An Extension of RSS-based Model Comparison Tests for Weighted Least Squares

H.T. Banks, Zackary R. Kenz, and W. Clayton Thompson
Center for Research in Scientific Computation
Center for Quantitative Sciences in Biomedicine
North Carolina State University
Raleigh, NC 27695-8212

August 22, 2012

Abstract

We consider residual-sum-of-squares-based model comparison tests using weighted least squares formulations and asymptotic statistics. An biological example from cell proliferation investigations is presented in some detail to illustrate usefulness of the new results.

AMS subject classifications: 34A55, 62F03, 62F12, 93E24.

Key words: Model comparison tests, ordinary least squares, weighted least squares, asymptotic statistics.

Report Documentation Page		*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **22 AUG 2012**	2. REPORT TYPE	3. DATES COVERED **00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE **An Extension of RSS-based Model Comparison Tests for Weighted Least Squares**	5a. CONTRACT NUMBER
	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) **North Carolina State University,Center for Research in Scientific Computation,Center for Quantitative Sciences in Biomedicine,Raleigh,NC,27695-8212**	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
We consider residual-sum-of-squares-based model comparison tests using weighted least squares formulations and asymptotic statistics. An biological example from cell proliferation investigations is presented in some detail to illustrate usefulness of the new results.

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**	**Same as Report (SAR)**	**19**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

1 Introduction

Model comparison by hypothesis testing is a tool in inverse or parameter estimation problems to distinguish mathematically relevant features in a biological or physical system under study. Consider that one has some dynamical system which describes a biological process

$$\frac{dy}{dx} = g(x, y(x; q), q) \tag{1}$$
$$y(0) = y_0.$$

Given a set of observations of some component(s) of the system under study at various instances in time, it is a standard inverse problem to determine the value of the model parameter q, out of a set Q of admissible parameters, which best describes the observed data. But an equally important problem is to establish, on a quantitative basis, that the biophysical mechanisms represented in the mathematical model are necessary to describe the observed data, and that no extraneous mechanisms have been included. This is the role of model comparison tests.

For certain classes of models, potentially extraneous mechanisms can be eliminated from the model by a simple restriction on the underlying parameter space, while the form of the mathematical model (1) remains unchanged. For instance, consider the convection-diffusion model

$$\frac{\partial u}{\partial t} + \mathcal{V}\frac{\partial u}{\partial x} = \mathcal{D}\frac{\partial^2 u}{\partial x^2},$$

where $\mathcal{D}$ is the diffusion coefficient and $\mathcal{V}$ is the bulk velocity of a transporting fluid. Such a model has been used to describe the transport of labeled sucrose in the brain tissue of cats [4]. Yet if the cerebral transport of sucrose is primarily driven by diffusion, then the convection term in the model above is extraneous. Given data involving the transport of labeled sucrose, one would like to compare the accuracy of two mathematical models. In the first, both convection and diffusion play a role and one must estimate the parameters $\mathcal{V}$ and $\mathcal{D}$ in an inverse problem. In the second, it is assumed that convection is not present ($\mathcal{V} = 0$), and only the coefficient of diffusion $\mathcal{D}$ is estimated. Thus, in effect one has two mathematical models which must be compared in terms of how accurately and parsimoniously they describe an available data set. Equivalently, this problem can be cast in the form of a hypothesis test: one would like to test the null hypothesis $\mathcal{V} = 0$, against the alternative hypothesis that $\mathcal{V} \neq 0$. Under the null hypothesis, the restricted mathematical model is contained within the mathematical model subject to the unconstrained parameter space (in the sense that estimating both $\mathcal{V}$ and $\mathcal{D}$ simultaneously includes the possibility that $\mathcal{V} = 0$). For this reason, the unconstrained model must necessarily fit the data (in an appropriate sense) at least as accurately as the constrained model. The role of the hypothesis test is to compare these two models on a quantitative basis to investigate whether the model with $\mathcal{V} \neq 0$ provides a significantly (in some sense) better fit to the data.

As the above example makes clear, there is some similarity between model comparison testing and hypothesis testing for a set of mathematical models which have the same mathematical form but differ in the restrictions placed on the underlying parameter space. In such a case, one may resort to hypothesis testing as a technique for model comparison. In this note, we are interested in the situation where the parameter set for the restricted model, Q_H, can be identified as a linear subset of the admissible parameter set of the unrestricted model, Q (this notation will be defined explicitly in Section 3). In addition to the study of convective mechanisms in models of fluid transport in cat brains [4], hypothesis tests have also been used to investigate mortality and anemotaxis in models of insect dispersal [5, 6]. These examples are summarized (focusing on the role of hypothesis testing) in [2, 8]. More recently, model comparison by hypothesis testing has been conducted to investigate the division dependence of death rates in a model of a dividing cell population [10], to evaluate the necessity of components in viscoelastic stenosis models [26], and to serve as the basis for developing a methodology to determine if a stenosis is present in a vessel based on input amplitude [3]. Thus the discussions and methods presented here have widespread applications in biological modeling as well as in more general scientific modeling efforts.

While a number of techniques for quantitative model comparison exist, many of these rely upon assumptions which are highly restrictive or can be hard to verify in practice. For instance, one can use likelihood ratio tests (e.g., [16]) in order to evaluate the relative likelihoods of any pair of mathematical models. However, these tests require the existence and complete mathematical specification of probability density functions describing the

likelihood of a set of observations. Alternatively, one can use information theoretic model comparison tests such as Akaike's Information Criterion (AIC) or similar tests (BIC, TIC, etc.) [14]. Still, such tests typically require a complete likelihood specification. Though a least squares analog of the AIC test exists, it is only valid if all observations are considered to be sampled from identical normal distributions with a constant variance (i.e., *i.i.d.* normal).

Unlike likelihood estimation, least squares estimation generally requires only the first two statistical moments of the observations to be specified [13, Ch. 3]. Additionally, the mathematical formulation is intuitive, and the resulting estimates are more robust to error misspecification (when compared to likelihood estimation) [17, Ch. 2]. As such, there is an obvious appeal for model comparison tests which are easy to apply in a least squares framework. Given the above discussion regarding the relationship between model comparison and hypothesis testing for certain classes of problems, we focus our attention on model comparison in a least squares framework.

Hypothesis testing for least squares problems was considered at length by Gallant [19]. Subsequently, these results were reconsidered by Banks and Fitzpatrick [2] under a slightly different set of assumptions to develop a hypothesis test which could be computed from the residual sum of squares (RSS) after fitting the models to a data set in a least squares framework. Unfortunately, this work was limited to cases in which the observations were assumed to be characterized by constant variance. Both the works of Gallant as well as Banks and Fitzpatrick rely upon an asymptotic theory for nonlinear least squares to demonstrate their results.

In this presentation we begin by defining a general least squares estimation problem. Next, we review features of the asymptotic theory relevant to model comparison by hypothesis testing. Finally, we further extend the results of Banks and Fitzpatrick (in a manner suggested possible by Gallant) to a more general variance model. An example involving cell proliferation models with nonconstant variance in the experimental data is used to illustrate application of the new model comparison techniques.

2   Least Squares Estimation

We begin with the dynamical system (1) above, which describes some hypothesized mathematical relationship between an observed process $y \in \mathbb{R}^k$, an independent variable (covariate) $x \in \mathbb{R}^m$, and a set of parameters $q \in Q$. For notational convenience, we assume without loss of generality that $k = 1$. Let $f(x; q)$ be the solution to (or observation for) the dynamical system. Suppose also that observations of the underlying process are noisy (that is, subject to random error and/or fluctuations). Following standard inverse problem procedures [13, 15, 17, 24], we may relate the mathematical model to the noisy data via a statistical model,

$$Y_k = f(x_k; q^*) + \mathcal{E}_k. \tag{2}$$

Here, $\{\mathcal{E}_k\}_{k=1}^n$ are random variables which account for any error (measurement error, modeling error, microfluctuations, etc.) between the mathematical model and the recorded observations at values $\{x_k\}_{k=1}^n$ of the independent variable. Hence the $\{Y_k\}_{k=1}^n$ are also random variables. It is assumed here that the values $\{x_k\}$ are known exactly. The parameter q^* is the hypothetical 'true' parameter which is assumed to generate the data for the observed process. It has been tacitly assumed that the underlying biological or physical system is perfectly represented by the mathematical model. (In fact, any mathematical model is only an approximation to reality. For the sake of argument, one can assume that the mathematical model is sufficiently accurate so that any modeling error is subsumed within the error random variables $\mathcal{E}_k$. One can relax this assumption, though doing so greatly complicates [18, 19] the analysis presented below.) It is further assumed that the error random variables are independent, though not necessarily identically distributed, with central moments

$$E[\mathcal{E}_k] = 0, \quad 1 \le k \le n,$$
$$Var(\mathcal{E}_k) = \sigma^2 w(x_k)^2, \quad 1 \le k \le n.$$

With the assumptions and definitions above, we may define the *weighted least squares cost function*,

$$J_n^{WLS}(q) = \frac{1}{n} \sum_{k=1}^n \left(\frac{Y_k - f(x_k; q)}{w(x_k)} \right)^2. \tag{3}$$

The *weighted least squares estimator* is defined to be the random variable q_n^{WLS} which minimizes the weighted least squares cost function for a given set of random variables $\{Y_k\}$, that is,

$$q_n^{WLS} = \arg \min_{q \in Q_{ad}} J_n^{WLS}(q), \tag{4}$$

where $Q_{ad} \subset Q$ is the set of admissible values for the parameter q. We remark that q_n^{WLS} is a random variable which is dependent upon the random variables $\{Y_k\}$ (and hence the error random variables $\{\mathcal{E}_k\}$). This dependence is suppressed in the notation above (with the exception of the subscript n) for simplicity and clarity, although we can write $q_n^{WLS} = q_n^{WLS}(\{Y_k\})$ when necessary. In the event $w(x_k) = 1$ for all $1 \leq k \leq n$, the formulas above reduce down to the familiar ordinary least squares formulae

$$J_n^{OLS}(q) = \frac{1}{n} \sum_{k=1}^{n} (Y_k - f(x_k; q))^2, \tag{5}$$

$$q_n^{OLS} = \arg \min_{q \in Q_{ad}} J_n^{OLS}(q). \tag{6}$$

Finally, we remark that it is often convenient to use a vectorized notation for the least squares problems. Define $\vec{f}(q) = (f(x_1; q), \ldots, f(x_n; q))^T$. Then the statistical model (2) may be rewritten as

$$\vec{Y} = \vec{f}(q) + \vec{\mathcal{E}} \tag{7}$$

where

$$E[\vec{\mathcal{E}}] = \vec{0}$$
$$Cov(\vec{\mathcal{E}}) = \sigma^2 diag(w(x_1)^2, \ldots, w(x_n)^2)$$
$$= \sigma^2 W.$$

Then clearly we can rewrite

$$J_n^{WLS}(q) = \frac{1}{n} \left(\vec{Y} - \vec{f}(q) \right)^T W^{-1} \left(\vec{Y} - \vec{f}(q) \right) \tag{8}$$

$$J_n^{OLS}(q) = \frac{1}{n} \left(\vec{Y} - \vec{f}(q) \right)^T \left(\vec{Y} - \vec{f}(q) \right). \tag{9}$$

The goal of an inverse or parameter estimation problem is to use a set of observations $\{y_k\}$ (realizations of the random variables $\{Y_k\}$) in order to determine an *estimate* of the true underlying parameter q^*. Intuitively, since $E[Y_k] = f(x_k; q^*)$ for all $1 \leq k \leq n$, we would expect the least squares estimator (either q_n^{WLS} or q_n^{OLS}) to be 'close' to q^* in an appropriate sense. In practice, one uses the realized data $\{y_k\}$ to determine an estimate $\hat{q}_n^{WLS}$ or $\hat{q}_n^{OLS}$ which is a realization of the appropriate least squares estimator. In this paper, we are less interested in the computational issues associated with finding an estimate and more interested in characterizing the statistical properties of the estimator itself. Since the estimator is a random variable, it is described not by a numerical value but rather by a probability distribution, and we may use results from asymptotic theory to characterize the distribution of the estimator.

Going forward, for notational simplicity the dependence of the estimators and cost functions on n may be suppressed (e.g., using $J^{OLS}(q)$ instead of $J_n^{OLS}(q)$). This is the case particularly during the discussions of the cell proliferation example in Section 5.

3 Asymptotic Theory and Model Comparison for Ordinary Least Squares

We now review some existing results on the asymptotic properties of nonlinear ordinary least squares estimators, particularly those relevant to hypothesis testing. The notation and organization of this section closely follows that

of Banks and Fitzpatrick [2]. As mentioned previously, the work of Banks and Fitzpatrick was partially inspired by the work of Gallant [19], and many of their results and assumptions are similar. In the discussion that follows, we comment on any differences between the two approaches that we feel are noteworthy. While Gallant's work is remarkably general, allowing for a misspecified model and a general estimation procedure (both least mean distance estimators and method of moments estimators are included), we do not consider such generalities here. The comments below are limited to least squares estimation with a correctly specified model (recall the discussion following Equation (2) above).

In this section we follow [2] in focusing exclusively on asymptotic properties and model comparison for *i.i.d.* (absolute) error models. The theoretical results of Gallant similarly focus on *i.i.d.* errors, though some mathematical tools are discussed which help to address more general error models [19, Ch. 2; pgs. 156–157]. In fact, these tools are used in a rigorous fashion in the next Section to extend the results of [2].

We begin by establishing some notation and precisely defining the hypothesis testing problem. It is assumed that the error random variables are defined on some probability triple $(\Omega, \mathcal{F}, P)$ and take their values in a space E. By construction, it follows that the data $\{Y_k\}$ as well as the estimators q_n^{WLS} and q_n^{OLS} are also random variables defined on this probability triple, and hence are functions of $\omega \in \Omega$ so that we may write $Y_k(\omega)$, $q_n^{OLS}(\omega)$, etc., as necessary.

As in the discussion of Section 2, we consider parameters $q \in Q_{ad} \subset Q$ for a topological space Q. Given the model $f(x; q)$ defined above, it is assumed that the sampled values of the independent variables satisfy $x_k \in X$ for all k, and for some metric space X. Moreover, if the set X is totally ordered (see Assumption (A3) below), then one can define the empirical distribution function,

$$\mu_n(x) = \frac{1}{n} \sum_{k=1}^{n} \Delta_{x_k}(x)$$

where

$$\Delta_{x_k}(x) = \begin{cases} 0, & x \leq x_k \\ 1, & otherwise \end{cases}.$$

Clearly, $\mu_n \in \mathcal{P}(X)$, the space of probability measures on X.

Finally, we return to the original goal of model comparison by hypothesis testing. For now, we use the notation of ordinary least squares estimation, in keeping with the results presented in [2]. We have already defined above the standard ordinary least squares estimator

$$q_n^{OLS} = \arg \min_{q \in Q_{ad}} J_n^{OLS}(q).$$

As alluded to in the introduction, we might also consider a restricted version of the mathematical model in which the unknown true parameter is assumed to lie in a subset $Q_H \subset Q_{ad} \subset Q$ of the admissible parameter space. We assume this restriction can be written as a linear constraint, $Hq^* = h$, where H is a known linear function and h is a known vector. Thus the restricted parameter space is

$$Q_H = \{q \in Q_{ad} : Hq = h\}.$$

Then the null and alternative hypotheses are

$$H_0 : q^* \in Q_H$$
$$H_A : q^* \notin Q_H.$$

We may define the restricted estimator

$$q_{n,H}^{OLS} = \arg \min_{q \in Q_H} J_n^{OLS}(q), \tag{10}$$

and the corresponding estimate $\hat{q}_{n,H}^{OLS}$. It is clear that $J_n^{OLS}(\hat{q}_n^{OLS}) \leq J_n^{OLS}(\hat{q}_{n,H}^{OLS})$, since $Q_H \subset Q_{ad}$. This fact forms the basis for a model comparison test based upon the residual sum of squares.

With these definitions, attention can now be directed toward establishing a quantitative basis upon which the null hypothesis H_0 should either be accepted or rejected. Again, the results presented below are paraphrased from [2], and comments have been included to indicate the alternative approach due to [19]. No proofs are given here, though the interested reader can find a complete set of proofs in [2] and [19]. First, we consider the following set of assumptions:

(A1) The error random variables $\{\mathcal{E}_k\}$ are independent and identically distributed random variables with distribution function $P(\mathcal{E})$. Moreover, $E[\vec{\mathcal{E}}] = \vec{0}$ and $Cov(\vec{\mathcal{E}}) = \sigma^2 I_n$ where I_n is the $n \times n$ identity matrix.

(A2) Q is a separable, finite-dimensional topological space (i.e., $Q \subset \mathbb{R}^p$), and Q_{ad} is a compact subspace of Q with $q^* \in int(Q_{ad})$.

(A3) X is a compact subset of $\mathbb{R}^m$.

(A4) The function $f(\cdot; \cdot) \in C^2(Q, C(X))$.

(A5) There exists a finite measure μ on X such that

$$\frac{1}{n}\sum_{k=1}^{n} g(x_k) = \int_X g(x)d\mu_n(x) \to \int_X g(x)d\mu(x)$$

for all continuous functions g. That is, the probability distributions μ_n converge weakly to μ, or μ_n converges to μ in the weak* topology (when $P(X)$ is viewed as a subset of $C^*(X)$, the dual of the space of continuous functions).

(A6) The functional

$$J^*(q) = \sigma^2 + \int_X \left(f(x; q^*) - f(x; q)\right)^2 d\mu(x)$$

has a unique minimizer at $q^* \in Q_{ad}$.

(A7) The matrix

$$\begin{aligned}
\mathcal{J} &= \frac{\partial^2 J^*(q^*)}{\partial q^2} \\
&= 2\int_X \left(\frac{\partial f(x; q^*)}{\partial q}\frac{\partial f(x; q^*)^T}{\partial q} + \frac{\partial^2 f(x; q^*)}{\partial q^2}\left(f(x; q^*) - f(x; q^*)\right)\right) d\mu(x) \\
&= 2\int_X \left(\frac{\partial f(x; q^*)}{\partial q}\frac{\partial f(x; q^*)^T}{\partial q}\right) d\mu(x)
\end{aligned}$$

is positive definite.

(A8) The matrix H, of dimension $r \times p$, has full rank r.

The most notable difference between the assumptions above (which are those of [2]) and those of [19] is assumption (A5). In its place, Gallant states the following.

(A5$'$) Define the probability measure

$$\nu(S) = \int_X \int_E I_S(\mathcal{E}, x)dP(\mathcal{E})d\mu(x)$$

for a set $S \subset E \times X$ and μ defined as above. Then almost every realized pair (ϵ_k, x_k) is a *Cesaro sum generator* with respect to ν and a dominating function $b(\mathcal{E}, x)$ satisfying $\int_{X \times E} b \, d\nu < \infty$. That is,

$$\lim_{n \to \infty} \frac{1}{n}\sum_{k=1}^{n} g(\epsilon_k, x_k) = \int_X \int_E g(\mathcal{E}, x)d\nu(\mathcal{E}, x)$$

6

almost always, for all continuous functions g such that $|g(\mathcal{E}, x)| < b(\mathcal{E}, x)$. Moreover, it is assumed that for each $x \in X$, there exists a neighborhood N_x such that

$$\int_E \sup_{N_x} b(\mathcal{E}, x) dP(\mathcal{E}) < \infty.$$

The assumption (A5$'$) is stronger than the assumption (A5) as it supposes not only the existence of a dominating function, but also involves the behavior of the probability distribution P, which is generally unknown in practice. The practical importance of the dominating function b arises in the proof of consistency for the least squares estimator (see Theorem 3.1 below). As has been noted elsewhere (see, e.g., [1], [24, Ch. 12]), the strong consistency of the estimator is proved by arguing that $J^*(q)$ is the almost sure limit of $J_n^{OLS}(q)$. Thus, if $J_n^{OLS}(q)$ is 'close' to $J^*(q)$ and $J^*(q)$ is uniquely minimized by q^*, it makes intuitive sense that q_n^{OLS}, which minimizes $J_n^{OLS}(q)$, should be close to q^*. This task is made difficult by the fact that the null set (from the 'almost sure' statement) may depend on the parameter q. In [2], the almost sure convergence of $J_n^{OLS}(q)$ to $J^*(q)$ is demonstrated constructively; that is, by building a set $A \in \mathcal{F}$ (which does not depend on q) with $P(A) = 1$ such that $J_n^{OLS}(q) \to J^*(q)$ for each $\omega \in A$ and for each $q \in Q$. This construction relies upon the separability of the parameter space Q (assumption (A2)) as well as the compactness of the space X (assumption (A3)). The alternative approach of Gallant uses a consequence of the Glivenko-Cantelli theorem [19, pg. 158] to demonstrate a uniform (with respect to q) strong law of large numbers. The proof relies upon the dominated convergence theorem, and hence the dominating function b. As a result, Gallant does not need the space Q to be separable or the space X to be compact. It should be noted, however, that in most practical applications of interest Q and X are compact subsets of Euclidean space so that relaxing these assumptions provides little advantage.

The choice of dominating function $b(\mathcal{E}, x)$ deserves further comment. Define the functional $h : E \times X \to \mathbb{R}$ by

$$h(\mathcal{E}, x) = (\mathcal{E} + f(x, q^*) - f(x, q))^2.$$

Then $b(\mathcal{E}, x)$ is chosen so that h, as well as its derivatives $\frac{\partial h}{\partial q_i}$, $\frac{\partial^2 h}{\partial q_i \partial q_j}$ and the product terms $\frac{\partial h}{\partial q_i} \frac{\partial h}{\partial q_j}$ (for all combinations of $1 \le i, j \le p$), are dominated by b for all $q \in Q$ [19, pg. 256]. The domination of the function h by the ν-integrable function b can be used to obtain a number of desirable results, including uniform convergence and the interchange of differentiation and integration [19, Lemma 3]. As a result, one does not need a positive definite hessian matrix $\mathcal{J}$ to prove the asymptotic normality of scores [19, Thm. 4] (contra. [2, Thm. 4.1]). Rather, one only needs $\mathcal{J}$ to be nonsingular in order for the conclusion of Theorem 3.2 below to be well-defined.

While the list of assumptions above is extensive, we remark that the set is not overly restrictive. Assumptions (A2) and (A3) are naturally satisfied for most problem formulations (although the requirement $q^* \in int(Q_{ad})$ may be occasionally problematic [2, Remark 4.4]). Assumptions (A4) and (A8) are easily checked. Though assumption (A1) may be difficult to verify, it is much less restrictive than, say, a complete likelihood specification. Moreover, residual plots [13, Ch. 3] can aid in assessing the reliability of the assumption.

Assumption (A5) is more difficult to check in practice as one does not know the limiting distribution μ. Of course, this is simply an assumption regarding the manner in which data is sampled (in the independent variable space X)–namely, that it must be taken in a way that 'fills up the space' in an appropriate sense. Similarly, assumptions (A6) and (A7) cannot be verified directly as one knows neither μ nor q^*. In many practical applications of interest, μ is Lebesgue measure, and one can assess the assumptions at q_n^{OLS} (which is hopefully close to q^*). Of course, if assumption (A7) holds, then assumption (A6) must hold at least for a small region around q^*, though possibly not on all of Q_{ad}.

As noted above, assumption (A7) is not strictly necessary if one uses assumption (A5$'$) in the place of (A5). Given assumptions (A2)–(A4), it follows that the function h (and its relevant derivatives; see above) is bounded (and hence dominated by a ν-measurable function) provided the *space E* in which the random variables $\mathcal{E}_k$ take their values is *bounded*. On one hand, this has the desirable effect of weakening the assumptions placed on the Hessian matrix $\mathcal{J}$. Yet the assumption that E is bounded *precludes* certain error models, in particular *normally distributed errors*.

Finally, we remark that while we only consider the linear restriction $Hq^* = h$, it is possible to extend the results above to include nonlinear restrictions of the form $\tilde{H}(q^*) = \tilde{h}$ [19, pg. 47ff]. In such a case, one is interested

in the restricted parameter space

$$\tilde{Q}_H = \left\{ q \in Q_{ad} : \tilde{H}(q) = \tilde{h} \right\}.$$

Assuming the null hypothesis is true, one can construct the linearization

$$L(q) = \tilde{H}(q^*) + \frac{\partial \tilde{H}(q^*)}{\partial q}(q - q^*)$$

$$= \tilde{h} + \frac{\partial \tilde{H}(q^*)}{\partial q}(q - q^*).$$

Then

$$\tilde{Q}_H \approx \left\{ q \in Q_{ad} : L(q) = \tilde{h} \right\}.$$

But the condition $L(q) = \tilde{h}$ is equivalent to the condition that

$$\frac{\partial \tilde{H}(q^*)}{\partial q} q = \frac{\partial \tilde{H}(q^*)}{\partial q} q^*.$$

Thus, under the null hypothesis, the nonlinear restriction $\tilde{H}(q^*) = \tilde{h}$ is asymptotically locally equivalent to the linear restriction of assumption (A8) with $H = \frac{\partial \tilde{H}(q^*)}{\partial q}$ and $h = \frac{\partial \tilde{H}(q^*)}{\partial q} q^*$. As such, we only consider the problem of testing a linear hypothesis. For nonlinear hypotheses, the results presented here are accurate to the order of the linear approximations above.

We now give several results which summarize the asymptotic properties of the ordinary least squares estimator.

Theorem 3.1. *Given assumptions (A1)–(A6), $q_n^{OLS} \to q^*$ with probability one. That is,*

$$P\left(\left\{\omega \in \Omega \middle| q_n^{OLS}(\omega) \to q^* \right\}\right) = 1.$$

This theorem states that the ordinary least squares estimator is consistent. We remark that the finite dimensionality of the parameter space Q (see assumption (A2)) is not necessary in the proof of this theorem, and it is sufficient for the function f to be continuous from Q into $C(X)$ rather than having two continuous derivatives.

Given Theorem 3.1, the following theorems may also be proven.

Theorem 3.2. *Given assumptions (A1)–(A7), as $n \to \infty$*

$$\sqrt{n}\left(q_n^{OLS} - q^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\sigma^2 \mathcal{J}^{-1}\right),$$

where the convergence is in distribution, and occurs with probability one.

Theorem 3.3. *Given assumptions (A1)–(A8) and assuming H_0 is true, then as $n \to \infty$*

$$\sqrt{n}\left(q_{n,H}^{OLS} - q_n^{OLS}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathcal{J}^{-1} V \mathcal{J}^{-1}\right)$$

where

$$V = 2\sigma^2 H^T (H \mathcal{J}^{-1} H^T)^{-1} H.$$

Again, the convergence occurs with probability one.

In effect, Theorem 3.3 states that the restricted and unrestricted OLS estimators will converge to one another provided H_0 is true. Now we can turn to the main result of interest. Define

$$U_n^{OLS} = \frac{n\left(J_n^{OLS}(q_{n,H}^{OLS}) - J_n^{OLS}(q_n^{OLS})\right)}{J_n^{OLS}(q_n^{OLS})}.$$

Theorem 3.4. *Under assumptions (A1)–(A8) and assuming H_0 is true, then as $n \to \infty$*

$$U_n^{OLS} \xrightarrow{\mathcal{D}} \chi^2(r)$$

with probability one, where $\chi^2(r)$ is a Chi-square distributed with r degrees of freedom.

Proofs of Theorems 3.1–3.4 can be found in [2]. We remark again that very similar results (under slightly differing assumptions) can be found in [19] where it is shown that the statistic

$$\tilde{U}_n^{OLS} = \frac{(n-p)\left(J_n^{OLS}(q_{n,H}^{OLS}) - J_n^{OLS}(q_n^{OLS})\right)}{J_n^{OLS}(q_n^{OLS})}$$

is asymptotically distributed as $\chi^2(r)$ [19, pg. 236]. The two model comparison statistics U_n^{OLS} and $\tilde{U}_n^{OLS}$ are asymptotically equivalent, the difference in scaling factors amounting to a bias correction in the estimation of the variance parameter σ^2. See [2, Thm. 4.6] and [19, pg. 236] for details.

The practical importance of this theorem is that one has a quantitative basis upon which to assess the null hypothesis. Because the unrestricted model contains the restricted model as a special case, we must have $J_n^{OLS}(\hat{q}_{n,H}^{OLS}) \geq J_n^{OLS}(\hat{q}_n^{OLS})$ and thus $\hat{u}_n^{OLS} \geq 0$ where $\hat{u}_n^{OLS}$ is a realization of U_n^{OLS}. Heuristically, U_n^{OLS} is used to indicate the relative likelihood that the results obtained from a given realization of the data are the result of random chance. A full discussion of the use and interpretation of the test statistic is given in [2, pg. 518–519] and [13, Sec. 3.5.1]. Significantly, the terms $J_n^{OLS}(\hat{q}_n^{OLS})$ and $J_n^{OLS}(\hat{q}_{n,H}^{OLS})$ used to compute $\hat{u}_n^{OLS}$ are a natural byproduct of the optimizations (6) and (10) and can be returned directly by most optimization software. It should also be noted that the limiting distribution for U_n^{OLS} is independent of the model parameterization. As such, the results are robust to any reparameterization or change of variables in the mathematical model $f(x; q)$.

4 Extension to Weighted Least Squares

The results presented above are quite useful, but they only apply to the class of problems in which the error random variables are independent and identically distributed with constant variance σ^2. While the assumption of independence is common, there are many practical cases of interest in which these random variables are not identically distributed. As discussed in Section 2, one may encounter a weighted least squares problem in which $E[\mathcal{E}_k] = 0$ for all k but $Var(\mathcal{E}_k) = \sigma^2 w(x_k)^2$. In such a case, the results above (namely, assumption (A1)) fail to apply directly. See, e.g., [7, Sec. 2] and references therein as well as [13] for more information on the development of error models.

In order to extend the results presented above to independent, heteroscedastic error models (and hence, to weighted least squares problems), we turn to a technique suggested by Gallant [19, pg. 124] in which one defines a change of variables in an attempt to normalize the heteroscedasticity of the random variables. As has been noted previously, Gallant used this technique under a different set of assumptions in order to obtain results similar to those presented above. This change of variables technique will allow us to extend the results above, originally from [2], in a rigorous fashion.

Consider the following assumptions.

(A1′a) The error random variables $\{\mathcal{E}_k\}$ are independent and have central moments which satisfy $E[\vec{\mathcal{E}}] = \vec{0}$, $Cov(\vec{\mathcal{E}}) = \sigma^2 W = \sigma^2 diag(w(x_1)^2, \ldots, w(x_n)^2)$.

(A1′b) The distributions of the random variables $\{\mathcal{E}_k\}$ are completely determined by the first two central moments.

(A1′c) The function $w : X \to \mathbb{R}^+$ is continuous.

(A7′) The matrix

$$\tilde{\mathcal{J}} = 2 \int_X \frac{1}{w(x)^2} \left(\frac{\partial f(x; q^*)}{\partial q} \frac{\partial f(x; q^*)^T}{\partial q} \right) d\mu(x)$$

is positive definite.

Theorem 4.1. *Under assumptions (A1'a)–(A1'c), (A2)–(A6), (A7'), and (A8),*

1. $q_n^{WLS} \to q^*$ *with probability one, and*

2. $\sqrt{n}\left(q_n^{WLS} - q^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\sigma^2 \tilde{\mathcal{J}}^{-1}\right).$

Assuming further that H_0 is true,

3. $\sqrt{n}\left(q_{n,H}^{WLS} - q_n^{WLS}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \tilde{\mathcal{J}}^{-1}\tilde{V}\tilde{\mathcal{J}}^{-1}\right)$

where

$$\tilde{V} = 2\sigma^2 H^T (H\tilde{\mathcal{J}}^{-1}H^T)^{-1} H.$$

Again, the convergence occurs with probability one. Finally, defining

$$U_n^{WLS} = \frac{n\left(J_n^{WLS}(q_{n,H}^{WLS}) - J_n^{WLS}(q_n^{WLS})\right)}{J_n^{WLS}(q_n^{WLS})},$$

4. $U_n^{WLS} \xrightarrow{\mathcal{D}} \chi^2(r)$ *with probability one.*

Proof. We first recall the statistical model (2)

$$Y_k = f(x_k; q^*) + \mathcal{E}_k,$$

which can be written in vector form (7) as

$$\vec{Y} = \vec{f}(q^*) + \vec{\mathcal{E}},$$

where now $E[\vec{\mathcal{E}}] = \vec{0}$ and $Cov(\vec{\mathcal{E}}) = \sigma^2 W$. Define $L = diag(w(x_1), \dots, w(x_n))$. It follows that $LL^T = L^2 = W$ (L is the Cholesky decomposition of W). Also, L^{-1} exists and can be applied to both sides of (7) to obtain

$$L^{-1}\vec{Y} = L^{-1}\vec{f}(q^*) + L^{-1}\vec{\mathcal{E}}$$

$$\text{or}$$

$$\vec{Z} = \vec{h}(q^*) + \vec{\eta}, \tag{11}$$

where $\vec{Z}$, $\vec{h}$ and $\vec{\eta}$ have the obvious definitions. The OLS cost functional for the transformed model and data (11) is

$$\begin{aligned}
J_n^{OLS}(q) &= \frac{1}{n}\left(\vec{Z} - \vec{h}(q)\right)^T \left(\vec{Z} - \vec{h}(q)\right) \\
&= \frac{1}{n}\left(L^{-1}\vec{Y} - L^{-1}\vec{f}(q)\right)^T \left(L^{-1}\vec{Y} - L^{-1}\vec{f}(q)\right) \\
&= \frac{1}{n}\left(\vec{Y} - \vec{f}(q)\right)^T L^{-1}L^{-1} \left(\vec{Y} - \vec{f}(q)\right) \\
&= \frac{1}{n}\left(\vec{Y} - \vec{f}(q)\right)^T W^{-1} \left(\vec{Y} - \vec{f}(q)\right).
\end{aligned}$$

In other words, the ordinary least squares cost function with respect to the transformed model and data (11) is exactly the weighted least squares cost function (8) for the original model and data. Thus, in order to prove Theorem 4.1, we will show that the rescaled model (11) satisfies the assumptions of Theorems 3.1–3.4.

Clearly, $E[\vec{\eta}] = \vec{0}$ and $Cov(\vec{\eta}) = \sigma^2 L^{-1}WL^{-1} = \sigma^2 I_n$. Thus the random variables η_k (the components of $\vec{\eta}$) are independent and identically distributed (by assumption (A1'b)) with constant variance, and thus assumption (A1) is satisfied. Assumptions (A2), (A3), (A5), and (A8) are unchanged. For assumption (A4), we must show

that the function $h(x; q) = f(x; q)/w(x) \in C^2(Q, C(X))$. This follows from assumption (A1'c) and the fact that $w(x)$ does not depend on q. For assumption (A6), we must show

$$\tilde{J}^*(q) = \sigma^2 + \int_X (h(x; q^*) - h(x; q))^2 \, d\mu(x)$$

$$= \sigma^2 + \int_X \left(\frac{f(x; q^*) - f(x; q)}{w(x)} \right)^2 d\mu(x)$$

has a unique minimum at $q = q^*$. Clearly, $\tilde{J}^*(q^*) = \sigma^2$. Since the function $J^*(q)$ (see assumption (A6)) has a unique minimum at $q = q^*$ and $w(x) > 0$, it follows immediately that $\tilde{J}^*(q^*) > \sigma^2$ if $q \neq q^*$ so that $\tilde{J}^*(q)$ has a unique minimum at $q = q^*$. Assumption (A7) is satisfied for the formulation (11) directly by assumption (A7'). $\qquad\square$

In fact, the proof of Theorem 4.1 applies to any set of observations in which a change of variables can be used to produce a set of error random variables which are independent and identically distributed. The weighted least squares problem addressed in the above discussion arises from an observation process in which the errors are assumed to be independent but are not necessarily identically distributed. By rescaling the observations in accordance with their variance (which is assumed to be known) one obtains error random variables which are identically distributed as well as independent. Even more generally, it is not strictly necessary that the observations be independent. For instance, one might have observations generated by an autoregressive process of order r. Then, by definition, some linear combination of r observational errors will give rise to errors which are independent and identically distributed. This linear combination is exactly the change of variables necessary to obtain a model which is suitable for ordinary least squares. Thus, even in the most general situation, when one has $Cov(\vec{\mathcal{E}}) = R$, one may still use the Cholesky decomposition in the manner discussed above, provided one has sufficient assumptions regarding the underlying error process. See [19, Ch. 2] for details.

As in Section 3, the assumption (A7') is the most problematic to verify in practice. In the proof above for the weighted least squares problem, the assumption (A1) has been replaced with assumptions (A1'a)–(A1'c), a change which merely accounts for the heteroscedastic statistical model. Then the assumptions (A2)–(A6) and (A8) for the rescaled model (11) can be verified directly from the original assumptions (A2)–(A6) and (A8) for the ordinary least squares formulation, as shown above. The only exception is the assumption (A7), which cannot be established directly from the ordinary least squares assumptions, hence the need for assumption (A7'). On one hand, the existence of a unique minimum (assumption (A6)) is sufficient to prove that the matrix $\mathcal{J}$ or $\tilde{\mathcal{J}}$ must be positive *semi*-definite, so that the assumption (A7) or (A7') may not be overly restrictive. Alternatively, as has been noted before, one can relax the assumptions (A7) or (A7') by assuming the existence of a dominating function b. Moreover, provided the weights $w(x)$ satisfy the slightly stronger requirement $w(x) \geq \delta > 0$ for all $x \in X$, then one can use the dominating function b (from the ordinary least squares problem) to obtain a new dominating function $\tilde{b}(\mathcal{E}, x) = \delta^{-1} b(\mathcal{E}, x)$ which is also ν-integrable. Even in this case, though, one must still make the additional assumption that $\tilde{\mathcal{J}}$ is invertible.

All of the results above also readily extend to systems governed by partial differential equations; the essential elements of the theory are the form of the discrete observation operator for the dynamical system and the statistical model as given in (2).

5   Example: Hypothesis Testing in Models of Cell Proliferation

We now consider a practical application of model comparison for a weighted least squares problem. We consider the problem of modeling flow cytometry data for a dividing population of lymphocytes labeled with the intracellular dye CFSE. As cells divide, the highly fluorescent intracellular CFSE is partitioned evenly between two daughter cells. A flow cytometer measures the fluorescence intensity of labeled cells as a surrogate for the mass of CFSE within a cell, thus providing an indication of the number of times a cell has divided. Most commonly, the data is presented in histograms showing the distribution of CFSE in the measured population of cells at each measurement time. A sample data set is shown in Figure 1. See [12, 25] and the references therein for a detailed overview of the experimental procedure.
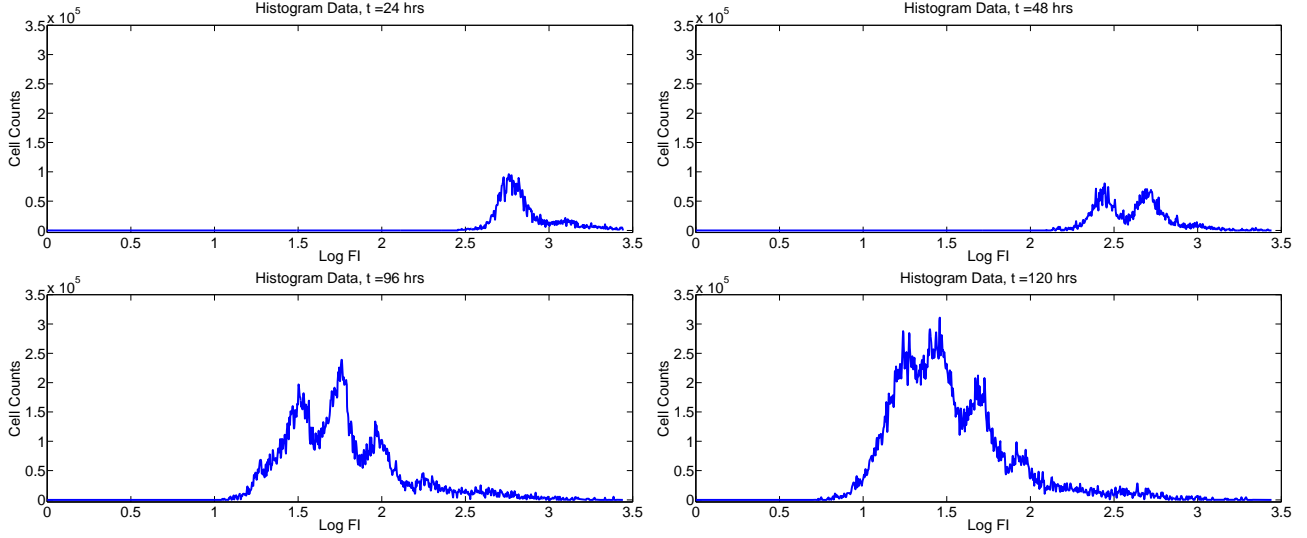
Figure 1: Typical histogram data for a CFSE-based proliferation assay showing cell count vs. log fluorescence intensity; originally from [22].

The mathematical modeling problem is to develop a system of equations which accounts for the manner in the population distribution of CFSE evolves as cells divide, die, and slowly lose fluorescence intensity as a result of protein turnover. In doing so, one must account for not only the division and death dynamics of the population of cells, but also how these processes relate to the observed distribution of fluorescence intensities. Let $\tilde{n}_i(t, \tilde{x})$ represent the label-structured density (cells per unity of intensity, cells/UI) of cells at time t having completed i divisions and having fluorescence intensity $\tilde{x}$ from *intracellular CFSE* (that is, not including the contribution of cellular autofluorescence). Then it can be shown that this density can be modeled in terms of two independently functioning components, one accounting for the division dynamics of the population of cells, and one accounting for the intracellular processing of CFSE [20, 23]. It has been shown that the slow loss of fluorescence intensity as a result of intracellular protein turnover can be well-modeled by a Gompertz growth/decay process [9, 25]. It has also been shown that the highly flexible cyton model [21] can very accurately describe the evolving generation structure (cells per number of divisions undergone) [12]. Thus we consider the model

$$\tilde{n}(t, \tilde{x}) = N_i(t)\bar{n}_i(t, \tilde{x}) \tag{12}$$

The functions $\bar{n}_i(t, \tilde{x})$ each satisfy the partial differential equation

$$\frac{\partial \bar{n}_i(t, \tilde{x})}{\partial t} - ce^{-kt}\frac{\partial[\tilde{x}\bar{n}_i(t, \tilde{x})]}{\partial \tilde{x}} = 0 \tag{13}$$

with initial condition

$$\bar{n}_i(0, \tilde{x}) = \frac{2^i \Phi(2^i \tilde{x})}{N_0},$$

where $\Phi(x)$ is the label-structured density of the cells at $t = 0$, and N_0 is the number of cells in the population at $t = 0$, all of which are assumed to be undivided. The functions $N_i(t)$ are described by the cyton model,

$$N_0(t) = N_0 - \int_0^t \left(n_0^{div}(s) - n_0^{die}(s)\right) ds$$

$$N_i(t) = \int_0^t \left(2n_{i-1}^{div}(s) - n_i^{div}(s) - n_i^{die}(s)\right) ds, \tag{14}$$

12

where $n_i^{div}(t)$ and $n_i^{die}(t)$ indicate the numbers per unit time of cells having undergone i divisions that divide and die, respectively, at time t. In this regard, the cyton model (and hence, the mathematical model as a whole) can be considered as a large class of models differentiated by the mechanisms one uses to describe the terms $n_i^{div}(t)$ and $n_i^{die}(t)$, which we consider in more detail below. We remark that the mathematical model (12) can be equivalently characterized by the system of partial differential equations

$$\frac{\partial \tilde{n}_0}{\partial t} - ce^{-kt}\frac{\partial \tilde{x}\tilde{n}_0}{\partial \tilde{x}} = \left(n_0^{div}(t) - n_0^{die}(t)\right)\bar{n}_0(t, \tilde{x})$$

$$\frac{\partial \tilde{n}_1}{\partial t} - ce^{-kt}\frac{\partial \tilde{x}\tilde{n}_1}{\partial \tilde{x}} = \left(2n_0^{div}(t) - n_1^{div}(t) - n_1^{die}(t)\right)\bar{n}_1(t, \tilde{x})$$

$$\vdots$$

Given the solution $\tilde{n}(t, \tilde{x}) = \sum \tilde{n}_i(t, \tilde{x})$ in terms of the fluorescence resulting from intracellular CFSE, one must add in the contribution of cellular autofluorescence to obtain a population density structured by total fluorescence intensity, which is the quantity measured by a flow cytometer. This density can be computed as

$$n(t, x) = \int_0^\infty \tilde{n}(t, \tilde{x})p(x - \tilde{x})d\tilde{x},$$

where $p(x_a)$ is a density function describing the distribution of cellular autofluorescence [20].

Let $\phi_0(t)$ and $\psi_0(t)$ be probability density functions (in units 1/hr) for the time to first division and time to death, respectively, for an undivided cell. Let F_0, the initial precursor fraction, be the fraction of undivided cells which would hypothetically divide in the absence of any cell death. It follows that

$$n_0^{div}(t) = F_0 N_0 \left(1 - \int_0^t \psi_0(s)ds\right)\phi_0(t)$$

$$n_0^{die}(t) = N_0 \left(1 - F_0 \int_0^t \phi_0(s)ds\right)\psi_0(t). \tag{15}$$

Similarly, one can define probability density functions $\phi_i(t)$ and $\psi_i(t)$ for times to division and death, respectively, for cells having undergone i divisions (he cellular 'clock' begins at the completion of the previous division), as well as the progressor fractions F_i of cells which would complete the ith division in the absence of cell death. Then

$$n_i^{div}(t) = 2F_i \int_0^t n_{i-1}^{div}(s)\left(1 - \int_0^{t-s}\psi_i(\xi)d\xi\right)\phi_i(t-s)ds$$

$$n_i^{die}(t) = 2\int_0^t n_{i-1}^{div}(s)\left(1 - F_i\int_0^{t-s}\phi_i(\xi)d\xi\right)\psi_i(t-s)ds. \tag{16}$$

It is assumed that the probability density functions for times to division are lognormal, and that undivided cells and divided cells have two separate density functions. Similarly, it is assumed that divided cells die with a lognormal probability density. Thus

$$\phi_0(t) = \frac{1}{t\sigma_0^{div}\sqrt{2\pi}}\exp\left(-\frac{(\log t - \mu_0^{div})^2}{2(\sigma_0^{div})^2}\right)$$

$$\phi_i(t) = \frac{1}{t\sigma_i^{div}\sqrt{2\pi}}\exp\left(-\frac{(\log t - \mu_i^{div})^2}{2(\sigma_i^{div})^2}\right) \quad (i \geq 1)$$

$$\psi_i(t) = \frac{1}{t\sigma_i^{die}\sqrt{2\pi}}\exp\left(-\frac{(\log t - \mu_i^{die})^2}{2(\sigma_i^{die})^2}\right) \quad (i \geq 1).$$

We treat cell death for undivided cells as a special case. It is assumed that the density function describing cell death for undivided cells is

$$\psi_0(t) = \frac{F_0}{t\sigma_0^{die}\sqrt{2\pi}}\exp\left(-\frac{(\log t - \mu_0^{die})^2}{2(\sigma_0^{die})^2}\right) + (1 - p_{idle})(1 - F_0)\beta e^{-\beta t}, \tag{17}$$

where p_{idle} is the fraction of cells which are will neither die nor divide over the course of the experiment. The form of the function above arises from the assumption that progressing cells die with a lognormal probability density. Nonprogressing cells die at an exponential rate, except for the fraction of idle cells. For the purposes of this paper, we are interested in testing the null hypothesis $p_{idle} = 0$ against the alternative hypothesis $p_{idle} \neq 0$. A more detailed discussion of the mathematical model can be found in [11].

With the mathematical model established, we now turn our attention to a statistical model of the data. Let N_k^j be a random variable representing the number of cells measured at time t_j with log-fluorescence intensity in the region $[z_k, z_{k+1})$. Let q be the vector of parameters of the model (12), so that we may rewrite $n(t, x) = n(t, x; q)$. Define the operator

$$I[n](t_j, z_k; q) = \int_{z_k}^{z_{k+1}} 10^z \ln(10) n(t, 10^z; q) dz,$$

which represents the computation of cell counts from the structured density $n(t, x)$ transformed to the logarithmic coordinate $z = \log_{10} x$. Then a statistical model linking the random variables N_k^j to the mathematical model is

$$N_k^j = \lambda_j I[\hat{n}](t_j, z_k^j; q_0) + \lambda_j \left(I[\hat{n}](t_j, z_k^j; q_0) \right)^\gamma \mathcal{E}_{kj} \tag{18}$$

where the λ_j are scaling factors [25, Ch. 4] and the random variables $\mathcal{E}_{kj}$ satisfy assumption (A1). In principle, it is possible to use a multi-stage estimation procedure to determine the values of the statistical model parameters λ_j and γ [17] but we do not consider that here. It is assumed that $\lambda_j = 1$ for all j.

In the case that $\gamma = 0$, one has precisely absolute error and a statistical model suitable for the ordinary least squares cost function (5) and estimator (6) to estimate the unknown parameter q_0. The results of using the OLS estimator can be found in Figure 2 for the null and alternative hypotheses. The ordinary least squares cost of fitting the model (12) under the null hypothesis is $J^{OLS}(\hat{q}_H^{OLS}) = 4.0423 \times 10^{11}$. The ordinary least squares cost under the alternative hypothesis is $J^{OLS}(\hat{q}^{OLS}) = 3.6228 \times 10^{11}$ [11]. We would like to determine whether or not the small decrease in cost associated with the alternative hypothesis is outweighed by the larger number of parameters (17 versus 16) required. As seen in Figure 2, the fits of the two models to the data are very similar, so that one cannot justify a conclusion solely on the basis of the slightly reduced cost. Rather, one should use a model comparison test.

However, as the theory of Sections 3 and 4 suggests, one cannot rigorously apply a model comparison test unless the underlying statistical model is accurate. It follows from the form of Equation (18) that when $\gamma = 0$ the residuals $r_k^j = I[n](t_j, z_k^j; \hat{q}^{OLS}) - n_k^j$ (where the n_k^j are realizations of the data random variables N_k^j) correspond to realizations of the error terms $\mathcal{E}_k^j$. As such, these residuals should appear random when plotted as a function of the magnitude of the model solution [13, Ch. 3]. However, we see in Figure 3 (top) that this is not the case. There is a noticeable increase in the variance of the residuals as the magnitude of the model solution increases. Thus we must conclude that the constant variance model $\gamma = 0$ is not correct, and that the ordinary least squares cost function is not adequate as a basis for model comparison. A constant coefficient of variation (CCV) model ($\gamma = 1$, usually called *relative error*) leading to a generalized least squares (GLS) problem formulation was also considered in [25]. Again residual analysis suggested that this was not the correct formulation.

In fact, the most appropriate theoretical value of γ appears to be $\gamma = 1/2$ [25] so that the weighted least squares cost function (3) and estimator (4) should be used. Because the weights must be bounded away from zero (see assumption (A1'c)), we define the weights to be

$$w(t_j, z_k^j) = \begin{cases} \left(I[n](t_j, z_k^j; q^*) \right)^{1/2}, & I[n](t_j, z_k^j; q^*) > I^* \\ (I^*)^{1/2}, & I[n](t_j, z_k^j; q^*) \leq I^* \end{cases}.$$

The cutoff value $I^* > 0$ is determined by the experimenter so that the resulting residuals appear random. In the work that follows, $I^* = 1 \times 10^4$. Because w_k^j depends upon the unknown parameter q^*, an iterative optimization scheme must be used (see [13, Ch. 3] or [17, Ch. 2] for details). Using an estimate of q^* (say, the ordinary least squares estimate $\hat{q}^{OLS}$), one computes the weights according to the formula above (with q^* replaced by $\hat{q}^{OLS}$) and then uses these (fixed) weights to compute the weighted least squares estimate $\hat{q}^{WLS}$. One can then use this updated estimate to compute new weights, and continue the process iteratively until convergence is achieved.
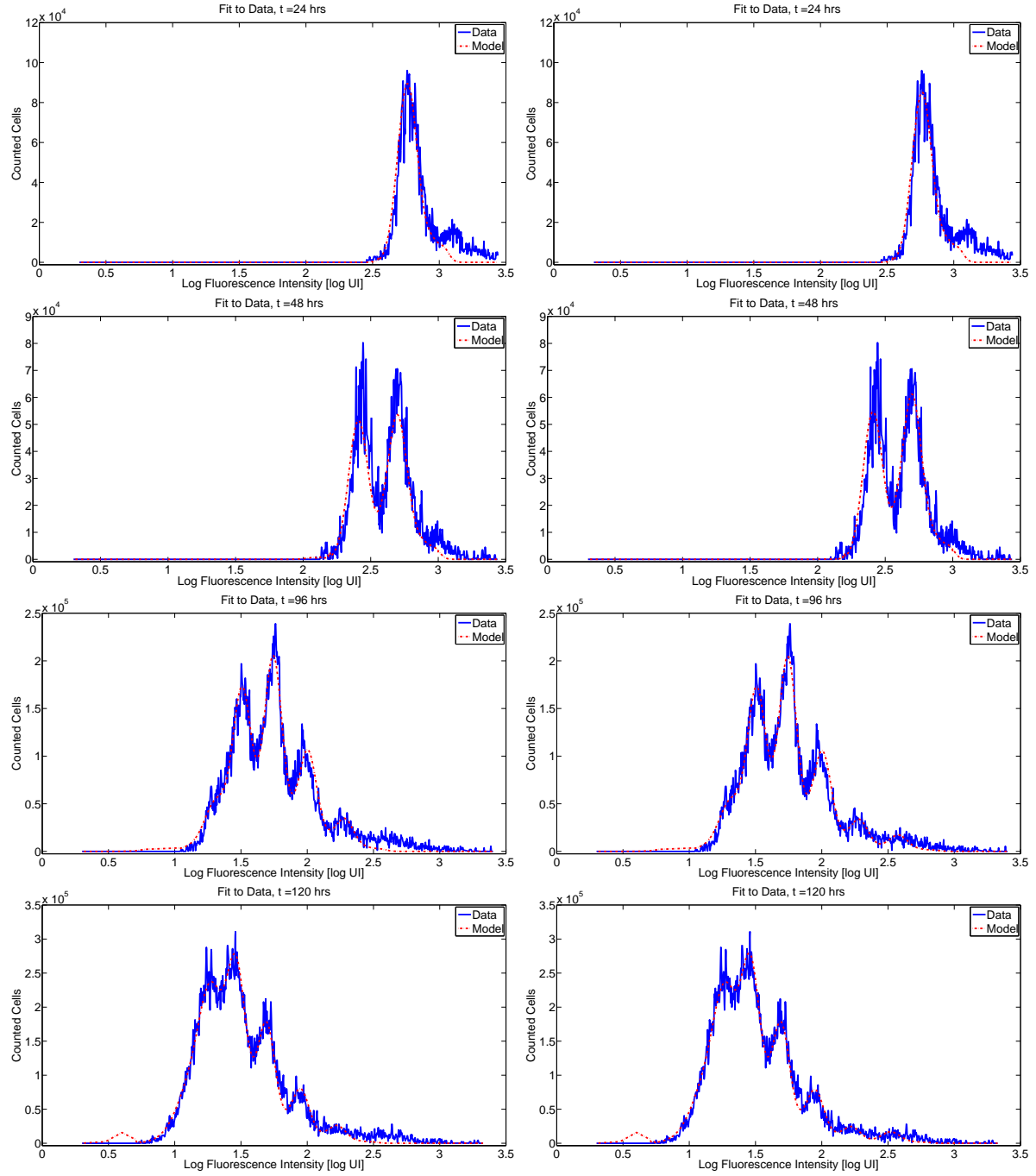
Figure 2: Left: results of fitting the model (12) under the null hypothesis using ordinary least squares. Right: results of fitting the model (12) under the alternative hypothesis using ordinary least squares. Both fits to data are superb, and it is difficult to distinguish between the two models from these graphics.
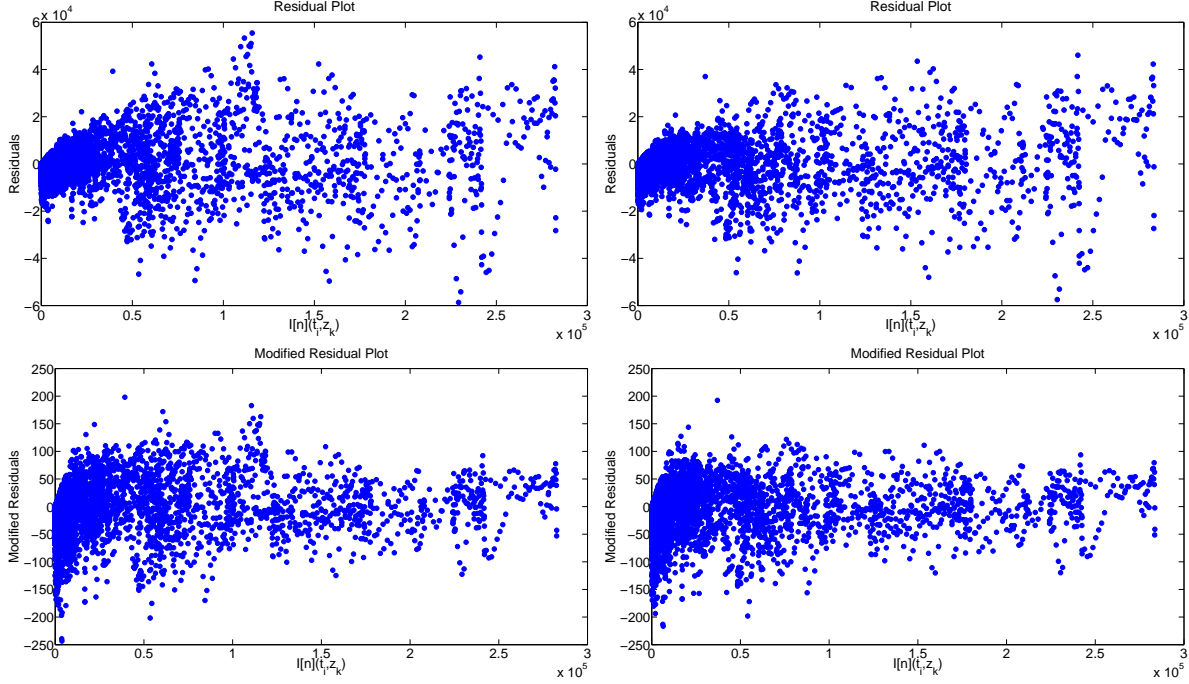
Figure 3: Top: residual plots for the mathematical model under the null (left) and alternative (right) hypotheses, showing the residuals in terms of the magnitude of the model solution. Bottom: modified residual plots for the mathematical model under the null (left) and alternative (right) hypotheses, showing the modified residuals in terms of the magnitude of the model solution. While the variance of the standard residuals is observed to grow with the magnitude of the model solution, we see that the variance of the modified residuals is largely constant.

Residual plots obtained using the theoretical value of $\gamma = 1/2$ are shown in Figure 3 (bottom). (The weighted least squares fit to data is not shown, as the results are sufficiently similar to the ordinary least squares fits of Figure 2.) While the variance of the standard residuals is observed to increase with the magnitude of the model solution, the variance of the modified residuals is approximately constant, providing some evidence that the statistical model (18) (with $\gamma = 1/2$) is correct. As such, we are prepared to use the model comparison test statistic to analyze the null hypothesis.

Under the null hypothesis, the weighted least squares cost functional is $J^{WLS}(\hat{q}_H^{WLS}) = 10.3040 \times 10^6$. Under the alternative hypothesis, the weighted least squares cost functional is $J^{WLS}(\hat{q}^{WLS}) = 8.8394 \times 10^6$. Thus the model comparison statistic is $\hat{u}^{WLS} \approx 711.3$ (there are 4293 data points). We conclude then, that there is *overwhelming* evidence to reject the null hypothesis in favor of the alternative hypothesis. In the present application, this strongly suggests that the fraction of cells which remain idle over the course of the experiment is a significant mathematical feature which should be included in an accurate model of proliferation assay data. As with the cat brain diffusion-convection model discussed in the introduction, hypothesis testing has again allowed us to better justify the use of key model components.

This example illuminates some desirable features of the model comparison statistic. The statistic is very easy to calculate from the minimized least squares cost functionals and provides a theoretically sound basis on which to test a hypothesis. We also see that the test statistic can provide meaningful comparisons between mathematical models that are not easily distinguishable from the fits of those models to data (Figure 2) or from the ordinary least squares cost functions.

6 Concluding Remarks

In this note we have outlined how hypothesis testing can be used as a tool for model comparison for certain classes of models. Beginning with an ordinary least squares framework, we reviewed assumptions and theorems which form the basis for a rigorous quantitative theory of model comparison based upon the statistical properties of least squares estimators. Two similar approaches, one from [2] and one from [19], have been considered and the relative merits of the different assumptions involved have been reviewed. The former approach is more general in the sense that one does not need to find a dominating function $b(\mathcal{E}, x)$ in order to apply the theoretical results presented. Yet the tradeoff is that one must place a stronger (and much harder to verify in practice) assumption on the hessian matrix $\mathcal{J}$ when compared to the latter approach. It appears that the most advantageous approach will depend largely upon the specific context in which the theory is used.

It has also been shown that the techniques for model comparison by hypothesis testing can be extended from ordinary least squares to weighted least squares through a change of variables designed to account for a heteroscedastic error model. The result is *significant* as it *extends the utility of the model comparison test* to include situations in which the errors characterizing the observation process are *independent but not identically distributed*. As noted above, the same technique can be used to extend the result to account for *non-independent errors* as well, provided one has some additional information regarding the correlation between observations.

It is reassuring that the primary result concerning the model comparison statistic (either U_n^{OLS} or U_n^{WLS}) is essentially unchanged–one simply computes the relative difference between the cost functional minimized over the admissible parameter space and the cost functional minimized over the linearly restricted parameter space. In effect, one must only take care to properly account for the variance of the observation errors in the formulation of the cost function, and the resulting model statistic will be Chi-square distributed. The test is similar in form to ANOVA-type statistical tests [2] and depends only on the residual sums of squares (RSS) for the restricted and unrestricted models. This is convenient as the test is independent of model parameterization, and the RSS can be returned as a byproduct of optimization by most software packages. Because the test is designed for least squares type estimation, only the first two central moments of the statistical model need to be specified, as opposed to a conditional density function in maximum likelihood estimation procedures. The major computational drawbacks of the method are that one must compute two separate inverse problems (as opposed to use of the Wald test statistic [19]) and one can only compare nested models pairwise (as opposed to use of Akaike's Information Criterion [14]).

The unspecified form of the function $w(x)$ is meant to indicate its wide applicability. An important subclass of weighted least squares problems, sometimes termed generalized least squares [13, Ch. 3] occurs when the magnitude of the observation error is assumed to be directly proportional to the magnitude of the true model value. In that case, $w(x) = f(x; q^*)$. As seen in the example of Section 5, this results in an interesting computational problem. In the theory above it is tacitly assumed that the weights $w(x_k)$ are known exactly. However, one does not know the true parameter value q^*. While computational strategies exist to handle such situations [13, 17], the effects of this additional level of approximation are not considered in the theory presented here, and it is assumed that the weights, even when computed iteratively, are known exactly. More generally, the weight functions may depend upon some additional nuisance parameters, $w(x) = w(x; \tau)$ (in the example of Section 5, $\tau = (\{\lambda_j\}, \gamma)$) which must be estimated using a multi-stage procedure [17, 19].

There are some open problems in extending this theory further. For instance, assumption (A2) effectively limits the application of the results presented to models in which the underlying parameter space is a subset of finite-dimensional Euclidean space. Yet in some situations, for instance the nonparametric estimation of probability measures [7], the 'parameters' to be estimated are contained within an infinite-dimensional function space. While computational methodologies have been developed for such situations, the statistical properties of the least squares estimators have not. Thus one cannot rely upon any asymptotic results for the quantification of uncertainty, nor can one test hypotheses in the manner demonstrated above.

It should be clear by this point that model comparison tests have immense potential for use as a key tool when working with biological modeling applications. The results presented here and the references herein provide a rigorous foundation for careful use of these tests in practice. The wide applicability of the results presented here will continue to find use in biological modeling projects, as well as provide motivation for further theoretical study and expansion of model comparison testing ideas.

Acknowledgements

References

[1] Takeshi Amemiya, Nonlinear regression models, Ch. 6 in *Handbook of Econometrics, Volume I*, Z. Griliches and M. D. Intriligator, Eds. North Holland, Amsterdam, 333–389.

[2] 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.

[3] H.T. Banks, S. Hu, Z.R. Kenz, C. Kruse, S. Shaw, J.R. Whiteman, M.P. Brewin, S.E. Greenwald, and M.J. Birch, Material parameter estimation and hypothesis testing on a 1D viscoelastic stenosis model: Methodology, CRSC-TR12-09, North Carolina State University, April 2012; *J. Inverse and Ill-Posed Problems*, submitted.

[4] H.T. Banks and P. Kareiva, Parameter estimation techniques for transport equations with applications to population dispersal and tissue bulk flow models, *J. Math. Biol.*, **17** (1983), 253–272.

[5] H.T. Banks, P. Kareiva, and P.D. Lamm, Modeling insect dispersal and estimaing parameters when mark-release techniques may cause initial disturbances, *J. Math. Biol.*, **22** (1985), 259–277.

[6] H.T. Banks, P. Kareiva, and K. Murphy, Parameter estimation techniques for interaction and redistribution models: a predator-prey example, *Oecologia*, **74** (1987), 356–362.

[7] H.T. Banks, Z.R. Kenz, and W.C. Thompson, A review of selected techniques in inverse problem nonparametric probability distribution estimation, CRSC-TR12-13, North Carolina State University, May 2012; *J. Inverse and Ill-Posed Problems*, accepted.

[8] H.T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhauser, Boston, 1989.

[9] H.T. Banks, Karyn L. Sutton, W. Clayton Thompson, G. Bocharov, Marie Doumic, Tim Schenkel, Jordi Argilaguet, Sandra Giest, Cristina Peligero, and Andreas Meyerhans, A new model for the estimation of cell proliferation dynamics using CFSE data, CRSC-TR11-05, North Carolina State University, Revised July 2011; *J. Immunological Methods,* **373** (2011), 143–160; DOI:10.1016/j.jim.2011.08.014.

[10] H.T. Banks, Karyn L. Sutton, W. Clayton Thompson, Gennady Bocharov, Dirk Roose, Tim Schenkel, and Andreas Meyerhans, Estimation of cell proliferation dynamics using CFSE data, CRSC-TR09-17, North Carolina State University, August 2009; *Bull. Math. Biol.*, **70** (2011), 116–150.

[11] H.T. Banks and W. Clayton Thompson, A division-dependent compartmental model with cyton and intracellular label dynamics, CRSC-TR12-12, North Carolina State University, Revised August 2012; *Intl. J. Pure and Appl. Math*, **77** (2012), 119–147.

[12] H.T. Banks and W. Clayton Thompson, Mathematical models of dividing cell populations: Application to CFSE data, CRSC-TR12-10, North Carolina State University, April 2012; *J. Math. Modeling of Natural Phenomena*, to appear.

[13] H.T. Banks and H.T. Tran, *Mathematical and Experimental Modeling of Physical and Biological Processes*, CRC Press, Boca Raton London New York, 2009.

[14] K.P. Burnham and D.R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, (2nd Edition), Springer, New York, 2002.

[15] R.J. Carroll and D. Ruppert, *Transformation and Weighting in Regression*, Chapman Hall, London, 2000.

[16] G. Casella and R. Berger, *Statistical Inference*, Thomson Learning, Pacific Grove, CA, 2002.

[17] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman Hall, London, 2000.

[18] B.G. Fitzpatrick, Statistical methods in parameter identification and model selection. Ph.D. Thesis, Division of Applied Mathematics, Brown University, Providence, RI, 1988.

[19] A. Ronald Gallant, *Nonlinear Statistical Models*, John Wiley and Sons, New York, 1987.

[20] J. Hasenauer, D. Schittler, and F. Allgöwer, A computational model for proliferation dynamics of division- and label-structured populations, `arXive.org`, arXiv:1202.4923v1, 22 February 2012.

[21] E.D. Hawkins, M.L. Turner, M.R. Dowling, C. van Gend, and P.D. Hodgkin, A model of immune regulation as a consequence of randomized lymphocyte division and death times, *Proc. Natl. Acad. Sci.*, **104** (2007), 5032–5037.

[22] T. Luzyanina, D. Roose, T. Schenkel, M. Sester, S. Ehl, A. Meyerhans, and G. Bocharov, Numerical modelling of label-structured cell population growth using CFSE distribution data, *Theoretical Biology and Medical Modelling*, **4** (2007), Published Online.

[23] D. Schittler, J. Hasenauer, and F. Allgöwer, A generalized model for cell proliferation: Integrating division numbers and label dynamics, *Proc. Eighth International Workshop on Computational Systems Biology (WCSB 2011)*, June 2011, Zurich, Switzerland, p. 165–168.

[24] G.A. Seber and C.J. Wild, *Nonlinear Regression*, Wiley, Hoboken, NJ, 2003.

[25] W. Clayton Thompson, *Partial Differential Equation Modeling of Flow Cytometry Data from CFSE-based Proliferation Assays,* Ph.D. Dissertation, Dept. of Mathematics, North Carolina State University, Raleigh, December, 2011.

[26] D. Valdez-Jasso, H.T. Banks, M.A. Haider, D. Bia, Y. Zocalo, R.L. Armentano, and M.S. Olufsen, Viscoelastic models for passive arterial wall dynamics, CRSC TR08-24, North Carolina State University, December 2008; *Adv. in Applied Math. and Mech.*, **1.2** (2009), 151–165.