Estimation in regression models with externally estimated parameters

R. TODD OGDEN*
Department of Biostatistics, Columbia University,
6th floor, 722 West 168th Street, New York, NY 10032, USA
to166@columbia.edu

THADDEUS TARPEY
Department of Mathematics and Statistics,
Wright State University, Dayton, OH 45435, USA

SUMMARY
In many regression applications, some of the model parameters are estimated from separate data sources. Typically, these estimates are plugged into the regression model and the remainder of the parameters is estimated from the primary data source. This situation arises frequently in compartment modeling when there is an external input function to the system. This paper provides asymptotic and bootstrap-based approaches for accounting for all sources of variability when computing standard errors for estimated regression model parameters. Examples and simulations are provided to motivate and illustrate the ideas.

Keywords: Bootstrap; Compartment models; PET imaging.

1. INTRODUCTION
It is a general statistical practice to report standard errors associated with estimated model parameters. Typically, model parameters are estimated by a single data source, but there are many examples in which two or more sources are used. A common example occurs in chemistry and physics applications where nonlinear models are defined in terms of well-known constants (e.g. Planck’s constant and Faraday’s constant) whose estimates have been previously established. The uncertainties associated with such physical constants are generally very small. However, in other applications, some of the model parameters must be estimated by external data sources and the uncertainties associated with these estimates can be quite large. We shall call the parameters estimated by external data sets ‘input’ parameters. In routine applications, the variability associated with the estimation of input parameters is often ignored.

The following illustrate examples of regression models requiring input parameters that are estimated by separate data sets.

EXAMPLE 1 (POSITRON EMISSION TOMOGRAPHY IMAGING FOR NEURORECEPTOR MAPPING)
Positron emission tomography (PET) imaging of the brain often involves kinetic modeling (see, e.g.

*To whom correspondence should be addressed.
Gunn et al., 2001) or one of its variants (Ogden, 2003; Gunn et al., 2002). The primary data source is measurements made by the PET camera of the concentration of a ligand throughout the brain. In such an application, the objective is to estimate the density of the target receptor throughout the brain. This is accomplished by modeling the kinetics of the ligand and estimating the relevant rate constants. The density of the receptor may be estimated by computing a function of the estimated rate parameters. These estimated densities may then be compared among different patient populations or treatment groups. Since there is substantial noise involved with imaging and modeling, in addition to large intrasubject variability, these comparisons often lack power. With standard error estimates, power can be dramatically increased.

In general kinetic modeling, the ligand concentration at a given brain location at time \( t \), denoted \( C(t) \), is given in terms of a convolution:

\[
C(t) = g(t; \beta) \otimes C_p(t; \alpha),
\]

where \( g(t; \beta) \) represents the kinetics of the brain compartmental system and involves unknown rate parameters \( \beta \). The input function \( C_p(t; \alpha) \) represents the concentration of the ligand in the plasma at time \( t \) and the input parameters \( \alpha \) are estimated separately using data obtained directly from blood samples drawn during the imaging experiment.

**Example 2 (Zinc Metabolism)** Foster et al. (1979) used a kinetic model to study the early phases of zinc metabolism in humans using data from Aamodt et al. (1979). In this study, a subject ingests a zinc isotope and the concentration of zinc is modeled over time in various compartments. The objective is to construct a model for the movement of zinc throughout the body. Results from such a study are often displayed as a table of estimated rate parameters. Standard errors would help in assessing the identifiability of each parameter and provide a guide in determining an appropriate model.

Assuming that the rate at which zinc enters the urine is proportional to the amount of zinc in the plasma, the amount of administered zinc in the urine compartment is modeled as

\[
f(t) = e^{\beta \int_0^t g(s; \alpha) \, ds}.
\]

The plasma over time is modeled as a sum of four exponentials:

\[
g(t; \alpha) = \sum_{i=1}^{4} \alpha_{2i-1} \exp(-\alpha_{2i}t).
\]

In applications, the plasma concentration curve is fitted using standard nonlinear least-squares methodology on combined data from several subjects. Subsequently, estimates for the input parameters \( \alpha \) are plugged into (1.2) in order to estimate subject-specific parameters for the urine compartment.

**Example 3 (Environmental Studies of Toxicants)** The concentration of a toxicant in an organism measured over time \( C(t) \) is often modeled as

\[
C(t) = \frac{\beta_u}{\beta_d} C_e (1 - e^{\beta_d t}),
\]

where the parameters \( \beta_u \) and \( \beta_d \) are the uptake and elimination rates of the toxicant, respectively, in the organism and \( C_e \) is an input parameter to the model representing the toxicant level in the environment (e.g. Bailer et al., 2000; Wheeler and Bailer, 2003; Lotufo et al., 2000; Rand et al., 1995). The primary objective in such studies is often to determine the bioconcentration factor (BCF), defined as the ratio...
of $\beta_u$ to $\beta_d$. The BCF is a key concern of environmental scientists as it quantifies how much chemical accumulates in an organism relative to the environmental exposures. The constant $C_e$ is estimated separately from water or sediment samples and the uncertainty in estimating $C_e$ is generally ignored when ascertaining the uncertainty in the estimated BCF. However, the validity of statistical inferences in environmental studies concerning changes of chemical accumulations in organisms requires accurate standard errors for estimated BCFs, which in turn requires the incorporation of the variability in estimating $C_e$.

All these examples fall under the general heading of compartment models (e.g. Matis and Wehrly, 1994; Jacquez, 1985) which include input parameters that are estimated from separate data sources.

The problem of estimating input parameters from a separate data set is closely related to models where separate data are available for covariates measured with error. For example, Higgins et al. (1997) consider a two-step approach using maximum likelihood estimation where time-varying covariates are estimated in the first step and are then used to obtain estimates of the parameters of primary importance. A parametric bootstrap approach is used to obtain standard errors that account for the uncertainty of estimating the covariates. The covariates in their setup are analogous to our input parameters. Wu (2002) considers a similar model with the added complication of censoring, proposing a joint modeling approach that simultaneously estimates the covariates and model parameters using a Monte Carlo expectation-maximization (EM) algorithm. The estimated standard errors are then obtained by approximating the information matrix.

In Section 2, a general regression model is provided for situations involving input parameters estimated from separate data sources. Section 3 gives asymptotic results for the sampling distribution of the estimated parameters when the input parameters are estimated. The resulting asymptotic covariance matrix can be used to provide approximate standard errors of model parameters. Section 4 describes a bootstrap alternative to asymptotic standard error estimation which is particularly useful for small- or moderate-sized samples. The proposed methodologies are explored for two situations via simulation in Section 5, and some general discussion is given in Section 6.

2. THE BASIC MODEL

The three examples mentioned in Section 1 are special cases of the following general regression model: Let $Y_1, \ldots, Y_n$ denote a sample from a known regression model depending on unknown parameter vectors $\alpha$ and $\beta$:

$$Y_i = f(x_i; \alpha, \beta) + \epsilon_i, \quad i = 1, \ldots, n. \quad (2.1)$$

In (2.1), the $x_i$ are the fixed, known values of predictors and the $\epsilon_i$ are a sequence of i.i.d. mean zero random variables with $\text{Var}(\epsilon_i) = \sigma^2$. The parameters $\alpha$ are input parameters which are to be estimated from a separate data source. Primary interest is in estimating $\hat{\beta}$ and computing reliable standard errors for $\hat{\beta}$.

It is common practice to estimate $\beta$ from the $Y$ data in (2.1) using standard regression techniques with $\alpha$ replaced by its estimated value $\hat{\alpha}$. This is appropriate for the purposes of point estimation, but without taking into account the variability in $\hat{\alpha}$, estimated standard errors of $\hat{\beta}$ will tend to be underestimated, perhaps severely so. If the final analysis consists of comparing estimated $\beta$ values for a sample of individual subjects across, for instance, treatment groups, then reliable standard errors can dramatically enhance the power of such comparisons.

Let $(x_1, Y_1), \ldots, (x_n, Y_n)$ represent data observed from (2.1). If $\alpha$ were known, then the covariance matrix of $\hat{\beta}$ can be estimated using standard least-squares theory with $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, where $X$ is the matrix of partial derivatives of $f$ with respect to $\beta$ evaluated at $\hat{\beta}$. Because $\alpha$ is not known in practice, this covariance matrix will tend to produce variances that are too small.
The scenario of interest here is that an estimate \( \hat{\alpha} \) of \( \alpha \), as well as an estimated variance–covariance matrix for \( \hat{\alpha} \), is available from another source. For this situation, the standard least-squares approach is to determine the value of \( \beta \) minimizing \( \sum_{i=1}^{n} (y_i - f(x_i; \hat{\alpha}, \beta))^2 \). Defining

\[
H(\hat{\alpha}, Y) = \arg \min_{\beta} \sum_{i=1}^{n} (Y_i - f(x_i; \hat{\alpha}, \beta))^2,
\]

the estimator of \( \beta \) in practice is \( \hat{\beta} = H(\hat{\alpha}, Y) \) whose value therefore depends on realizations of two other random variables. If the estimator of \( \beta \) in the case of known \( \alpha \) is approximately unbiased for \( \beta \) and the function \( H(\hat{\alpha}, Y) \) is well approximated by a first-order Taylor series about \( \alpha \) and \( \alpha \) must be estimated, then

\[
E[\hat{\beta}] = E[H(\hat{\alpha}, Y)] \\
\approx E \left[ H(\alpha, Y) + \frac{\partial H(\alpha, Y)}{\partial \alpha} (\alpha - \hat{\alpha}) \right] \\
\approx \beta + E \left[ \frac{\partial H(\alpha, Y)}{\partial \alpha} \right] E[\alpha - \hat{\alpha}],
\]

showing that \( \hat{\beta} \) is approximately unbiased provided that \( \hat{\alpha} \) is approximately unbiased for \( \alpha \) and independent of \( Y \), and that the expectations exist.

However, the variability in the \( \hat{\beta} \) is amplified when an estimator of \( \alpha \) is used in place of its true value. If \( \hat{\alpha} \) is used in the estimation of \( \beta \), then the well-known variance decomposition formula can be used to relate \( \text{Var}(\hat{\beta}) \) to \( \text{Var}(\hat{\beta}|\hat{\alpha}) \):

\[
\text{Var}(\hat{\beta}) = \text{Var}(E[\hat{\beta}|\hat{\alpha}]) + E[\text{Var}(\hat{\beta}|\hat{\alpha})]. \tag{2.2}
\]

Note that \( \text{Var}(\hat{\beta}) - E[\text{Var}(\hat{\beta}|\hat{\alpha})] \) is a positive semidefinite matrix, indicating that the conventional approach based on \( \text{Var}(\hat{\beta}|\hat{\alpha}) \) underestimates the true variance–covariance matrix of \( \hat{\beta} \) on average.

3. ASYMPTOTIC RESULTS

In this section, the asymptotic distribution for the least-squares estimator of \( \beta \) in (2.1) is derived when the \( \alpha \) parameters are estimated from a separate data set. The asymptotic results derived here are related to previous work on pseudo maximum likelihood by Gong and Samaniego (1981) and on estimating equations by Carroll et al. (1995). Gong and Samaniego (1981) derive asymptotic normality results similar to our result below except that they consider maximum likelihood estimation with a single data source where a set of nuisance parameters is estimated using a method different from maximum likelihood. Carroll et al. (1995, pp. 267–269) consider the estimation of \( \beta \) based on a random sample of \( n \) individuals, from which estimators of \( \alpha \) are also computed.

We shall assume that the function \( f \) in (2.1) is differentiable with respect to the parameters. Let \( \alpha = (\alpha_1, \ldots, \alpha_p)' \) and \( \beta = (\beta_1, \ldots, \beta_q)' \). Denote by \( \alpha^* \) and \( \beta^* \) the ‘true’ values of the parameters. Suppose that \( \hat{\alpha} \) is a strongly consistent estimator of \( \alpha \) that is based on another data set of sample size \( n_1 \) and that

\[
\sqrt{n_1}(\hat{\alpha} - \alpha^*) \xrightarrow{\text{d}} N(0, \Psi),
\]

where \( \xrightarrow{\text{d}} \) represents convergence in distribution. We shall assume that the sample sizes \( n_1 \) and \( n \) go to infinity at the same rate with \( \frac{n}{n_1} \to c \), where \( 0 < c < \infty \). Furthermore, we assume that \( \hat{\alpha} \) is independent
of $\epsilon_1, \ldots, \epsilon_n$. Some necessary additional notation follows:

$$\Omega_1 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\partial f(x_i; \alpha^*, \beta^*)}{\partial \beta} \frac{\partial f(x_i; \alpha^*, \beta^*)}{\partial \beta'} \right]$$

and

$$\Omega_2 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\partial f(x_i; \alpha^*, \beta^*)}{\partial \alpha} \frac{\partial f(x_i; \alpha^*, \beta^*)}{\partial \beta'} \right],$$

where $\frac{\partial f(x_i; \alpha^*, \beta^*)}{\partial \beta}$ represents the column vector of length $q$ with the $j$th element given by $\frac{\partial f(x_i; \alpha^*, \beta^*)}{\partial \beta_j}$ evaluated at the point $\alpha = \alpha^*$ and $\beta = \beta^*$. (The derivative with respect to $\beta'$ is the transpose of this.) The matrices $\Omega_1$ and $\Omega_2$ have dimensions $q \times q$ and $p \times q$, respectively. We assume that the above limits exist and that $\Omega_1$ is nonsingular. We also assume appropriate regularity conditions ensuring that $\beta|$ is consistent for $\beta$ in the case that $\alpha$ is known [for additional details, see Chapter 12 of Seber and Wild (1989)]. The following result demonstrates the asymptotic normality of $\hat{\beta}$. The proof of the theorem, given in the supplementary material (www.biostatistics.oxfordjournals.org), follows the general approach taken by Amemiya (1983).

**Theorem 3.1** Assuming the appropriate regularity conditions, $\sqrt{n}(\hat{\beta} - \beta^*)$ is asymptotically normal with mean 0 and variance–covariance matrix

$$\sigma^2 \Omega_1^{-1} + c \Omega_1^{-1} \Omega_2 \Psi \Omega_2 \Omega_1^{-1}.$$

The square root of the diagonal of the asymptotic covariance matrix can be used to obtain standard errors for the rate parameters. In practice, this asymptotic covariance matrix must be estimated. In particular, the estimated covariance matrix of $\hat{\alpha}$ can be used in place of $\Psi$. Similarly, $\Omega_1$ and $\Omega_2$ can be estimated by taking the appropriate partial derivatives and evaluating them using the parameter estimates and averaging over all observations.

It is worth noting that the variance–covariance matrix of $\hat{\beta}$ based on the asymptotic formula above has two components just as in (2.2). The first component is ordinarily computed as the estimated variance–covariance matrix in any nonlinear least-squares setting—the ‘naive’ estimator of the standard error in which the variability of the estimated input function is neglected. The second component arises from the variability in estimating $\alpha$. This component depends on the factor $c = \lim_{n_1 \to \infty} n/n_1$ which can be estimated by the ratio of the actual sample sizes. Thus, if the input parameters are estimated with a large sample size $n_1$ relative to $n$, then the constant $c$ will be small and the variability of $\hat{\alpha}$ will play a diminished role in affecting the variability of $\hat{\beta}$.

Theorem 3.1 can be generalized to handle situations where the i.i.d. assumption may not hold (e.g. nonconstant variance or longitudinal studies) or where the input data set is not independent of the primary data set. Consider the following (possibly nonlinear) model for the input data set

$$z_i = h(w_i; \alpha) + \xi_i, \quad i = 1, \ldots, n_1,$$

where $\alpha$ represents the input parameters as before, the $\xi_i$ are mean zero errors, and the function $h$ is twice differentiable with respect to $\alpha$. Stack the input errors $\xi = (\xi_1, \ldots, \xi_{n_1})'$ on the errors $\epsilon$ from (2.1) and let the covariance matrix for the joint distribution of $(\xi', \epsilon')'$ be given by

$$\text{cov} \left( \begin{array}{c} \epsilon \\ \xi \end{array} \right) = \left( \begin{array}{cc} \sigma^2 V_{1,n} & V_{12,n,n_1} \\ V_{12,n,n_1}' & \sigma_1^2 V_{2,n_1} \end{array} \right).$$
Let \( G_n = [\partial f/\partial \beta] \) denote the \( n \times q \) matrix of partial derivatives of the nonlinear regression function \( f \) evaluated at \( x_1, \ldots, x_n \). Similarly, let \( H_n = [\partial h/\partial \alpha] \) denote the \( n_1 \times p \) matrix of partial derivatives of \( h \) with respect to \( \alpha \) evaluated at \( w_1, \ldots, w_{n_1} \). Implementing the asymptotic results of Amemiya (1983, p. 354) for nonlinear regression with correlated errors and incorporating the covariance between the input data and primary data, it follows that as \( n_1 \) and \( n \) go to infinity, \( \sqrt{n}(\hat{\beta} - \beta) \) is asymptotically normal with mean 0 and covariance matrix

\[
\Omega_{1}^{-1}\left[ \sigma^2 G'V_1G + c\sigma^2 \Omega_2'(H'H)^{-1}H'V_2H(H'H)^{-1}\Omega_2 \\
- G'V_{12}H(H'H)^{-1}\Omega_2 - \Omega_2'(H'H)^{-1}H'V_{12}G \right] \Omega_1^{-1},
\]

where

\[
G'V_1G = \lim_{n \to \infty} \frac{1}{n} G_n^t V_{1,n} G_n,
\]

\[
(H'H)^{-1}H'V_2H(H'H)^{-1} = \lim_{n_1 \to \infty} n_1 (H_{n_1} H_{n_1})^{-1} H_{n_1}^t V_{2,n_1} H_{n_1} (H_{n_1} H_{n_1})^{-1},
\]

\[
G'V_{12}H(H'H)^{-1} = \lim_{n,n_1 \to \infty} G_n' V_{12,n,n_1} H_{n_1} (H_{n_1} H_{n_1})^{-1}.
\]

The proof of (3.1) is outlined in the supplementary material (www.biostatistics.oxfordjournals.org).

4. Bootstrap Standard Errors

An alternative to the asymptotic approach from the previous section is to use the bootstrap (Efron and Tibshirani, 1993) for estimating the standard errors. In some applications, the asymptotic results may be difficult to obtain if the computation of the partial derivative matrices needed for the asymptotic covariance matrix is complex. Also, standard errors based on asymptotic computations may not provide good estimates for data sets of small to moderate size. The implementation of the bootstrap approach must incorporate the variability of the estimation of the input parameters into the calculation of standard errors for the rate parameters.

If the points \( x_i \) are fixed by design, then it is more appropriate to compute bootstrap samples of the \( Y \) data using the residuals from the fit rather than resampling the raw data (see Efron and Tibshirani, 1993, Section 9.5). The proper resampling scheme to generate resampled input parameter estimates depends on the nature of the input data. If raw input data are not available, a parametric bootstrap can be applied, provided both an estimate of \( \alpha \) and its estimated variance–covariance matrix are available. The choice of parametric bootstrap algorithm should depend upon the nature of the input data, but if it can be reasonably assumed that \( \hat{\alpha} \) is normally distributed with mean \( \alpha^* \) and variance–covariance matrix \( \frac{1}{n_1} \Psi \), then each resampled set of estimated parameters could be generated from an \( N(\hat{\alpha}, \frac{1}{n_1} \Psi) \) distribution (replacing \( \Psi \) with its estimate where appropriate).

There are a number of variants of the general bootstrap algorithm that would be appropriate for this situation, depending on whether original data or residuals are to be resampled, or if a parametric bootstrap is to be utilized. For instance, if data are thought to be serially correlated, a moving blocks bootstrap algorithm (Künsch, 1989; Liu and Singh, 1992) may be appropriate. This would involve resampling blocks of adjacent residuals rather than sampling from the residuals without regard to their order as in the usual bootstrap algorithm. This helps to preserve the correlation structure inherent in the original data.
For any selected set of bootstrap schemes, the general algorithm is:

**Bootstrap Algorithm 1 (one-stage bootstrap)**

1. For each $k = 1, \ldots, B$,
   
   (a) obtain a bootstrap sample of the $a$ data and calculate the resulting $\hat{a}_k$ value;  
   (b) obtain a bootstrap sample of the $Y$ data and compute $\hat{\beta}_k$ from these data by plugging in the value of $\hat{a}_k$ in the expression for $f(x_i; a, \beta)$.

2. The bootstrap standard error is the standard deviation of components of $\hat{\beta}_1, \ldots, \hat{\beta}_B$.

An alternative approach would be to conduct a two-stage (nested) bootstrap. This would allow for the estimation of each component of the variance decomposition formula (2.2), which may be useful for purposes of design as well as for ascertaining the individual sources of variability of the parameter estimate. In the case that the resampling for the $Y$ data is based on residuals from a fitted curve involving $\hat{a}$, this will also allow for a richer set of residuals for the $Y$ data (since a new set of residuals would be computed for each bootstrapped value of $\hat{a}$). This algorithm is given by:

**Bootstrap Algorithm 2 (two-stage bootstrap)**

1. The outer loop: For each $k = 1, \ldots, B_1$,
   
   (a) obtain a bootstrap sample of the $a$ data and calculate the resulting $\hat{a}_k$ value.
   
   (b) The inner loop: For each $\ell = 1, \ldots, B_2$,
   
      (i) using the computed value of $\hat{a}_k$, obtain a bootstrap sample of the $Y$ data and compute $\hat{\beta}_{\ell(k)}$.
   
   (c) Compute $\hat{\mu}_k$, the average of the $\hat{\beta}_{\ell(k)}$ values.
   
   (d) Compute $\hat{\Gamma}_k$, the sample variance–covariance matrix of the $\hat{\beta}_{\ell(k)}$ values.

2. Estimate $\text{Var}(E[\hat{\beta}|\hat{a}])$ by computing the sample variance–covariance matrix of the $\hat{\mu}_k$ values.

3. Estimate $E[\text{Var}(\hat{\beta}|\hat{a})]$ by computing the matrix average of the $\hat{\Gamma}_k$s.

4. The bootstrap estimate of the (unconditional) variance–covariance matrix of $\hat{\beta}$ is the sum of these two components.

Note that in the two-stage algorithm, each $\hat{\mu}_k$ is an estimate of the corresponding $E[\hat{\beta}|\hat{a}_k]$ and $\hat{\Gamma}_k$ is an estimate of $\text{Var}(\hat{\beta}|\hat{a}_k)$.

5. **Applications**

We return to Examples 1 and 2 from Section 1 to illustrate the computation of standard errors when input parameters are estimated via an external data set. Simulations were performed to examine the behavior of estimated standard errors using the asymptotic formula as given in the theorem in Section 3, both bootstrap approaches as described in Section 4, and the ‘naive’ approach (in which the variability in estimating the input parameters is neglected). We begin with the relatively simple model of zinc metabolism in urine as described in Example 2 of Section 1.
5.1 Zinc metabolism

To provide a simple illustration of the methods described in this paper, data on zinc concentration in urine (1.2) and in plasma (1.3) were simulated using parameter values based on the results reported in Foster et al. (1979). As is common in these situations, we assumed a multiplicative error and performed the fitting using a log-transformed model:

\[ Y_i = \beta + \log \left( \int_0^{t_i} g(s; \alpha) \, ds \right) + \epsilon_i. \]

In this simple case, the estimate of \( \beta \) is just the arithmetic mean of the values \( Y_i - \log \left( \int_0^{t_i} g(s; \alpha) \, ds \right) \).

The input parameters in (1.3) were estimated using nonlinear least squares. Gaussian errors were generated throughout, and all simulations were run using Matlab software. A plasma input function was estimated based on averaged data from 17 subjects who ingested a zinc isotope. Concentration of the isotope in other compartments (liver, urine, etc.) was measured for the same 17 subjects at several time points over 5 days and averaged across subjects. For the simulation, parameter values were chosen to conform with the data reported in Aamodt et al. (1979). In particular, input parameter values for the plasma were set at \( \alpha = (0.790, 176, 0.175, 73.4, 0.022, 5.87, 0.013, 0.053)' \), and the value of the rate parameter \( \beta \) was set to \(-1.46\). The standard deviation for the plasma data was set to 0.0015 and that for the urine data was set to 0.02.

The sampling distribution of \( \hat{\beta} \) was simulated in two cases and is shown in Figure 1. The dashed curve in Figure 1 shows the distribution of the estimated rate parameter in the case when the input parameters

![Fig. 1. Distributions of simulated values of \( \hat{\beta} \) for the case in which the input function is known (dashed curve) and estimated (solid curve). The vertical line indicates the true value of \( \beta \).](https://academic.oup.com/biostatistics/article-abstract/7/1/115/243111)
are known without error while the solid curve shows the distribution of \( \hat{\beta} \) when the input parameters are estimated. In both cases, \( \hat{\beta} \) is approximately unbiased, but its variability is much greater when the input function is estimated.

Next, simulations were run in order to compare the standard errors of the naive, asymptotic, and bootstrap approaches. Two separate simulations were run for the zinc model differing only in the number of time points at which plasma samples were to be taken. In the first run, the time points were set to the original time points as in Aamodt et al. (1979); in the second (to allow for better comparison of the asymptotic expression for the standard error), the plasma sampling rate was increased.

The first simulation run used the original time points from the Foster et al. paper (for both the plasma and the urine data), giving \( n_1 = 16 \) time points for the plasma data and \( n = 12 \) urine samples over the course of about 5 days. A total of 1000 simulated data sets were generated. For the one-stage bootstrap algorithm, we used \( B = 200 \). For the two-stage bootstrap, we set \( B_1 = B_2 = 200 \). The results of this simulation are displayed in Figure 2. Corresponding summary statistics are provided in Table 1. The five smoothed histograms in Figure 2 were generated from the 1000 simulated data sets. Note that these are all on the standard deviation scale. The solid curve represents the naive standard error estimates (neglecting the variability of the input data). The short-length dashed curve represents the distribution of the average of the \( \hat{\Gamma}_{i}^{*} \) values as described in Section 4, the bootstrap estimate of \( E[\text{Var}(\hat{\beta}|\hat{\alpha})] \). The dotted curve shows the distribution of the estimates of \( \text{Var}(E[\hat{\beta}|\hat{\alpha}]) \), the sample variance–covariance matrix of the \( \hat{\mu}_{i}^{*} \) values. Combining these two gives the medium-length dashed curve representing the total bootstrap estimate of the standard errors from the two-stage algorithm. Finally, the long-length dashed curve shows the

![Figure 2](https://academic.oup.com/biostatistics/article-abstract/7/1/115/243111)

**Fig. 2.** Smoothed histograms of the results from the simulation study based on the original data as reported in Section 5.1. The vertical line represents the ‘true’ standard error for this situation, as computed by a separate simulation study of 10 000 simulated data sets.
Table 1. Summary statistics of simulation results presented in Figure 2

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive standard error</td>
<td>0.0082</td>
<td>0.0077</td>
<td>0.0027</td>
</tr>
<tr>
<td>( \sqrt{\text{Var}(E(\hat{\beta}</td>
<td>\hat{\alpha}))} )</td>
<td>0.0220</td>
<td>0.0217</td>
</tr>
<tr>
<td>( \sqrt{E(\text{Var}(\hat{\beta}</td>
<td>\hat{\alpha}))} )</td>
<td>0.0090</td>
<td>0.0085</td>
</tr>
<tr>
<td>Two-stage standard error</td>
<td>0.0238</td>
<td>0.0236</td>
<td>0.0052</td>
</tr>
<tr>
<td>One-stage standard error</td>
<td>0.0236</td>
<td>0.0233</td>
<td>0.0051</td>
</tr>
</tbody>
</table>

estimated distribution of standard errors computed using the one-stage bootstrap algorithm. The vertical line represents the ‘true’ standard error, computed as the standard deviation of the \( \hat{\beta} \) values from a separate simulation of 10,000 data sets.

It is clear in Figure 2 that the naive approach to standard error estimation badly underestimates the truth. The two-stage bootstrap algorithm demonstrates that most of the overall variability of \( \hat{\beta} \) is due to the variability in the estimates of \( \hat{\alpha} \), so the \( \text{Var}(E(\hat{\beta}|\hat{\alpha})) \) term tends to dominate the overall estimate of the standard error. There is little difference between the one-stage and the two-stage bootstrap standard errors in this example. Figure 2 shows that both bootstrap approaches show some negative bias. This may be due in part to the small sample size, in particular, the necessity to estimate eight plasma parameters from only 16 data points using nonlinear regression.

The asymptotic formula given in Section 3 when applied to the simulation data using original time points gave very large and unreliable standard error estimates for about 1–2% of the simulated data sets. This is due to numerical instability of estimating the derivatives for the plasma function based on a small sample size, as the denominator of the estimated derivatives can take values near 0. If these values are excluded, the distribution of the asymptotic standard error estimates closely resembles the bootstrap results.

In order to compare the performance of the asymptotic formula with the bootstrap approach, a second simulation study was conducted using \( n = 12 \) urine samples [again, as in the original Aamodt et al. (1979) paper] but \( n_1 = 80 \) plasma samples (computed by equally subdividing each interval in the original plasma data). In this new setting, 1000 simulated data sets were generated and standard error estimates were computed using both bootstrap algorithms as well as the asymptotic formulas. The results of this simulation are shown in Figure 3 and summary statistics are given in Table 2.

Figure 3 displays the same information as Figure 2, but for the new, more frequently sampled plasma data. In this situation, the bootstrap standard error is more balanced between its two components and the two components taken together provide a nearly unbiased estimate of the ‘true’ standard error. The one-stage bootstrap algorithm is close but exhibits some negative bias. Again, as may be expected, the naive estimator conforms closely to the \( E[\text{Var}(\hat{\beta}|\hat{\alpha})] \) term.

For the setting with increased plasma sampling (\( n_1 = 80 \)), the distribution of the bootstrap estimates nearly coincides with that of the asymptotic estimates, and both are approximately unbiased for the true standard error.

5.2 PET imaging

In the PET imaging application described in Example 1 of Section 1, the simplest model in common use is a two-compartment model [(one compartment for the plasma and one for a brain region of interest (ROI)]. For model (1.1), the \( g(t; \beta) \) function is a simple exponential, so the concentration of the ligand in a given brain location is given by \( C(t) = k_1 e^{-k_2 t} \otimes C_p(t; \alpha) \) for unknown rate parameters \( k_1 \) and \( k_2 \). The plasma
Estimation in regression models with externally estimated parameters

Fig. 3. Smoothed histograms of the results from the simulation study with increased plasma sampling as reported in Section 5.1. The vertical line represents the 'true' standard error for this situation, as computed by a separate simulation study of 10,000 simulated data sets.

Table 2. Summary statistics of simulation results presented in Figure 3

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive standard error</td>
<td>0.0062</td>
<td>0.0061</td>
<td>0.00141</td>
</tr>
<tr>
<td>$\sqrt{\text{Var}(E(\hat{\beta}</td>
<td>\hat{\alpha}))}$</td>
<td>0.0084</td>
<td>0.0084</td>
</tr>
<tr>
<td>$\sqrt{E(\text{Var}(\hat{\beta}</td>
<td>\hat{\alpha}))}$</td>
<td>0.0064</td>
<td>0.0063</td>
</tr>
<tr>
<td>Two-stage standard error</td>
<td>0.0106</td>
<td>0.0106</td>
<td>0.00106</td>
</tr>
<tr>
<td>One-stage standard error</td>
<td>0.0103</td>
<td>0.0103</td>
<td>0.00112</td>
</tr>
</tbody>
</table>

concentration function is commonly modeled by a piecewise expression:

$$C_p(t) = a_0 t I(0 \leq t < t^*) + \sum_{i=1}^{3} a_i e^{-\lambda_i t} I(t \geq t^*).$$  \hspace{1cm} (5.1)

Plasma data (consisting of measured concentration based on total radioactivity counts corrected for rate of decay and metabolism) are taken over time from arterial samples while the subject is being scanned:

$$X_i = C_p(s_i) + \eta_i, \quad i = 1, \ldots, n_1.$$  

The parameters involved in the expression of $C_p(t)$ in (5.1) are estimated using nonlinear least squares.
The convolution defining $C(t)$ can be worked out analytically as

$$C(t) = \begin{cases} 
  \alpha_0 \left( k_1 / k_2 \right)^2 \left( e^{-k_2 t} + k_2 t - 1 \right), & t < t^*, \\
  \alpha_0 \left( k_1 / k_2 \right)^2 \left( e^{-k_2 (t-t^*)} (k_2 t^* - 1) + e^{-k_2 t} \right) \\
  + k_1 \sum_{i=1}^3 \alpha_i (k_2 - \lambda_i)^{-1} \left( e^{-\lambda_i t} - e^{-\lambda_i t^*} e^{-k_2 (t-t^*)} \right), & t \geq t^*. 
\end{cases}$$

Through the PET imaging modality, concentration is measured for each of several time-ordered ‘frames’ at each location on a 3D grid. In some applications, aggregate measures are computed for several anatomically defined ROI by combining all data from voxels contained within the ROI. Assuming the two-compartment model, the brain data are given by

$$Y_i = C(t_i) + \epsilon_i, \quad i = 1, \ldots, n.$$ 

One outcome measure of interest is the ‘total volume of distribution,’ defined as $V_T = K_1 / k_2$.

Simulation parameters are set to mimic real data from PET imaging experiments as in Parsey et al. (2000). The ‘true’ plasma and brain regression functions are displayed in Figure 4.

This PET example is more complicated than the zinc example and requires a number of modifications to the bootstrap procedures. First, in practice, the plasma data points are measured together with an estimate of their standard errors. These errors are used to determine weights in a weighted least-squares fitting algorithm that substantially improves the quality of the fit and the stability of estimates. The weights, taken from an actual study, are used to generate the plasma data in the simulation. Additional improvement of the bootstrap procedure was achieved by centering the residuals at zero and dividing the residuals by $\sqrt{1 - h_{ii}}$, where $h_{ii}$ represents the nonlinear regression analogue to the ‘hat’ matrix; see Davison and Hinkley (1997, Chapter 6) for a discussion of this procedure in the linear regression case.

Figure 5 and Table 3 summarize the results of the simulation. As in the zinc example, the standard error of the naive estimator badly underestimates the true standard error. On the other hand, both the bootstrap and asymptotic results are reasonably close to the true standard error.

For the purposes of illustration, these methods were applied to data from the ventral striatum of a PET imaging subject. For this study with an estimated $V_T$ of 43.8, the bootstrap estimate of the standard error was 11.1 and the asymptotic estimate was 12.6. The naive estimates of the standard error neglecting the plasma effect are 9.01 for the bootstrap and 8.60 for the asymptotic estimator. This example illustrates

![Fig. 4. ‘True’ plasma (left panel) and ‘true’ brain (right panel) concentrations, taken from real PET data.](https://academic.oup.com/biostatistics/article-abstract/7/1/115/243111)
that in current PET imaging practice, the naively estimated standard errors for parameter estimates such as $V_T$ are underestimating the true standard error.

6. DISCUSSION

This paper proposes two general methods for the estimates of standard errors in regression models defined in terms of input parameters that are estimated separately. The usual method of estimation neglects the variability in estimating the input parameters. In contrast, the methods described here take the variability of the estimated input parameters into account and consequently yield much better estimates of the true variability in the estimated model parameters.

The asymptotic formulas gave reasonably good standard error estimates even for relatively small sample sizes. Computationally, the asymptotic standard errors can be computed very fast. The quick computation makes the asymptotic standard error approach very attractive in applications like PET imaging (Section 5.2) that involve model fittings at hundreds of thousands of voxels. However, a drawback of the asymptotic approach is that the coding routine can require painstaking calculations and is prone to error.

On the other hand, the bootstrap approaches tend to work quite well for a variety of settings. Coding the bootstrap is straightforward, but, of course, they require considerably more computational expense than the asymptotic approach.

Table 3. Summary statistics of simulation results presented in Figure 5

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive standard error</td>
<td>3.2</td>
<td>3.0</td>
<td>0.95</td>
</tr>
<tr>
<td>Bootstrap standard error</td>
<td>4.3</td>
<td>4.2</td>
<td>1.22</td>
</tr>
<tr>
<td>Asymptotic standard error</td>
<td>5.1</td>
<td>4.9</td>
<td>1.24</td>
</tr>
</tbody>
</table>

Fig. 5. Smoothed histograms of one-stage bootstrap and asymptotic standard errors for the PET simulation study as described in Section 5.2. The vertical line represents the ‘true’ standard error for this situation, as computed by a separate simulation study of 10,000 simulated data sets.
ACKNOWLEDGMENTS

The authors are grateful to the editor, associate editor, and referee for helpful and insightful comments that have led to a substantial improvement of this paper.

REFERENCES


[Received May 19, 2004; first revision October 20, 2004; second revision April 8, 2005; third revision June 14, 2005; fourth revision June 22, 2005; accepted for publication July 13, 2005]