Standard Errors and Confidence Intervals in Inverse Problems: Sensitivity and Associated Pitfalls

H.T. Banks, Stacey L. Ernstberger and Sarah L. Grove

Center for Research in Scientific Computation
North Carolina State University
Raleigh, NC 27695-8205

March 2, 2006

Abstract

We review the asymptotic theory for standard errors in classical ordinary least squares (OLS) inverse or parameter estimation problems involving general nonlinear dynamical systems where sensitivity matrices can be used to compute the asymptotic covariance matrices. We discuss possible pitfalls in computing standard errors in regions of low parameter sensitivity and/or near a steady state solution of the underlying dynamical system.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
02 MAR 2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Standard Errors and Confidence Intervals in Inverse Problems: Sensitivity and Associated Pitfalls	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Center for Research in Scientific Computation,North Caroline State University,Raleigh,NC,27695-8205**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**27**	

1 Introduction

There is a growing interest in measures of uncertainty in many areas of mathematical modeling as more sophisticated models are employed in biology [4, 3, 8], chemistry, sociology [7], etc., as well as in the usual physical and engineering sciences. In particular, there has been recent renewed efforts by applied mathematicians in the area of inverse or parameter estimation problems to attach measures of reliability to estimated quantities using experiment data sets. Statisticians have for some time [9, 11, 15, 16, 19] recognized the importance of such problems and have developed theory and methodology to treat these. In particular, a rather extensive area of maturity is the computation of standard errors (SE) and confidence intervals (CI) for ordinary least squares (OLS) problems using what engineers and applied mathematicians call *sensitivity matrices* [1, 10, 12, 13, 14, 18]. These can be computed or approximated in a number of ways and for quite sophisticated unknown "parameters" including functions and even probability measures [2, 3, 5, 6]. A most frequently used framework for the computation of SE is the *asymptotic theory of estimators* which, roughly speaking, guarantees under certain conditions that as one uses more and more data in the inverse problems, that one can approximate the estimator or sampling distribution by a Gaussian with computable mean and covariance. As with most statistical theories, this asymptotic theory holds under quite specific hypotheses on the statistical model assumed for the observation error (which may include both model and measurement error) as well as assumptions on regularity of the underlying problem and *the method in which the increasing amount of data is collected*. In practice, most investigators are concerned with the form of the error (e.g., independent identically distributed with constant variance or constant coefficient of variation) but do not give much concern to either the regularity hypotheses or the method of data collection. This can sometimes lead to surprising and seemingly perplexing results for unwary users. The purpose of this note is to recall these results, discuss some pitfalls, and illustrate with an example that shares qualitative features with many systems encountered in wide-spread scientific applications. We focus specifically on the phenomenon that increased sampling data often does not result in improvements in the estimates or in their associated statistical reliability as embodied in their SE or CI. For example, when additional data is sampled from a region near an equilibrium or steady state, the estimated values of the problem generally do not improve. In fact, the confidence intervals for

estimated values often increase in size as more data is taken. To discuss the causes for this, we will illustrate ideas with the logistic growth population model.

2 Summary of asymptotic theory for errors

We first give a general summary of the asymptotic theory for standard errors. We assume that n scalar longitudinal observations (the extension to vectors is completely straight-forward) are represented by the statistical model

$$Y_j \equiv f_j(\beta_0) + \epsilon_j, \;\; j = 1, 2, \ldots, n, \tag{1}$$

where $f_j(\beta_0)$ is the model for the observations in terms of the state variables and $\beta_0 \in \mathbb{R}^p$ is a "set" of theoretical "true" parameter values (assumed to exist in a standard statistical approach). For example, if one is given a differential equation system $\dot{x} = \Gamma(t, x(t), \beta)$, sampled at times $t_j, j = 1, 2, \ldots, n$, then $f_j(\beta) = x(t_j, \beta)$.

We assume for our statistical model of the observation or measurement process (1) that the errors ϵ_j, $j = 1, 2, \ldots, n$, are independent identically distributed (*i.i.d.*) random variables with mean $E[\epsilon_j] = 0$ and constant variance $var[\epsilon_j] = \sigma_0^2$, where of course σ_0^2 is unknown. We then have that the observations Y_j are *i.i.d.* with mean $E[Y_j] = f_j(\beta_0)$ and variance $var[Y_j] = \sigma_0^2$.

We consider estimation of parameters using an ordinary least squares (OLS) approach. Thus we seek to use data $\{y_j\}$ for the observation process $\{Y_j\}$ with the model to seek a value $\hat{\beta}^n$ that minimizes

$$J_n(\beta) = \sum_{j=1}^{n} |f_j(\beta) - y_j|^2. \tag{2}$$

Since Y_j is a random variable, we have that the estimator $\hat{\beta}^n_{OLS}$ is also a random variable with a distribution called the *sampling distribution*. Knowledge of this sampling distribution provides uncertainty information (e.g., standard errors) for the numerical values of $\hat{\beta}^n$ obtained using a specific data set $\{y_j\}$ (i.e., a realization of $\{Y_j\}$) when minimizing $J_n(\beta)$.

Under reasonable assumptions on smoothness and regularity (the smoothness requirements for model solutions are readily verified using continuous

dependence results for differential equations in most examples; the regularity requirements include, among others, conditions on how the observations are taken (more on this later) as sample size increases, i.e., $n \to \infty$), the standard nonlinear regression approximation theory ([11], [15], [16], and Chapter 12 of [19]) for asymptotic (as $n \to \infty$) distributions can be invoked. This theory yields that the sampling distribution $\hat{\beta}^n(Y)$ for the estimate $\hat{\beta}^n$, where $Y = \{Y_j\}_{j=1}^n$, is approximately a p-multivariate Gaussian with mean $E[\hat{\beta}^n(Y)]$ and covariance matrix $cov[\hat{\beta}^n(Y)] \approx \Sigma_0 = \sigma_0^2[\chi^T(\beta_0)\chi(\beta_0)]^{-1}$. Here $\chi(\beta) = F_\beta(\beta)$ is the $n \times p$ sensitivity matrix with elements

$$\chi_{jk}(\beta) = \frac{\partial f_j(\beta)}{\partial \beta_k} \qquad \text{and} \qquad F_\beta(\beta) \equiv (f_{1\beta}(\beta), \ldots, f_{n\beta}(\beta))^T.$$

That is, for n large, the sampling distribution approximately satisfies

$$\hat{\beta}^n_{OLS}(Y) \sim \mathcal{N}_p(\beta_0, \sigma_0^2[\chi^T(\beta_0)\chi(\beta_0)]^{-1}) := \mathcal{N}_p(\beta_0, \Sigma_0). \tag{3}$$

There are typically several ways to compute the matrix F_β. First, the elements of the matrix $\chi = (\chi_{jk})$ can always be estimated using the forward difference

$$\chi_{jk}(\beta) = \frac{\partial f_j(\beta)}{\partial \beta_k} \approx \frac{f_j(\beta + h_k) - f_j(\beta)}{h_k},$$

where h_k is an p-vector with nonzero entry in only the k^{th} component. Alternatively, if the $f_j(\beta)$ correspond to longitudinal evaluations $x(t_j, \beta)$ of solutions $x \in \mathbb{R}^{n^*}$ to a parameterized n^*-vector differential equation system $\dot{x} = \Gamma(t, x(t), \beta)$, then one can use the $n^* \times p$ matrix sensitivity equations (see [6] and the references therein)

$$\frac{d}{dt}\left(\frac{\partial x}{\partial \beta}\right) = \frac{\partial \Gamma}{\partial x}\frac{\partial x}{\partial \beta} + \frac{\partial \Gamma}{\partial \beta}$$

to obtain

$$\frac{\partial f_j(\beta)}{\partial \beta_k} = \frac{\partial x(t_j, \beta)}{\partial \beta_k}.$$

Finally, in some cases the function $f_j(\beta)$ may be sufficiently simple so as to allow one to derive analytical expressions for the components of F_β. This is the case for the problems addressed in this paper and hence for our efforts we chose this latter approach in the examples below.

Since β_0, σ_0 are not known, we must approximate them in

$$\Sigma_0 = \sigma_0^2[\chi^T(\beta_0)\chi(\beta_0)]^{-1}.$$

For this we follow standard practice and use the approximation

$$\Sigma_0 \approx \Sigma(\hat{\beta}^n) = \hat{\sigma}^2[\chi^T(\hat{\beta}^n)\chi(\hat{\beta}^n)]^{-1} \tag{4}$$

where $\hat{\beta}^n$ is the parameter estimate obtained, and the approximation $\hat{\sigma}^2$ to σ_0^2 is given by

$$\sigma_0^2 \approx \hat{\sigma}^2 = \frac{1}{n-p}\sum_{j=1}^{n} |f_j(\hat{\beta}^n) - y_j|^2. \tag{5}$$

Standard errors to be used in confidence interval calculations are thus given by $SE_k(\hat{\beta}^n) = \sqrt{\Sigma_{kk}(\hat{\beta}^n)}$, $k = 1, 2, \ldots, p$ (see [9]).

Finally, in order to compute the confidence intervals (at the $100(1-\alpha)\%$ level) for the estimated parameters in our example, we define the confidence level parameters associated with the estimated parameters so that

$$P\{\hat{\beta}_k^n - t_{1-\alpha/2}SE_k(\hat{\beta}^n) < \beta_k^n < \hat{\beta}_k^n + t_{1-\alpha/2}SE_k(\hat{\beta}^n)\} = 1-\alpha, \tag{6}$$

where $\alpha \in [0,1]$, and $t_{1-\alpha/2} \in \mathbb{R}_+$. For a given α value (small, say $\alpha = .05$ for 95% confidence intervals), the critical value $t_{1-\alpha/2}$ is computed from the Student's t distribution t^{n-p} with $n-p$ degrees of freedom. The value of $t_{1-\alpha/2}$ is determined by $P\{T \geq t_{1-\alpha/2}\} = \alpha/2$ where $T \sim t^{n-p}$.

We note that when $f_j(\beta) = x(t_j, \beta)$ as mentioned above when one is taking longitudinal samples corresponding to solutions of a dynamical system, the $n \times p$ sensitivity matrix depends explicitly on where in time the observations are taken. Note also that the sensitivity matrix

$$\chi(\beta) = F_\beta(\beta) = \left(\frac{\partial f_j(\beta)}{\partial \beta_k}\right)$$

depends on the number n and nature (e,g.,how taken) of the sampling times $\{t_j\}$. Moreover, it is the matrix $[\chi^T\chi]^{-1}$ in (4) and the parameter $\hat{\sigma}^2$ in (5) that ultimately determine the SE and CI. At first investigation of (5), it appears that an increased number n of samples will drive $\hat{\sigma}^2$ (and hence the SE) to zero as long as this is done in a way to maintain a bound on the residual sum of squares in (5). But then we observe that the condition

number of the matrix $\chi^T\chi$ is also very important in these considerations and increasing the sampling could possibly adversely affect the inversion of $\chi^T\chi$. In this regard, we note that among the important hypotheses in the asymptotic statistical theory (see p. 571 of [19]) are that

$$\frac{1}{n}\chi^T(\beta)\chi(\beta) \to \mathcal{X}(\beta) \quad \text{as n} \to \infty$$

with $\mathcal{X}(\beta_0)$ **nonsingular**. It is this condition that is rather easily violated in practice when one is dealing with data from differential equation systems near an equilibrium or steady state. To illustrate this as well as related issues regarding *sensitivity* that are often ignored in statistical discussions, we turn to a simple example, the popular logistic growth or Verhulst-Pearl equation [17], where analytic expressions are readily found and used to discuss the ideas.

3 Logistic growth population models

We consider the dynamics

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right)$$

where K is the carrying capacity and r is the intrinsic growth rate. This is the Verhulst-Pearl logistic equation [17]. The exact solution is readily solved

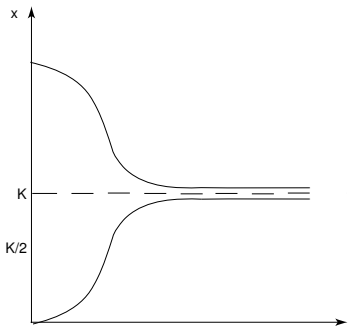

Figure 1: Solution to ODE

for and is given by

$$x(t) = \frac{K}{1 + \left(\frac{K}{x_0} - 1\right) e^{-rt}}.$$

This function has a flex point $\left(\frac{d^2x}{dt^2} = 0\right)$ at $\tilde{x} = \frac{K}{2}$.

We will examine the equivalent systems

$$\dot{x} = ax - bx^2, \tag{7}$$

which have solutions

$$\begin{aligned} x(t) &= \frac{a/b}{1 + (\frac{a/b}{x_0} - 1)e^{-at}} \\ &= \frac{a}{b + ke^{-at}} \end{aligned} \tag{8}$$

with $k = \frac{a}{x_0} - b$. The solutions $x(t)$ have an asymptote as $t \to \infty$ at $x^\infty = \frac{a}{b} = K$. Note that $a = r$ and hence $\frac{r}{K} = b$. When $x_0 > K$, the solutions $x(t)$ are monotone decreasing functions with an asymptote at K. Similar behavior (monotone increasing) is observed for solutions with $x_0 < K$.

Standard errors will be examined for this example using both analytical calculations as well as numerical computations. The analytical derivations will show that if observations are taken in specific regions pertaining to the solution curve, $\chi^T\chi$ could approach a singular matrix, resulting in the standard errors computed with the asymptotic theory becoming unbounded. We expect that the numerical computations performed using MATLAB will behave similar to the analytical results and therefore expect that the calculations will go badly if observations are not correctly chosen. This will reinforce the observation that in practical problems, sensitivity equations and standard errors can be used to inform experimental design (e.g., so that observations are taken appropriately to avoid such difficulties).

4 Analytical considerations

As stated earlier, we consider an ordinary least squares problem for the parameters $\beta = (a, b, x_0)$ with the approximate covariance matrix given by

$$\Sigma = \hat{\sigma}^2(\chi^T\chi)^{-1}$$

and the corresponding estimate of the standard error

$$\mathrm{SE}_j = \sqrt{\hat{\sigma}^2(\chi^T\chi)^{-1}_{jj}}, \qquad j = 1 \ldots n. \tag{9}$$

Recall that

$$\begin{aligned}
\Sigma &= \hat{\sigma}^2(\chi^T\chi)^{-1} \\
&= \frac{1}{n-p}\sum_{j=1}^{n}(f(t_j,\hat{\beta}) - y_j)^2(\chi^T\chi)^{-1} \\
&= \frac{1}{n-p}(\mathrm{RSS})(\chi^T\chi)^{-1}
\end{aligned}$$

where RSS denotes the residual sum of squares. In this example we have

$$\chi^T = \frac{\partial x}{\partial \beta}^T = \begin{pmatrix} \frac{\partial x(t_1)}{\partial a} & \cdots & \frac{\partial x(t_n)}{\partial a} \\ \frac{\partial x(t_1)}{\partial b} & \cdots & \frac{\partial x(t_n)}{\partial b} \\ \frac{\partial x(t_1)}{\partial x_0} & \cdots & \frac{\partial x(t_n)}{\partial x_0} \end{pmatrix}.$$

We will examine varying behavior in the model and the resulting statistical analysis depending on the region from which t_j is sampled. To this end, let R_0 be the region where $t \in [0, \tau_1]$, R_1 be the region corresponding to $t \in [\tau_1, \tau_2]$, and R_2 where $t \in [\tau_2, \infty)$ as depicted in Figure 2. We expect differences in the ability to estimate parameters depending on the region in which the solution is observed (sampled).

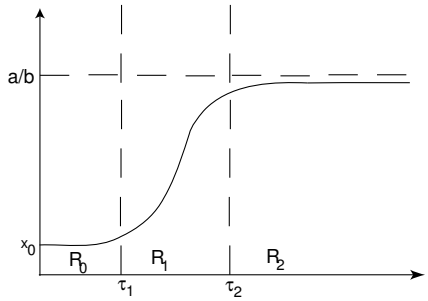

Figure 2: Partition of solution curve into distinct regions.

We will examine behavior and anticipate finding three distinct situations.

1. When data is sampled from region R_0, the solution is insensitive to the parameters a and b. If we attempt to estimate a or b from this data, we expect to obtain large standard errors as $\chi^T\chi$ approaches a singular matrix as $n \to \infty$. Thus, the parameters a and b, as well as the carrying capacity, $K = a/b$, cannot be estimated using data sampled from R_0 alone.

2. When sampling in R_1 we suspect that estimating β with decent standard errors is possible, and using more data (up to a point) will improve reliability (i.e., the corresponding SE).

3. When data is sampled from R_2, the solution is insensitive to x_0. If we try to estimate x_0 from this data, we expect to obtain large standard errors because $\chi^T\chi$ should again become ill-conditioned (approximately singular) as $n \to \infty$. We cannot estimate a and b independently, however we should be able to estimate the carrying capacity $K = a/b$ since $x(t) \to K = x^\infty$ as $t \to \infty$.

In order to consider the problem for data in different regions, we explicitly solve the equation

$$\dot{x} = ax - bx^2 \tag{10}$$

for x, and then examine the partial derivatives $\frac{\partial x}{\partial a}$, $\frac{\partial x}{\partial b}$, and $\frac{\partial x}{\partial x_0}$. Rewriting the equation as

$$\frac{dx}{x(a-bx)} = dt,$$

we obtain (by integrating and using partial fractions) the solution

$$x = \frac{a}{b + ke^{-at}}.$$

Thus, we readily see that $x(t) \to K = \frac{a}{b}$ as $t \to \infty$. Also note that $x(t) \to \frac{a}{b+k}$ as $t \to 0$. We will therefore denote $x(0) = x_0 = \frac{a}{b+k}$.

We easily compute the partial derivatives

$$\begin{aligned}
\frac{\partial x}{\partial a} &= \frac{b + (atk - b)e^{-at}}{(b + ke^{-at})^2}, \\
\frac{\partial x}{\partial b} &= \frac{-a(1 - e^{-at})}{(b + ke^{-at})^2}, \\
\frac{\partial x}{\partial x_0} &= \frac{a^2 e^{-at}}{x_0^2 (b + ke^{-at})^2}.
\end{aligned}$$

The matrix χ is thus composed of these partial derivatives evaluated at points $t = t_1, \ldots, t_n$, i.e.,

$$\chi = \begin{pmatrix} \frac{\partial x(t_1)}{\partial a} & \frac{\partial x(t_1)}{\partial b} & \frac{\partial x(t_1)}{\partial x_0} \\ \vdots & \vdots & \vdots \\ \frac{\partial x(t_n)}{\partial a} & \frac{\partial x(t_n)}{\partial b} & \frac{\partial x(t_n)}{\partial x_0} \end{pmatrix}.$$

For use in approximation Σ of the covariance matrix we thus obtain the matrix

$$\chi^T\chi = \sum_{j=1}^{n} \begin{pmatrix} \left(\frac{\partial x(t_j)}{\partial a}\right)^2 & \frac{\partial x(t_j)}{\partial a}\frac{\partial x(t_j)}{\partial b} & \frac{\partial x(t_j)}{\partial a}\frac{\partial x(t_j)}{\partial x_0} \\ \frac{\partial x(t_j)}{\partial b}\frac{\partial x(t_j)}{\partial a} & \left(\frac{\partial x(t_j)}{\partial b}\right)^2 & \frac{\partial x(t_j)}{\partial b}\frac{\partial x(t_j)}{\partial x_0} \\ \frac{\partial x(t_j)}{\partial x_0}\frac{\partial x(t_j)}{\partial a} & \frac{\partial x(t_j)}{\partial x_0}\frac{\partial x(t_j)}{\partial b} & \left(\frac{\partial x(t_j)}{\partial x_0}\right)^2 \end{pmatrix}.$$

We next consider the different regions in which we analyze Σ. We address the asymptotic regions R_0 and R_2 first.

If we sample data from R_0, where $t_j < \tau_1$ for $j = 1 \ldots n$, we have

$$\frac{\partial x(t_j)}{\partial a} \approx 0, \qquad \frac{\partial x(t_j)}{\partial b} \approx 0, \qquad \frac{\partial x(t_j)}{\partial x_0} \approx 1,$$

as a result of

$$\begin{aligned} \lim_{t\to 0} \frac{\partial x(t_j)}{\partial a} &= 0, \\ \lim_{t\to 0} \frac{\partial x(t_j)}{\partial b} &= 0, \\ \lim_{t\to 0} \frac{\partial x(t_j)}{\partial x_0} &= 1. \end{aligned}$$

Also note that

$$\sum_{j=1}^{n} \frac{\partial x(t_j)}{\partial x_0}\frac{\partial x(t_j)}{\partial x_0} = \sum_{j=1}^{n} 1 = n$$

for $t_j < \tau_1$ in R_0. Hence, in this region, we can make the approximation

$$\chi^T\chi \approx \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & n \end{pmatrix}.$$

In R_0 we expect not to be able to accurately determine a or b, however we should be able to estimate x_0. Moreover, we must expect difficulties when computing the standard errors for estimates of a and b. As additional data points are sampled, SE_a and SE_b will increase because $(\chi^T\chi)^{-1}$ approaches a matrix with unbounded values in the first two diagonal elements.

Next we consider the region R_2, where $t_j > \tau_2$ for $j = 1 \dots n$, and note that

$$\begin{aligned}
\lim_{t\to\infty} \frac{\partial x(t_j)}{\partial a} &= \frac{1}{b}, \\
\lim_{t\to\infty} \frac{\partial x(t_j)}{\partial b} &= -\frac{a}{b^2}, \\
\lim_{t\to\infty} \frac{\partial x(t_j)}{\partial x_0} &= 0.
\end{aligned}$$

This implies that in R_2

$$\frac{\partial x(t_j)}{\partial a} \approx \frac{1}{b}, \qquad \frac{\partial x(t_j)}{\partial b} \approx -\frac{a}{b^2}, \qquad \frac{\partial x(t_j)}{\partial x_0} \approx 0,$$

and hence,

$$\begin{aligned}
\chi^T\chi &\approx \sum_{j=1}^{n} \begin{pmatrix} \frac{\partial x(t_j)}{\partial a}\frac{\partial x(t_j)}{\partial a} & \frac{\partial x(t_j)}{\partial a}\frac{\partial x(t_j)}{\partial b} & 0 \\ \frac{\partial x(t_j)}{\partial b}\frac{\partial x(t_j)}{\partial a} & \frac{\partial x(t_j)}{\partial b}\frac{\partial x(t_j)}{\partial b} & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
&= n \begin{pmatrix} \frac{1}{b^2} & -\frac{a}{b^3} & 0 \\ -\frac{a}{b^3} & \frac{a^2}{b^4} & 0 \\ 0 & 0 & 0 \end{pmatrix}.
\end{aligned}$$

In this case the second column of $\chi^T\chi$ is a scalar multiple, $-\frac{a}{b}$, of column one, and thus it is not possible to estimate a and b independently. Similar to the standard errors corresponding to region R_0, SE_a and SE_b will increase in size with the additional sampling of data points. Also, in R_2 there is difficulty estimating x_0 with any accuracy. The ill-posedness (ill conditioning) of $\chi^T\chi$ can be expected when computing the standard error of x_0, and SE_{x_0} will also increase as more data points are sampled.

In region R_1, where $\tau_1 < t_j < \tau_2$ for $j = 1, \ldots n$, we note that the partial derivative estimates differ greatly from the estimates in regions R_0 and R_2, and in general the matrix $\chi^T\chi$ will be well-conditioned. When R_1 is included in the sampling region, we should be able to recover decent estimates for a, b, and x_0 along with reasonable corresponding SE.

5 Implementation

We first considered inverse problems with exact (no-noise) simulated data. MATLAB was used to compute numerical solutions to equation (7) for use as the model $\{f_j(\beta)\} = \{f(t_j, \beta)\}$ in (2). We create a simulated data set, $\{y_j\}_{j=1}^n$, using the solution given by (8) with a specific β, denoted β_0. That is, we evaluate the analytic solution (8) at t_j to obtain data y_j. We restrict the values of t_j to the region from which data is sampled, and provide an initial estimate β^0 for the MATLAB ODE solver.

The cost function uses $ode15s$ to approximate the solution and returns

$$J_n(\beta) = \sum_{j=1}^{n} |f(t_j, \beta) - y_j|^2$$

where $f(t, \beta) = x(t; a, b, x_0)$ is the numerical approximation to the solution. We use the MATLAB function $fminsearch$ to optimize over β in order to obtain the minimized cost $J_n(\hat{\beta}^n)$. Here $\hat{\beta}^n$ represents the optimized value of β using n data points.

6 Results

The following are the results from three sets of simulated data: a relatively flat curve with $\beta_0 = (0.5, 0.1, 0.1)$, a moderately sloped curve with $\beta_0 = (0.7, 0.04, 0.1)$, and a steep curve with $\beta_0 = (0.8, 0.01, 0.1)$. For all three curves, define R_0 to be the region where $t \in [0, 2]$, R_1 is the region where $t \in [2, 12]$, and R_2 is where $t \in [12, 16]$. Within each region, we record four different initial guesses for the vector $\beta^0 = (a, b, x_0)$. We will show a figure corresponding to each set of simulated data plotted with the estimated curve for one of the initial guesses. We will also examine the standard errors of $\hat{\beta}^n$ by uniformly sampling varied amounts of data with $n = n_0 + n_1 + n_2$ where n_i represents the amount of data sampled from region $R_i, i = 0, 1, 2$.

6.1 Region R_0

When minimizing the OLS functional with data taken from region R_0, the model with the resulting parameters closely approximates the data curve within R_0; however, as seen in Figure 3, it does not come close to the data set in other regions. We point out that that the graphs in Figure 3 correspond to the first entries from each section of Table 1.

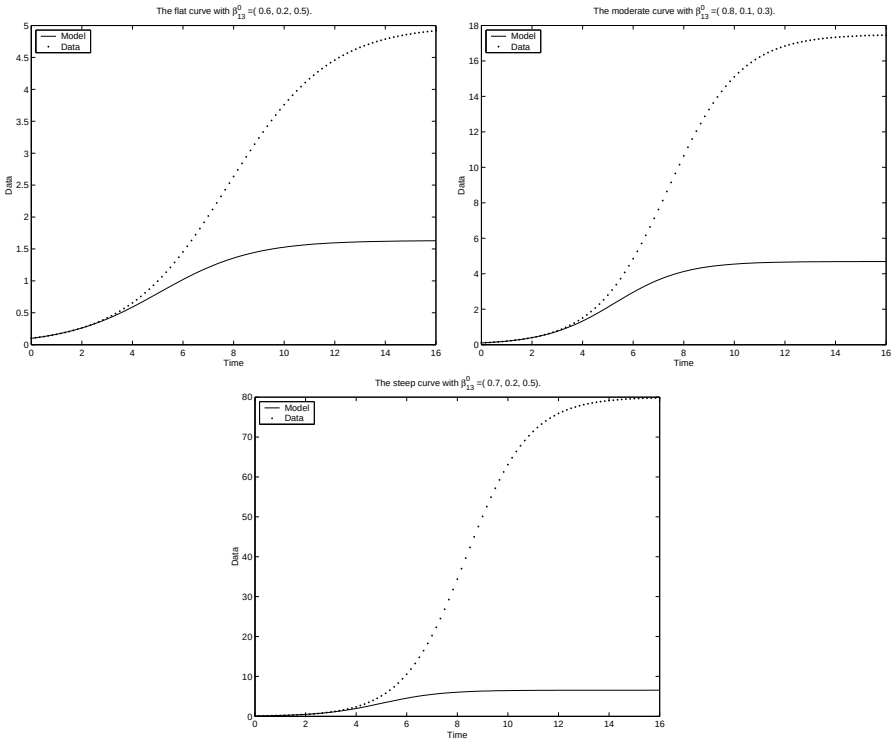

Figure 3: Simulated data compared to the solution with estimated parameters using data in the region R_0 for the a) flat curve with $\beta^0 = (0.6, 0.2, 0.5)$ b) moderate curve with $\beta^0 = (0.8, 0.1, 0.3)$ c) steep curve with $\beta^0 = (0.7, 0.2, 0.5)$.

When optimizing $J_n(\beta)$, we minimize the difference between model solution and data only in the region of interest. Note that the values for $J_{13}(\hat{\beta}^{13})$ using data from R_0 are close to zero in every case even when $\hat{\beta^{13}}$ is not close to β_0. Because of our previous analytical analysis, we are not surprised that

The flat curve with $\beta_0 = (0.5, 0.1, 0.1)$ where K=5.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{13}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{13}}$	$\boldsymbol{J_{13}(\hat{\beta}^{13})}$	$\boldsymbol{\hat{K}}$
(0.6,0.2,0.5)	5.8587	(0.5415,0.3317,0.0991)	$.3766e^{-5}$	1.6328
(2,0.5,1)	89.7757	(0.6842,1.1442,0.0966)	$.7833e^{-4}$	0.598
(0.7,0.3,0.001)	0.398	(0.6132, 0.7294, 0.0976)	$.2799e^{-5}$	0.8407
(5,5,5)	33.991	(2.391,10.7157,0.0691)	0.0058	0.2231

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ where K=17.5.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{13}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{13}}$	$\boldsymbol{J_{13}(\hat{\beta}^{13})}$	$\boldsymbol{\hat{K}}$
(0.8,0.1,0.3)	3.6193	(0.7281,0.1552,0.0991)	$.4887e^{-5}$	4.6907
(2,0.5,0.4)	46.2021	(0.9510,1.0667,0.0919)	$.395e^{-3}$	0.8915
(0.3,0.05,1)	14.9077	(0.6920,0.0076,0.1003)	$.3658e^{-6}$	91.5778
(5,5,5)	33.0287	(3.1546,10.0916,0.0407)	0.0239	0.3126

The steep curve with $\beta_0 = (0.8, 0.01, 0.1)$ where K=80.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{13}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{13}}$	$\boldsymbol{J_{13}(\hat{\beta}^{13})}$	$\boldsymbol{\hat{K}}$
(0.7,0.2,0.5)	6.0063	(0.8339,0.1273,0.0987)	$.1123e^{-4}$	6.5524
(0.9,0.005,0.2)	1.7988	(0.799,0.0064,0.1)	$.6183e^{-8}$	125.7549
(0.6,0.03,0.9)	28.6298	(0.8015,0.015,0.0999)	$.3084e^{-7}$	53.3319
(5,5,5)	32.4756	(3.6576,9.7304,0.0246)	0.0442	0.3759

Table 1: The initial and optimized parameters from region R_0.

the OLS procedure does not lead to accurate estimates for the values of a or b in this region, and we can see from Table 1 that the numerical results support this previous analysis. Moreover, as expected, the estimation procedure did return a reasonable estimate for the true value of x_0.

6.2 Region R_1

Note in Figure 4 that by minimizing with data from region R_1, the best fit model provides a decent approximation in R_1 as well as in the other regions. This is what we predicted based on the previous analysis regarding the $\chi^T\chi$ matrix. The graphs in the figure correspond to the initial entries from each section of Table 2. The solution is dynamically changing in R_1 with the asymptotes in the individual curves lying outside of this region. Hence it is

possible to obtain reasonable estimates for true values β_0 using data sampled from R_1.

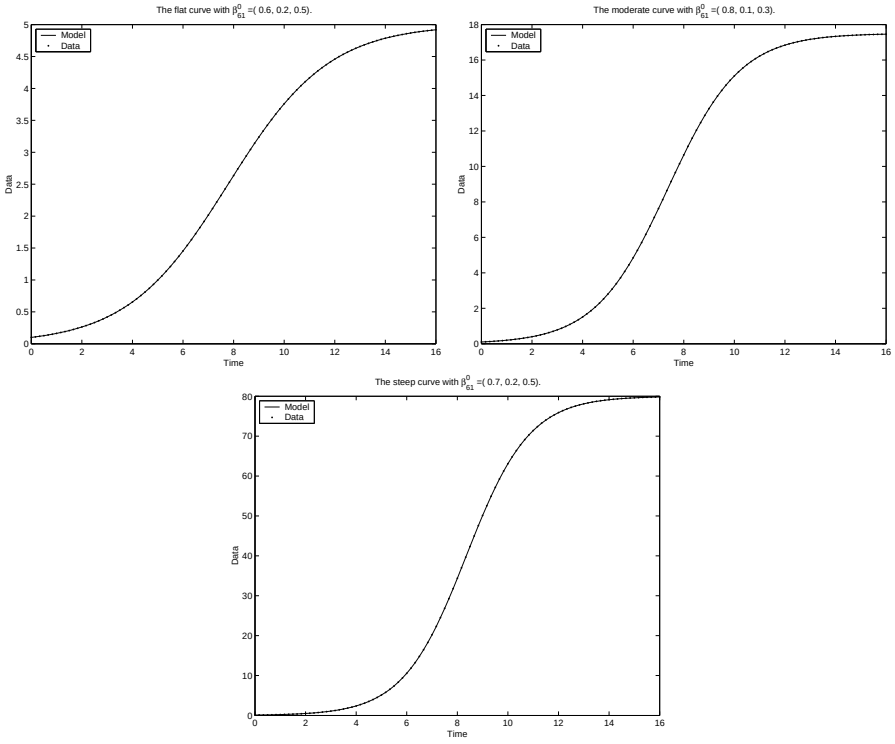

Figure 4: Simulated data compared to the solution with estimated parameters using data from the region R_1 for the a) flat curve with $\beta^0 = (0.6, 0.2, 0.5)$ b) moderate curve with $\beta^0 = (0.8, 0.1, 0.3)$ c) steep curve with $\beta^0 = (0.7, 0.2, 0.5)$.

The minimized cost, $J_{61}(\hat{\beta}^{61})$, is very small for reasonable initial guesses as is seen in Table 2. When we used the initial guess of $\beta^0 = (5.0, 5.0, 5.0)$, the procedure returned a poor estimate of the parameter values. However, when reasonable initial guesses were used, the values of β_0 were nearly exactly approximated. This illustrates the local convergent properties of the OLS procedure.

The flat curve with $\beta_0 = (0.5, 0.1, 0.1)$ where K=5.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{61}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{61}}$	$\boldsymbol{J_{61}(\hat{\beta}^{61})}$	$\boldsymbol{\hat{K}}$
(0.6,0.2,0.5)	68.5101	(0.5,0.1,0.1)	$.1449e^{-6}$	4.7
(2.0,0.5,1)	319.5682	(0.4999,0.1,0.1)	$.4067e^{-6}$	5
(0.7,0.3,0.001)	263.4515	(0.5,0.1,0.1)	$.1757e^{-6}$	4.9999
(5,5,5)	204.9234	(6.4942,2.9899,5.1836)	121.1291	2.1721

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ where K=17.5.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{61}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{61}}$	$\boldsymbol{J_{61}(\hat{\beta}^{61})}$	$\boldsymbol{\hat{K}}$
(0.8,0.1,0.3)	$1.3315e^3$	(0.6999,0.04,0.1)	$.1052e^{-4}$	17.4997
(2.0,0.5,0.4)	$3.2091e^3$	(0.7,0.04,0.1)	$.1117e^{-4}$	17.4980
(0.3,0.05,1)	$2.6895e^3$	(0.7,0.04,0.1)	$.8611e^{-6}$	17.4988
(5,5,5)	$5.2903e^3$	(7.6425,0.938,5.9823)	$.2173e^4$	8.1481

The steep curve with $\beta_0 = (0.8, 0.01, 0.1)$ where K=80.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{61}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{61}}$	$\boldsymbol{J_{61}(\hat{\beta}^{61})}$	$\boldsymbol{\hat{K}}$
(0.7,0.2,0.5)	$8.7989e^4$	(0.7999,0.01,0.1)	$.3903e^{-3}$	79.9890
(0.9,0.005,0.2)	$2.5627e^4$	(0.7998,0.01,0.1)	$.3798e^{-3}$	79.99
(0.6,0.03,0.9)	$4.7167e^4$	(0.7999,0.01,0.1)	$.3864e^{-3}$	79.9864
(5,5,5)	$9.6307e^4$	(7.9785,0.2681,4.9371)	$.458e^5$	29.7608

Table 2: The initial and optimized parameters using data from region R_1.

6.3 Region R_2

When carrying out the OLS estimation problem with data selected form R_2, we observe (Figure 5, Table 3) that each model curve with the estimated parameters is a good approximation to the data as the curves approach the carrying capacity K. However, when we compare the model with estimated parameters to the data over the entire region, we see that it is not in reasonable agreement. In particular, the only closely approximated part of the data is $x^\infty = K = \frac{a}{b}$.

Moreover, when we use data only from region R_2, the estimate for each parameter is poor, which is consistent with our earlier analytical considerations. Recall that the analysis of $\chi^T\chi$ for sampling in R_2 suggests that it might be possible to approximate the relationship between a and b without

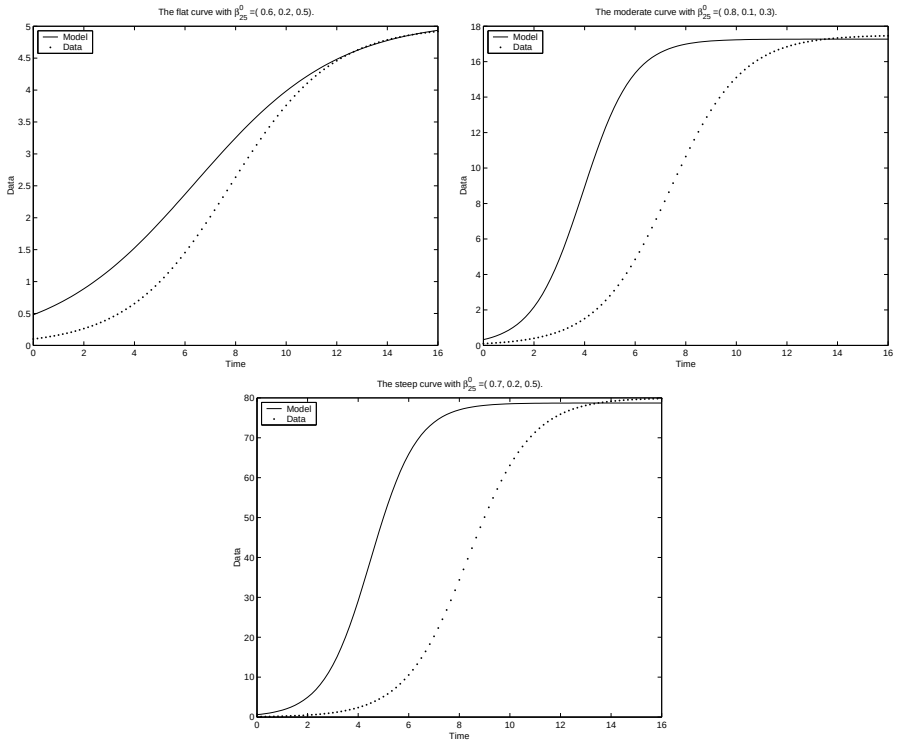

Figure 5: Simulated data plotted versus the model with estimated parameters using data from region R_2 for the a) flat curve with $\beta^0 = (0.6, 0.2, 0.5)$ b) moderate curve with $\beta^0 = (0.8, 0.1, 0.3)$ c) steep curve with $\beta^0 = (0.7, 0.2, 0.5)$.

independently estimating the parameters. Indeed, Table 3 reveals that we do obtain a reasonable approximation for K. Furthermore, note that we are not able to estimate x_0 with data from R_2, and as a result the optimized value $\hat{x_0}$ is near the initial guess instead of the true value. The algorithm cannot improve much on the initial guess for x_0 when using data from R_2 and hence leaves it essentially unchanged from the initial guess. Similarly, it appears that the optimized value $\hat{a}$ is relatively close to the initial guess for a_0, whereas $\hat{b}$ is adjusted accordingly to produce the optimized value of $\hat{K}$, the only quantity to which the data in R_2 provides any sensitivity.

The flat curve with $\beta_0 = (0.5, 0.1, 0.1)$ where K=5.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{25}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{25}}$	$\boldsymbol{J_{25}(\hat{\beta}^{25})}$	$\boldsymbol{\hat{K}}$
(0.6,0.2,0.5)	77.5157	(0.3539,0.0694,0.4791)	0.0025	5.1007
(2,0.5,1)	14.5852	(1.9241,0.4049,1.0643)	0.4672	4.7515
(0.7,0.3,0.001)	188.931	(0.8837,0.1805,0.0011)	0.0144	4.8971
(5,5,5)	352.3011	(7.3549,1.5480,5.3033)	0.4681	4.7513

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ where K=17.5.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{25}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{25}}$	$\boldsymbol{J_{25}(\hat{\beta}^{25})}$	$\boldsymbol{\hat{K}}$
(0.8,0.1,0.3)	$2.151e^3$	(1.0057,0.0582,0.3249)	0.7889	17.2699
(2,0.5,0.4)	$4.403e^3$	(2.7639,0.16,0.4508)	0.7971	17.2696
(0.3,0.05,1)	$3.4282e^3$	(0.5051,0.0287,0.9384)	0.0066	17.5785
(5,5,5)	$6.6184e^3$	(7.6593,0.4435,5.9774)	0.8069	17.2697

The steep curve with $\beta_0 = (0.8, 0.01, 0.1)$ where K=80.

$\boldsymbol{\beta^0}$	$\boldsymbol{J_{25}(\beta^0)}$	$\boldsymbol{\hat{\beta}^{25}}$	$\boldsymbol{J_{25}(\hat{\beta}^{25})}$	$\boldsymbol{\hat{K}}$
(0.7,0.2,0.5)	$1.4142e^5$	(1.0855,0.0138,0.597)	31.6162	78.7080
(0.9,0.005,0.2)	$2.5195e^5$	(0.7263,0.0091,0.2402)	0.0308	80.1117
(0.6,0.03,0.9)	$8.6528e^4$	(0.5574,0.0069,1.7452)	0.3821	80.4821
(5,5,5)	$1.5097e^5$	(8.0185,0.1019,6.1419)	32.1897	78.7013

Table 3: The initial and optimized parameters with data from region R_2.

6.4 Standard Errors

When we examine the standard errors for $\hat{\beta}^n$ as calculated in equation (9), we obtain a measure of the reliability (i.e., uncertainty associated with) the estimated parameter values. Specifically, as more points are sampled from the asymptotic portion of the curve, the values of the estimated parameters may or may not improve and the SE may or may not decrease in size. In addition to using noise free simulated data as in the previous computations, we will next introduce noise into the data and examine the corresponding resulting standard errors for the OLS estimation procedures.

As in the case for estimation using noise free data, by sampling data from R_0 alone, we are unable to estimate the values for a or b, and similarly, when data is sampled from R_2 alone, we are unable to estimate the values for a,

b, or x_0. Even if the number of points sampled from each region is doubled, we expect as before that the estimates will not improve. However, if the additional points are instead sampled from a portion of R_1, we suspect that the estimates for $\hat{\beta}^n$ might improve. In addition to more accurate estimates, the standard errors may decrease in size when R_1 is included in the sampling region.

The flat curve with $\beta_0 = (0.5, 0.1, 0.1)$ using $\beta^0 = (0.6, 0.2, 0.5)$.

$\boldsymbol{n_0}$	$\boldsymbol{n_1}$	$\hat{\boldsymbol{\beta}}^{\boldsymbol{n}}$	**Standard Errors**
25	0	(0.5418,0.3345,0.0991)	$(0.011, 0.0078, 0.0051)$
49	0	(0.5413,0.332,0.0991)	$(0.0216, 0.0153, 0.01)$
25	24	(0.5,0.1,0.1)	$(7.5111e^{-5}, 1.8985e^{-5}, 5.2494e^{-5})$

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ using $\beta^0 = (0.8, 0.1, 0.3)$.

$\boldsymbol{n_0}$	$\boldsymbol{n_1}$	$\hat{\boldsymbol{\beta}}^{\boldsymbol{n}}$	**Standard Errors**
25	0	(0.7244,0.1414,0.0992)	$(0.0049, 0.0011, 0.0025)$
49	0	(0.7241,0.1401,0.0992)	$(0.0094, 0.0020, 0.0049)$
25	24	(0.6998,0.0397,0.1)	$(2.2059e^{-4}, 1.4424e^{-5}, 1.5696e^{-4})$

The steep curve with $\beta_0 = (0.8, 0.01, 0.1)$ using $\beta^0 = (0.7, 0.2, .5)$.

$\boldsymbol{n_0}$	$\boldsymbol{n_1}$	$\hat{\boldsymbol{\beta}}^{\boldsymbol{n}}$	**Standard Errors**
25	0	(0.8325,0.1238,0.0987)	$(0.0074, 0.0012, 0.0036)$
49	0	(0.8334,0.1280,0.0987)	$(0.0155, 0.0026, 0.0075)$
25	24	(0.8,0.01,0.1)	$(7.0893e^{-5}, 1.0211e^{-6}, 5.7533e^{-5})$

Table 4: The estimated parameters for $\beta = (a, b, x_0)$ along with standard errors for the optimized $\hat{\beta}^n$.

6.4.1 Simulated Data without Noise

We first sampled 25 data points from the region R_0 of the data set with no noise. These points are obtained using a uniform grid with the increment of $1/12$ over the region $[0, 2]$. Note that in the corresponding results given in Table 4, $\hat{x_0}$ is close to the true value of x_0, whereas the estimates for a and b are not simultaneously very accurate with the inaccuracy increasing as the steepness of the data curve increases. This is consistent with our results from

Table 1. In attempts to improve the optimized parameter values as well as the standard errors, we first refine the grid size within R_0. Using the increment of $1/24$ over the region $[0, 2]$, we uniformly sampled 49 points, and observed that $\hat{\beta}^{49}$ is nearly equal to $\hat{\beta}^{25}$, although the standard errors doubled in size. Since there is no marked improvement in $\hat{\beta}^{n_0}$ by increasing the number of sampled points from region R_0, we instead doubled the sampling region, and now included points sampled from R_1. Using the increment of $1/12$ uniformly over the region [0,4], we sampled 49 points from region $R_0 \cup R_1$. Using this data set it is seen in Table 4 that $\hat{\beta}^{49}$ is a close approximation to the true value of β_0, and the standard errors are reduced by several orders of magnitude.

The flat curve with $\beta_0 = (0.5, 0.1, 0.1)$ using $\beta^0 = (0.6, 0.2, 0.5)$.

n_1	n_2	$\hat{\beta}^n$	**Standard Errors**
0	25	(0.3539,0.0694,0.4791)	$(0.0702, 0.0182, 0.1684)$
0	49	(0.3190,0.0621,0.6709)	$(0.1668, 0.0435, 0.4931)$
24	25	(0.5,0.1,0.1)	$(7.6421e^{-4}, 1.9291e^{-4}, 5.3446e^{-4})$

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ using $\beta^0 = (0.8, 0.1, 0.3)$.

n_1	n_2	$\hat{\beta}^n$	**Standard Errors**
0	25	(1.0057,0.0582,0.3249)	$(1.0794, 0.0663, 1.3714)$
0	49	(1.0058,0.0582,0.3249)	$(2.1739, 0.1335, 2.7618)$
24	25	(0.6996,0.04,0.1003)	$(5.0686e^{-3}, 3.3366e^{-4}, 3.6103e^{-3})$

The steep curve with $\beta_0 = (0.8, 0.01, 0.1)$ using $\beta^0 = (0.7, 0.2, 0.5)$.

n_1	n_2	$\hat{\beta}^n$	**Standard Errors**
0	25	(1.0855,0.0138,0.5970)	$(1.6669, 0.0224, 4.4595)$
0	49	(1.0855,0.0138,0.5970)	$(3.3488, 0.045, 8.9585)$
24	25	(0.7996,0.01,0.1002)	$(1.4895e^{-2}, 2.1398e^{-4}, 1.2125e^{-2})$

Table 5: The estimated parameters for $\beta = (a, b, x_0)$ along with standard errors for the optimized $\hat{\beta}^n$.

We turn to similar computations for data from R_2 and $R_2 \cup R_1$. We sampled 25 data points from region R_2, where these points are obtained using a uniform grid with the increment of $1/6$ over the region $[12, 16]$. Note in Table 5 that the values for $\hat{K} = \hat{a}/\hat{b}$ are reasonably close to the true

values of $K = 5, 17.5$, and 80, respectively for the three data curves. The corresponding estimates for a, b and x_0 are not accurate for any of the curves. This is consistent with our findings reported in Table 3. In an initial attempt to improve the optimized values as well as the standard errors, we refined the sampling grid within R_2. Using the increment of $1/12$ over the region $[12, 16]$, we uniformly sampled 49 points. We observe that $\hat{\beta}^{49}$ is nearly equivalent to $\hat{\beta}^{25}$, and as before with data in R_0, the standard errors double in size. Since there is no improvement in $\hat{\beta}^{n_2}$ by increasing the number of sampled points from region R_2, we next doubled the sampling region and included points sampled from R_1. Thus, using the increment of $1/6$ over the region [8,16], we sampled 49 points from the region $R_1 \cup R_2$. Using this data set, $\hat{\beta}^{49}$ is a reasonable approximation to the true value of β_0, and the standard errors are significantly reduced.

6.4.2 Simulated Data with Noise

We considered two different sets of simulated data with noise corresponding to the moderate curve, i.e., $\beta_0 = (0.7, 0.04, 0.1)$. One set corresponds to the data in R_0 which is perturbed by adding Gaussian noise $\epsilon_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_0)$ for $j = 1, 2, \ldots, n$, with $\sigma_0 = 0.005$. The other data set corresponds to data in R_2 with similar added noise except with $\sigma_0 = 0.5$. These sigma values represent an added noise level of approximately ten percent at the lowest bound in the observed regions. Figure 6 depicts the data set with the highest level of noise.

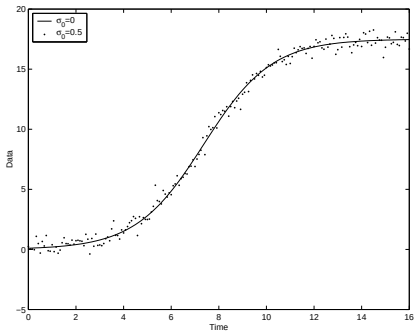

Figure 6: Noisy data with $\sigma_0 = 0.5$ compared to the exact solution with $\beta_0 = (0.7, 0.04, 0.1)$.

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ using $\beta^0 = (0.8, 0.1, 0.3)$.

n_0	n_1	$\hat{\beta}^n$	Standard Errors
25	0	(0.7244,0.1455,0.0988)	$(0.052, 0.0116, 0.0269)$
49	0	(0.7245,0.1452,0.0988)	$(0.1060, 0.0236, 0.0548)$
25	24	(0.7086,0.0517,0.0988)	$(0.0497, 0.0041, 0.033)$
49	48	(0.703,0.0435,0.0994)	$(0.091, 0.0065, 0.0631)$
0	25	(0.7058,0.0491,0.0994)	$(0.0133, 0.0011, 0.009)$
0	49	(0.6958,0.0362,0.1008)	$(0.0252, 0.0015, 0.0184)$

Table 6: The standard errors for the optimized $\hat{\beta}^n$ with $\sigma_0 = 0.005$ using data sampled from the interval $[0, 4]$.

We uniformly sampled data with a step size of 1/12 from the intervals [0,2], [0,4], and [2,4]. Then the sampling frequency is doubled over the same set of intervals to determine if refining the grid will result in better estimates with reasonable standard errors. By examining the [0,2] interval, we determine that the $\hat{\beta}^{25}$ does not accurately estimate β_0 and the standard errors double as a result of the grid refinement. As seen in Table 6, the estimates and the standard errors improve somewhat when sampling from [0,4]. However, doubling the sampling frequency increases the standard errors without significant improvement to the estimates. When sampling from the interval [2,4], we observed improved parameter estimates with reduced standard errors. Thus we see that including the R_0 region within the data set results in worse estimates and larger standard errors than sampling from R_1 alone.

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ using $\beta^0 = (0.8, 0.1, 0.3)$.

n_1	Interval	$\hat{\beta}^n$	Standard Errors
25	[2,4]	(0.7058,0.0491,0.0994)	$(0.0133, 0.0011, 0.009)$
49	[2,4]	(0.6958,0.0362,0.1008)	$(0.0252, 0.0015, 0.0184)$
49	[2,6]	(0.6981,0.0394,0.1006)	$(0.0355, 0.0023, 0.0255)$
73	[2,8]	(0.6999,0.04,0.1001)	$(0.0794, 0.0052, 0.0564)$

Table 7: The standard errors for the optimized $\hat{\beta}^n$ with $\sigma_0 = 0.005$ using data sampled from various intervals within R_1.

Instead of only increasing the sampling frequency over a fixed interval,

we also considered increasing the sampling region to determine if there is an improvement in parameter estimation. Based on our findings summarized in Table 6, we focused on sampling data points from region R_1 alone. We have observed that when a given parameter estimate is already close to the true value, the standard error corresponding to that parameter becomes discernably worse as the sampling region increases. In this case, even though n is increased in a region of high sensitivity to all parameters, the OLS procedure is not improved with additional data since the additional data contains only information already obtained at the lower sampling level, i.e., the ill-conditioning of $\chi^T\chi$ is increasing. These results are displayed in Table 7.

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ using $\beta^0 = (0.8, 0.1, 0.3)$.

n_1	n_2	$\hat{\beta}^n$	**Standard Errors**
0	25	(1.0058,0.0582,0.3249)	$(2.7953, 0.1716, 3.5515)$
0	49	(1.0057,0.0582,0.3249)	$(6.3606, 0.3906, 8.0808)$
24	25	(0.5286,0.0297,0.4397)	$(3.7015, 0.2516, 10.375)$
48	49	(0.673,0.0384,0.123)	$(10.2462, 0.679, 8.8898)$
25	0	(0.5175,0.0281,0.4307)	$(1.0073, 0.0671, 2.8477)$
49	0	(0.4713,0.0249,0.5921)	$(2.1167, 0.1414, 8.0601)$

Table 8: The standard errors for the optimized $\hat{\beta}^n$ with $\sigma_0 = 0.5$ using data sampled from the interval $[8, 16]$.

We uniformly sampled data with a step size of 1/6 from the intervals [12,16], [8,16], and [8,12] in Table 8. By doubling the sampling frequency over the same set of intervals, we again questioned if refining the grid will result in better estimates with reasonable standard errors. On the [12,16] interval, we determined that the OLS procedure does not return a good estimate for β_0. When the sampling frequency is doubled, the poor parameter estimate capability remains and the standard errors double. We included data points from the upper region of R_1 by sampling from the interval [8,16] and there is slight improvement in the parameter estimates and the standard errors. However, when we restrict the sampling region to [8,12] the standard errors improve significantly. Therefore including R_2 in the data set increases the standard error with minimal improvement to the parameter estimates.

In Table 9 we again increased the sampling region instead of primarily increasing the sampling frequency over a fixed interval to determine if there

The moderate curve with $\beta_0 = (0.7, 0.04, 0.1)$ using $\beta^0 = (0.8, 0.1, 0.3)$.

n_1	Interval	$\hat{\beta}^n$	Standard Errors
25	[8,12]	(0.5175,0.0281,0.4307)	$(1.0073, 0.0671, 2.8477)$
49	[8,12]	(0.4713,0.0249,0.5921)	$(2.1167, 0.1414, 8.0601)$
49	[4,12]	(0.6720,0.0383,0.1254)	$(5.1571, 0.3409, 4.5525)$

Table 9: The standard errors for the optimized $\hat{\beta}^n$ with $\sigma_0 = 0.5$ using data sampled from various intervals within R_1.

is an improvement in parameter estimates. In light of the results from Table 8, we focused on sampling data points from region R_1 alone. As the sampling region expands to include more of the curve in R_1, a better estimate for the parameters β_0 is obtained. While we are able to improve our parameter estimates, the corresponding standard errors are again increased (compare with Table 7).

7 Concluding remarks

In summary we emphasize several points that are readily illustrated by our combination of analytical and computational analysis in this note. In inverse problems, parameter sensitivity is of fundamental importance in several regards:

(i) Lack of sensitivity can produce the inability to even estimate corresponding parameters;

(ii) As illustrated in Table 7, even when parameter sensitivity is adequate, increased sampling (i.e., larger n) does not necessarily significantly improve the estimates nor does it necessarily improve at all the reliability (i.e., the SE and CI) of the estimates obtained. In particular, ill-conditioning of the matrix $\chi^T\chi$ can increase as data points are added without adding new information (i.e., adding a row to χ that is almost linearly dependent on previous rows). This provides an increasing $(\chi^T\chi)^{-1}_{jj}$ and hence an increasing SE_j with increasing n even if $\hat{\sigma}^2 = \frac{1}{n-p}(RSS)$ is decreasing. This is especially likely with increased sampling near steady-states or equilibria of a dynamical system where

often little new information about dynamics is provided with additional sampling.

In practice one, of course, usually does not know the exact solution and may not know all the qualitative properties of the dynamics. But the analysis and computations in this note in the context of asymptotic statistical estimates for covariance matrices and standard errors suggest that one might profitably use the sensitivities in a problem to design experiments for additional data collection.

Acknowledgements

This research was supported in part (HTB and SLG) by the Joint DMS/ NIGMS Initiative to Support Research in the Area of Mathematical Biology under grant 1R01GM67299-01, in part (HTB and SLG) by the U.S. Air Force Office of Scientific Research under grant AFOSR-FA9550-04-1-0220, and in part (SLE) by the U.S. Dept of Homeland Security with a Fellowship in the DHS Scholarship and Fellowship Program, a program administered by the Oak Ridge Institute for Science and Education (ORISE) for DHS through an interagency agreement with the U.S Department of Energy (DOE). ORISE is managed by Oak Ridge Associated Universities under DOE contract number DE-AC05-06OR23100. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of DHS, DOE, or ORISE.

References

[1] H. M. Adelman and R.T. Haftka, Sensitivity analysis of discrete structural systems, *A.I.A.A. Journal*, **24** (1986), 823–832.

[2] H.T. Banks and D. M. Bortz, A parameter sensitivity methodology in the context of HIV delay equation models, CRSC-TR02-24, NCSU, August, 2002; *J. Mathematical Biology*, **50** (2005), 607–625.

[3] H.T. Banks and D. M. Bortz, Inverse problems for a class of measure dependent dynamical systems, CRSC-TR04-33, NCSU, September, 2004; *J. Inverse and Ill-posed Problems*, **13** (2005), 103–121.

[4] H.T. Banks, D.M. Bortz and S.E. Holte, Incorporation of variability into the mathematical modeling of viral delays in HIV infection dynamics, *Mathematical Biosciences*, **183** (2003), 63–91.

[5] H.T. Banks, D.M. Bortz, G.A. Pinter and L.K. Potter, Modeling and imaging techniques with potential for application in bioterrorism, Chapter 6 in *Bioterrorism: Mathematical Modeling Applications in Homeland Security*, (H.T. Banks and C. Castillo-Chavez, eds.), Frontiers in Applied Mathematics **FR28**, SIAM, Philadelphia, 2003, pp. 129–154.

[6] H. T. Banks and H. K. Nguyen, Sensitivity of dynamical system to Banach space parameters, CRSC Tech Rep., CRSC-TR05-13, N.C. State University, February, 2005; *J. Math Anal. Appl.*, in press.

[7] L.M.A. Bettencourt, A. Cintron-Arias, D.I. Kaiser and C. Castillo-Chavez, The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models, Preprint No. LAUR-05-0485, Los Alamos National Laboratory, January, 2005.

[8] S.M. Blower and H. Dowlatabadi, Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model as an example, *International Statistics Review*, **62** (1994), 229–243.

[9] G. Casella and R. L. Berger, *Statistical Inference,* Duxbury, California, 2002.

[10] J.B. Cruz, ed., *System Sensitivity Analysis*, Dowden, Hutchinson & Ross, Inc., Stroudsberg, PA, 1973.

[11] M. Davidian and D. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman & Hall, London, 1998.

[12] M. Eslami, *Theory of Sensitivity in Dynamic Systems: An Introduction*, Springer-Verlag, Berlin, 1994.

[13] P.M. Frank, *Introduction to System Sensitivity Theory*, Academic Press, Inc., New York, NY, 1978.

[14] M. Kleiber, H. Antunez, T.D. Hien and P. Kowalczyk, *Parameter Sensitivity in Nonlinear Mechanics: Theory and Finite Element Computations*, John Wiley & Sons, New York, NY, 1997.

[15] A. R. Gallant, *Nonlinear Statistical Models*, John Wiley & Sons, Inc., New York, 1987.

[16] R. I. Jennrich, Asymptotic properties of non-linear least squares estimators., *Ann. Math. Statist.*, **40** (1969), 633–643.

[17] M. Kot, *Elements of Mathematical Ecology*, Cambridge University Press, Cambridge, 2001, p. 7-9.

[18] A. Saltelli, K. Chan and E.M. Scott, eds., *Sensitivity Analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, NY, 2000.

[19] G. A. F. Seber and C. J. Wild, *Nonlinear Regression,* John Wiley & Sons, Inc., New York, 1989.