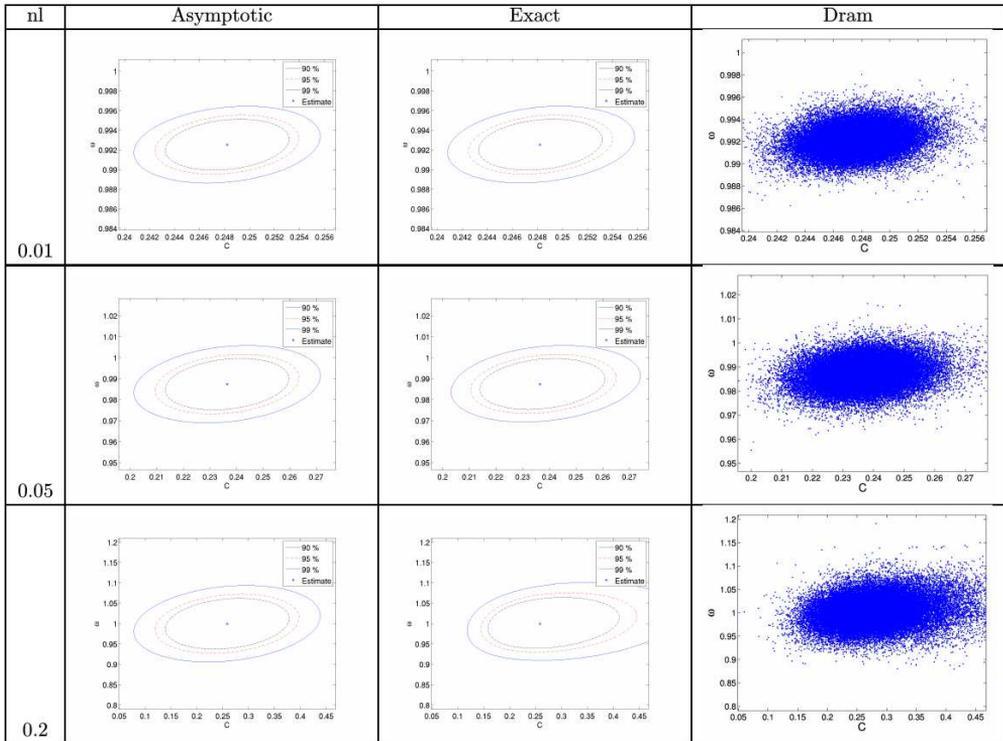


Table 11: We plot the results for the spring equation with the (C, ω) parametrization. The left column of plots are the asymptotic ellipsoids, the middle column of plots are the exact ellipsoids, and the right column of plots are the DRAM Monte Carlo plots. The top row of plots correspond to noise level $nl = 0.01$, the middle row to noise level $nl = 0.05$, and the bottom row to noise level $nl = 0.2$.

Science Foundation under Research Training Grant (RTG) DMS-1246991, NSF Undergraduate Biomathematics grant number DBI-1129214, and NSF grant number DMS-0946431.

References

- [1] Kaska Adoteye, H.T. Banks and Kevin B. Flores, Optimal design of non-equilibrium experiments for genetic network interrogation, CRSC-TR14-12, N. C. State University, Raleigh, NC, September, 2014; *Applied Math Letters*, **40** (2015), 84–89.

- [2] C. Andrieu and J. Thomas, A tutorial on adaptive MCMC, *Statistics and Computing*, **18** (2008), 343–373.
- [3] A. Cintron-Arias, H.T. Banks, A. Capaldi and A. Lloyd, A sensitivity matrix based methodology for inverse problem formulation, *J. Inverse Ill-Posed Problems*, **17** (2009), 545–564.
- [4] H.T. Banks, R. Baraldi, K. Cross, C. McChesney, L. Poag, E. Thorpe, and K.B. Flores, Uncertainty quantification in modeling HIV viral mechanics, CRSC-TR13-16, N. C. State University, Raleigh, NC, December, 2013; *Mathematical Biosciences and Engineering*, **12** (2015), **12** (2015), 937–964.
- [5] H.T. Banks and B.G. Fitzpatrick, Statistical methods for model comparison in parameter estimation problems for distributed systems, *J. Math. Biol.*, **28** (1990), 501–527.
- [6] H.T. Banks, S. Hu, K. Link, E.S. Rosenberg, S. Mitsuma and L. Rosario, Modeling immune response to BK virus infection and donor kidney in renal transplant recipients, CRSC-TR14-09, N. C. State University, Raleigh, NC, June, 2014; *J. Inverse Problems in Science and Engineering*, DOI:10.1080/17415977.2015.1017484.
- [7] H.T. Banks, Shuhua Hu, and W. Clayton Thompson, *Modeling and Inverse Problems in the Presence of Uncertainty*, CRC Press, Boca Raton, FL, 2014.
- [8] H.T. Banks and K.L. Rehm, Experimental design for distributed parameter vector systems, CRSC-TR12-17, N. C. State University, Raleigh, NC, August, 2012; *Applied Mathematics Letters*, **26** (2013), 10–14; <http://dx.doi.org/10.1016/j.aml.2012.08.003>.
- [9] H.T. Banks and K.L. Rehm, Parameter estimation in distributed systems: Optimal design, CRSC TR14-06, N. C. State University, Raleigh, NC, May, 2014; *Eurasian Journal of Mathematical and Computer Applications*, **2** (2014), 70–79.
- [10] H.T. Banks and H.T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes*, CRC Press, Boca Raton, FL, 2009.
- [11] Y. Chu and J. Hahn, Parameter set selection for estimation of nonlinear dynamic systems, *AIChE Journal*, **53** (2007), 2858–2870.

- [12] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman Hall/CRC Monographs on Statistics and Applied Probability, (1995).
- [13] A.R. Gallant, *Nonlinear Statistical Models*, Wiley, New York, 1987.
- [14] H. Haario, M. Laine, A. Mira, and E. Saksman, DRAM: Efficient adaptive MCMC, *Statistics and Computing*, **16.4** (2006), 339–354.
- [15] H. Haario, E. Saksman, and J. Tamminen, An adaptive Metropolis algorithm, *Bernoulli*, **7.2** (2001), 223–242.
- [16] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, New York, 2005.
- [17] Zackary R. Kenz, H. T. Banks, and Ralph C. Smith. Comparison of frequentist and Bayesian confidence analysis methods on a viscoelastic stenosis model. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1) (2013) 348–369.
- [18] Joaquim R. R. A. Martins, Ilan M. Kroo, and Juan J. Alonso, An automated method for sensitivity analysis using complex variables, *Proc. 38th Aerospace Sciences Meeting*, paper AIAA-2000-0689, Amer. Inst. Aeronautics and Astronautics (2000), pp. 1–12.
- [19] R. Kent Nagle, Edward Saff, and Arthur Snider, *Fundamentals of Differential Equations and Boundary Value Problems*, Fourth edition, Pearson Education, Boston San Francisco New York, 2004.
- [20] T.J. Rothenberg, Identification in parametric models, *Econometrica*, **39** (1971), 577-591.
- [21] G.A.F. Seber and A.J. Lee, *Linear Regression Analysis*, Wiley, Hoboken, 2003.
- [22] G.A. Seber and C.J. Wild, *Nonlinear Regression*, Wiley, Hoboken, 2003.
- [23] R.C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, Philadelphia, 2014.
- [24] A. Solonen, *Monte Carlo Methods in Parameter Estimation of Nonlinear Models*, MS Thesis, Lappeenranta University of Technology, Lappeenranta, Finland, 2006.

- [25] A. Solonen, P. Ollinaho, M. Laine, H. Haario, J. Tamminen and H. Järvinen, Efficient MCMC for climate model parameter estimation: parallel adaptive chains and early rejection, *Bayesian Analysis*, **7** (2012), 715–736.
- [26] Gilbert Strang, *Introduction to Linear Algebra, 3rd. Ed.*, Wellesley-Cambridge Press, Wellesley, MA, 2003.