.

The Nonlinear Mixed Effects Model with a Smooth Random Effects Density

Marie Davidian and A. Ronald Gallant

Institution of Statistics Mimeograph Series No. 2206

November 1991
Revised April 1992
Revised November 1992
Revised December 1992

The Nonlinear Mixed Effects Model with a Smooth Random Effects Density [1]

Marie Davidian
North Carolina State University

A. Ronald Gallant
North Carolina State University

November 1991
Revised April 1992
Revised November 1992
Revised December 1992

[1]Corresponding author: Marie Davidian, Department of Statistics, North Carolina State University, Campus Box 8203, Raleigh NC 27695-8203 USA, 1-919-515-2532.

Summary

The fixed parameters of the nonlinear mixed effects model and the density of the random effects are estimated jointly by maximum likelihood. The density of the random effects is assumed to be smooth but is otherwise unrestricted. The method uses a series expansion that follows from the smoothness assumption to represent the density and quadrature to compute the likelihood. Standard algorithms are used for optimization. Empirical Bayes estimates of random coefficients are obtained by computing posterior modes. The method is applied to data from pharmacokinetics, and properties of the method are investigated by application to simulated data.

Some key words: Maximum likelihood; nonlinear mixed effects models; nonparametric; pharmacokinetics.

1. Introduction

Data consisting of repeated measurements on each individual in a sample from a population arise in pharmacokinetics, pharmacodynamics, growth studies in agriculture and medicine, labor economics, and other applications. In these applications, standard models for the observations on a given individual are nonlinear in their parameters, as in the case of poly-exponential pharmacokinetic models characterizing drug plasma concentration over time. Although the functional form of the model remains the same for all individuals the parameter values vary from individual to individual. Often, the objective of an analysis is to characterize the population of these parameter values. Determination of parameter values for a given individual may also be of interest, as in setting individual dosage regimens in pharmacokinetics.

The nonlinear mixed effects model is a standard model for this situation. Intra-individual variation is accounted for by the nonlinear model and a distribution for measurement error. Inter-individual variation is accounted for by the assumption of a separate, random parameter for each individual. Since this variation may partially depend on individual attributes, the random parameters are often taken to be a function of these attributes, fixed effects, and random effects. The distribution of the random parameters describes the population and is of primary interest.

Estimates of the fixed effects and the first and second moments of the random effects are often used to describe the distribution. Some methods taking this approach use individual regression parameter estimates as building blocks (Steimer, Mallet, Golmard & Boisvieux, 1984; Beal & Sheiner, 1985; Racine-Poon, 1985; Davidian & Giltinan, 1993). Beal & Sheiner (1982), Lindstrom & Bates (1990), and Vonesh & Carter (1992) suggest methods based on linearization of the nonlinear mixed effects model in the random effects. Other methods make distributional assumptions. For instance, Gelfand, Hills, Racine-Poon & Smith (1990) describe a Gibbs sampling algorithm to generate samples of the random parameters based on a full hierarchical Bayesian specification.

The estimation methods discussed above make a parametric assumption regarding the distribution of the random effects, estimate only first and second moments, or require more observations per individual than are often available. Features such as multimodality or asymmetry will not be detected under standard parametric assumptions or from first and second moments. A parametric specification that can represent these characteristics without

a priori knowledge of them is difficult. Therefore, it is important to have methods that can estimate the entire distribution nonparametrically, even from sparse individual data (Mallet, 1986).

Mallet (1986) proposes a nonparametric maximum likelihood approach for estimating the distribution of the random effects. The distribution is unrestricted. A side effect is that the estimate of the distribution is discrete. The method is appealing since very little is assumed about the form of the population distribution. However, as pointed out by Mallet, Mentré, Steimer & Lokiec (1988), no estimates of the precision of the estimated population characteristics or distribution are available, and a separate maximization is required to estimate the fixed effects.

By sacrificing some generality in favor of a smoothness assumption, the density of the random effects can be estimated jointly with the fixed effects by maximum likelihood, and inference is possible. The method uses a series expansion that follows from smoothness assumptions and that is due to Gallant & Nychka (1987) to represent the density and uses quadrature to compute the likelihood. There is no reliance on linearizations or other approximations to the likelihood. Standard algorithms can be used for optimization. Empirical Bayes estimates of random parameters are obtained by computing posterior modes.

In Section 2, we specify a general nonlinear mixed effects model which makes no parametric assumption about the form of the random effects distribution. In Section 3, we describe the proposed estimation procedure. In Section 4, we illustrate the method by application to data from a clinical study of neonatal population pharmacokinetics of phenobarbital. In Section 5, we focus on the ability of the method to accurately track features of a population such as bimodality by application to simulated data.

2. Model and notation

Observed responses y_{ij}, $1 \leq j \leq J_i$, on individual i, $1 \leq i \leq n$, at settings x_{ij} of a vector of covariates are assumed to follow the intra-individual nonlinear regression model $y_{ij} = f(x_{ij}, \beta_i) + e_{ij}$. The J_i are bounded by some $J < \infty$. The total number of observations is $N = \sum_{i=1}^{n} J_i$.

The function f is known up to the unknown parameter β_i and the joint density of the errors e_{ij}, $p_e(e_{i1}, \ldots, e_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, \sigma, \beta_i)$ is known up to the unknown parameter (σ, β_i); σ and β_i are vectors of dimension p_σ and p_β respectively. This specification is flexible enough

to accommodate general intra-individual heterogeneity, in particular heteroskedasticity and correlation.

The p_β-dimensional parameter β_i is random and follows the inter-individual nonlinear regression model $\beta_i = g(w_i, \gamma, z_i)$, where the function g is known, w_i is a vector of individual attributes, γ is a p_γ-dimensional vector of unknown fixed effects, and z_i is an M-dimensional vector of inter-individual random effects with density h.

We assume that h belongs to a smooth class $\mathcal{H}$. The primary objective is estimation of and inference regarding the random effects density h and the fixed parameters $\tau = (\gamma, \sigma) \in \mathcal{R}^{p_\tau}$, $p_\tau = p_\gamma + p_\sigma$. Once τ and h are determined, the individual parameters β_i can be estimated by empirical Bayes.

To summarize, the quantities of interest are τ^o and h^o that denote the true values of the fixed parameters and random effects density respectively, β_i^o, $1 \le i \le n$, that denote the realized values of the random parameters, and z_i^o, $1 \le i \le n$, that denote the realized values of the random effects.

One approach to estimation of h^o and τ^o is by maximum likelihood. The log likelihood is

$$\ell(\tau, h) = \sum_{i=1}^{n} \log \int p(y_{i1}, \ldots, y_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, w_i, \tau, z)\, h(z)\, dz.$$

where $p(y_{i1}, \ldots, y_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, w_i, \tau, z_i)$ is the joint density of the observations on individual i. This density may be obtained by substituting $e_{ij} = y_{ij} - f\{x_{ij}, g(w_i, \gamma, z_i)\}$ into $p_e\{e_{i1}, \ldots, e_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, \sigma, g(w_i, \gamma, z_i)\}$ because the Jacobian of $(e_{i1}, \ldots, e_{iJ_i})$ with respect to $(y_{i1}, \ldots, y_{iJ_i})$ is the identity matrix of order J_i. We estimate h^o nonparametrically, simultaneously with τ^o, by maximizing the likelihood over h in $\mathcal{H}$ and τ in $\mathcal{R}^{p_\tau}$. The procedure is described in the next section.

3. Estimation method

The maximum likelihood estimate of (τ^o, h^o), $\tau^o = (\gamma^o, \sigma^o)$, may be computed as $(\hat{\tau}, \hat{h}) = \operatorname{argmin}_{\tau \in \Re^{p_\tau}, h \in \mathcal{H}} s_n(\tau, h)$, where $s_n(\tau, h) = (-1/N)\ell(\tau, h)$. The empirical Bayes estimates of the effects z_i^o, $1 \le i \le n$, are computed as

$$\hat{z}_i = \arg\max_{z \in \Re^M} p(y_{i1}, \ldots, y_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, w_i, \hat{\tau}, z)\, \hat{h}(z).$$

From these, the empirical Bayes estimates of the random parameters $\beta_i^o, 1 \le i \le n$, are obtained by evaluating $\hat{\beta}_i = g(w_i, \hat{\gamma}, \hat{z}_i)$. These computations require a characterization of

$\mathcal{H}$, a convenient representation of $h \in \mathcal{H}$, and an accurate and efficient means to compute integrals of the form $\int \psi(z)h(z)\,dz$. These are the topics of this section.

We follow Gallant & Nychka (1987) who propose a nonparametric estimator of h^o specifically designed to be used with maximum likelihood computations. Their class $\mathcal{H}$ of smooth densities is described in terms of a weighted Sobolev norm:

Sobolev norm. Denote a partial derivative of a function $f(z)$ on $\Re^M$ by

$$D^\lambda f(z) = \Big(\frac{\partial^{\lambda_1}}{\partial z_1^{\lambda_1}}\Big) \cdots \Big(\frac{\partial^{\lambda_M}}{\partial u_M^{\lambda_M}}\Big) f(z),$$

where $\lambda = (\lambda_1, \ldots, \lambda_M)$. Letting $|\lambda| = \sum_{k=1}^M |\lambda_k|$, the Sobolev norm of f with respect to a weight function $\mu(z)$ is

$$\begin{aligned} \|f\|_{m,p,\mu} &= \Big\{ \sum_{|\lambda| \le m} \int |D^\lambda f(z)|^p \mu(z)\,dz \Big\}^{1/p} \qquad 1 \le p < \infty \\ \|f\|_{m,\infty,\mu} &= \max_{|\lambda| \le m} \sup_{z \in \Re^k} |D^\lambda f(z)| \mu(z). \end{aligned}$$

The class $\mathcal{H}$, which is assumed to contain h^o, is defined as follows:

Parameter space $\mathcal{H}$. For some integer $m_0 > M/2$, some bound $\mathcal{B}_0$, some small $\epsilon_0 > 0$, some $\delta_0 > M/2$, and some strictly positive density function h_0 with $\|h_0\|_{m_0,2,\mu_0} < \mathcal{B}_0$, let $\mathcal{H}$ consist of those density functions h that have the form $h(z) = f^2(z) + \epsilon_0 h_0(z)$, with $\|f\|_{m_0,2,\mu_0} < \mathcal{B}_0$, where $\mu_0(z) = (1 + z'z)^{\delta_0}$.

In the definition of $\mathcal{H}$, $\mathcal{B}_0$ is an upper bound that both imposes a smoothness restriction on members of $\mathcal{H}$ and bounds the tails of densities in $\mathcal{H}$ from above. The fattest tails permitted by this bound are t-like with $h(z) \propto (1 + z'z)^{-\delta_0 - \eta}$ for some small $\eta > 0$. The smoothness restriction rules out kinks, jumps, and oscillatory behavior. It does not rule out skewed, leptokurtic, platykurtic, or multi-modal densities. The highest reasonable value for m_0 that one can assume for h^o determines the number $m_0 - M/2$ of derivatives of h^o that are estimated consistently.

The density h_0 is a lower bound that is imposed both to impose positivity and to bound the tails from below. In theory, one would choose $h_0(z) \propto \exp\{-(z'z)^{1+\eta}\}$ for some small $\eta > 0$ to allow $\mathcal{H}$ to contain densities with tails that are thinner than the normal density. In practice, simply add $\hat{\epsilon}$ to the integral $\int p(y_{i1}, \ldots, y_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, w_i, \tau, z)\, f^2(z)\,dz$ that is near the smallest value for which $\log \hat{\epsilon}$ can be computed without error; for example, $\hat{\epsilon} = 1 \times 10^{-300}$ on a machine with Institute of Electrical and Electronics Engineers double precision floats.

Representation of h. Writing a monomial as $z^\lambda = z_1^{\lambda_1} \cdots z_M^{\lambda_M}$, a density from $\mathcal{H}$ has the representation

$$h(z) = \Big\{ \sum_{|\lambda|<\infty} a_\lambda (R^{-1}z)^\lambda \Big\}^2 n_M(z|0, RR'),$$

where the term $\epsilon_0 h_0(z)$ is omitted as we do in applications, $n_M(\cdot|\mu, \Sigma)$ denotes the multivariate normal density, and R is an upper triangular matrix. Equality is in the sense of the norm $\|\cdot\|_{m_o - M/2, \infty, \mu}$, where $\mu(z) = (1 + z'z)^\delta$ for some $M/2 < \delta < \delta_0$.

Write a truncated expansion as $\{P_K(R^{-1}z)\}^2 n_M(z|0, RR')$, where $P_K(z) = \sum_{|\lambda|<K} a_\lambda z^\lambda$ denotes a polynomial of degree K on $\Re^M$. The truncated expansion will be a density if the coefficients $\{a_\lambda : 0 \leq |\lambda| \leq K\}$ are chosen so that $\int \{P_K(z)\}^2 n_M(z|0, I)\, dz = 1$. Equivalently, put $a_0 \equiv 1$ and write the truncated expansion as

$$h_K(z) = \frac{\{P_K(R^{-1}z)\}^2 n_M(z|0, RR')}{\int \{P_K(u)\}^2 n_M(u|0, I)\, du}.$$

The denominator $\int \{P_K(u)\}^2 n_M(u|0, I)\, du$ is a sum of products of the moments of the univariate standard normal distribution and is easily computed. Let $\theta_{(1)}$ be a vector whose elements are the coefficients $\{a_\lambda : 0 \leq |\lambda| \leq K\}$ arranged in some order, let $\theta_{(2)} = (r_{11}, r_{12}, r_{22}, r_{13}, r_{23}, r_{33}, \ldots, r_{MM})$, let $\theta = (\theta_{(1)}, \theta_{(2)})$, and let p_θ denote the dimension of the vector θ, which is determined solely by the degree K of P_K.

Let $\hat{K}_n$ represent some rule for choosing a truncation point in a sample of size n. For example, $\hat{K}_n$ might be a deterministic rule such as $\hat{K}_n = n^\alpha$ for some $0 < \alpha < 1$, or $\hat{K}_n$ might be an adaptive rule. Examples of adaptive rules are upward significance testing (Eastwood, 1991) and model selection according to the Schwarz criterion (Potscher, 1989). When such a rule has been specified, the $\hat{h}$ component of the maximum likelihood estimator $(\hat{\tau}, \hat{h}) = \text{argmin}_{\tau \in \Re^{p_\tau}, h \in \mathcal{H}} s_n(\tau, h_{\hat{K}_n})$, is a nonparametric estimator of h^o. If $\lim_{n\to\infty} \hat{K}_n = \infty$ almost surely then $\lim_{n\to\infty} ||\hat{\tau} - \tau^o|| = 0$ and $\lim_{n\to\infty} ||\hat{h} - h^o||_{m_o - M/2, \infty, \mu} = 0$ almost surely (Gallant & Nychka, 1987). The consistency norm $\|\cdot\|_{m_o - M/2, \infty, \mu}$ is a strong norm. Convergence with respect to this norm implies that the derivatives of h^o, moments of h^o, and other functionals such as $\sigma(h) = \int \psi(z) h(z)\, dz$ or $\sigma(h) = \max_z h$ are estimated consistently.

If h^o satisfies $\int z\, h^o(z)\, dz = 0$ then this constraint may be imposed on $\hat{h}$ without altering the consistency result. For $K > 0$, the off-diagonal elements of R can be constrained to be zero which attenuates estimated correlations but does not affect the consistency result.

A direct consequence of the series representation of h^o and the fact that adaptive rules are permitted is that the bound $\mathcal{B}_0$ that appears in the definition does not need to be imposed

on the estimate. This implies that K is the sole tuning parameter of the nonparametric estimator. Gallant & Nychka (1987) termed this estimator seminonparametric to suggest that the method lies midway between parametric and nonparametric methods: standard parametric algorithms are used yet the method has nonparametric properties. SNP is the acronym.

The optimization problem $(\hat{\tau}, \hat{h}) = \text{argmin}_{\tau \in \Re^{p_\tau}, h \in \mathcal{H}} s_n(\tau, h_{\hat{K}_n})$ is exactly the same as occurs in standard, finite dimensional maximum likelihood estimation. As the derivatives of h_K are easily obtained, standard algorithms such as NPSOL (Gill, Murray, Saunders & White, 1983) may be used to fit either the constrained or unconstrained version of the problem.

Confidence intervals can be computed for the elements of τ and functionals $\sigma(h)$ of h using maximum likelihood formulae since $\sigma(h_K)$ will be a function of θ; for example, if $\sigma(h) = \int z^\lambda h(z)\, dz$ then $\sigma(h_K)$ is the ratio of two polynomials in θ. In simpler settings than the one considered here, confidence intervals constructed from truncation estimators in this fashion are asymptotically correct (Andrews, 1991; Eastwood & Gallant, 1991; Eastwood, 1991; Gallant & Souza, 1991). Simulations by Eastwood & Gallant (1991) suggest that deterministic rules such as $\hat{K}_n = n^\alpha$ do not yield accurate confidence intervals in small samples, because they do not use the sample information to adapt to the roughness of h^o, whereas adaptive rules do.

Most adaptive rules are based on criteria that pick the value of K that minimizes an expression of the form $s_n(\hat{\tau}, \hat{h}_K) + c(N)(p_{net}/N)$, where $p_{net} = p_\tau + p_\theta - 1$ if the constraint $\int z h(z)\, dz = 0$ is not imposed and $p_{net} = p_\tau + p_\theta - M - 1$ if it is. The term $c(N)(p_{net}/N)$ is a penalty factor designed to compensate for small $s_n(\hat{\tau}, \hat{h}_K)$ achieved by fitting an over parameterized model.

These criteria have been extensively studied when $(-N)s_n(\hat{\tau}, \hat{h}_K)$ is replaced in the expression above by the optimized log likelihood of a linear regression model $E(y_i) = x_i'\beta$ for which the rule of formation $(x_{i1}, x_{i2}, \ldots, x_{ip}) = x_i'$ of the regressors is known, as in the case of lags in time series analysis or the case of successive terms of a Fourier series expansion. For the formula that converts a rule based on the residual sum of squares to a rule based on the log likelihood see Gallant (1987, p. 366). Under standard regularity conditions with the true value of p assumed to be finite, any criterion that satisfies $\lim_{n\to\infty} c(N)/N = 0$ will not underfit in large samples (Potscher, 1989). When the true value of p is assumed to

be finite but standard regularity conditions are violated because $(1/N)\sum_{i=1}^{N} x_i x_i'$ does not converge to a constant, the largest penalty factor that does not underfit in large samples is $c(N) = (1/2)\log N$ which corresponds to the Schwarz criterion (Potscher, 1989). Under standard regularity conditions with the true value of p assumed to be finite, the smallest penalty factor that does not overfit in large samples is $c(N) = \log\log N$ which corresponds to the Hannan-Quinn criterion (Hannan, 1987). If p is assumed to be infinite, the penalty factor that adds terms at an optimal rate puts $c(N)$ to a constant which is the Akaike criterion when $c(N) = 1$ (Eastwood & Gallant, 1991; Eastwood, 1991).

Our recommendation is to inspect plots such as Figure 2 and 3 for all models between those chosen by the Schwarz and Akaike criteria inclusively and make a visual selection. We cannot state the case for visual inspection better than Silverman:

> A natural method for choosing the smoothing parameter [K] is to plot out several curves and choose the estimate that is most in accordance with one's prior ideas about the density. For many applications this approach will be perfectly satisfactory. Indeed, the process of examining several plots of the data, all smoothed by different amounts, may well give more insight into the data than merely considering a single automatically produced curve. (Silverman, 1986, p. 44)

If one insists upon an automatic selection rule we recommend the Hannan-Quinn criterion because, upon checking several published time series applications of Hermite expansions, we found that the Hannan-Quinn criterion usually selected the same model that the authors of these articles had selected after extensive diagnostic testing. The Schwarz criterion nearly always chose a smaller model than the authors and the Akaike criterion nearly always selected a larger model.

As pointed out by a referee, one can also use a mixture of seminonparametric densities weighted by a prior distribution on K if desired.

One structural aspect of the truncation estimator h_K deserves comment. If $K = 0$ then h_K is the normal density; that is, the normal density is the leading term in the expansion of h^o. This is a substantial advantage in applications, especially in high dimensional applications, where the normal distribution is a reasonable first approximation making the estimator $\hat{h}_K$ an ideal candidate for nonparametric time series analysis, which is where it has seen most frequent use. See Gallant, Rossi & Tauchen (1992), for a time series application in which

the estimated density receives an extensive graphical examination; see their references for additional applications and comparisons with other nonparametric estimators. Also, the fact that the leading term of the series is the normal density provides a convenient means to test the hypothesis that h^o is normal. One can compare the optimized likelihood for $K > 0$ with that for $K = 0$ using one of the model selection criteria discussed above or the asymptotic χ^2 test. The asymptotic χ^2 statistic for a choice between specifications $K_H < K_A$ having $p_{net} = p_H$ and p_A, respectively, is $2N\{s_n(\hat{\tau}_H, \hat{h}_{K_H}) - s_n(\hat{\tau}_A, \hat{h}_{K_A})\}$ on $p_A - p_H$ degrees of freedom. When the asymptotic χ^2 test is used for model selection it behaves very much like the Akaike criterion when N or $p_A - p_H$ are large.

Imposing the constraint $\int z\,\hat{h}(z)\,dz = 0$ usually has little effect on estimates and can be convenient when reporting results. Sometimes, however, the constraint increases the value of K required to obtain an adequate fit. We recommend not imposing it unless the K selected, the estimates of τ, and the visual appearance of the estimated density remain essentially unchanged. When $K = 1$, $\int z\,\hat{h}(z)\,dz = 0$ imposes normality.

Putting the off-diagonal elements of R to zero improves numerical stability, especially when M is large. We recommend that it be imposed if estimates of τ and the visual appearance of the fitted density are little changed.

Now consider computation of an integral of the form $\int \psi(z) h_K(z)\,dz$ which is a ratio with numerator $\int \psi(z)\{P_K(R^{-1}z)\}^2 n_M(z|0, RR')\,dz$ and denominator $\int\{P_K(u)\}^2 n_M(u|0, I)\,du$. The denominator is easily computed as noted above. A change of variables puts the numerator in a form suited to Gauss-Hermite quadrature (Davis & Rabinowitz, 1975)

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi(\sqrt{2}Rt) P_K(\sqrt{2}t) \pi^{-M/2} \prod_{i=1}^{M} \exp(-t_i^2)\,dt_i,$$

where $t = (t_1, \ldots, t_M)'$. A Gauss-Hermite rule has the form $\int_{-\infty}^{\infty} \phi(s)\exp(-s^2)\,ds \approx \sum_{i=1}^{L} W_i \phi(s_i)$. Thus,

$$\int \psi(z)\{P_K(R^{-1}z)\}^2 n_M(z|0, RR')\,dz \approx \sum_{i_1=1}^{L} \cdots \sum_{i_M=1}^{L} \psi(\sqrt{2}Rs) P_K(\sqrt{2}s) \pi^{-M/2} \prod_{j=1}^{M} W_{i_j},$$

where $s = (s_{i_1}, \ldots, s_{i_M})'$. The abscissae s_i and weight factors W_i can be obtained from tabulations such as Table 25.10 of Abramowitz & Stegun (1964) or can be computed as needed using an algorithm due to Golub (1973); see also Golub & Welsch (1969). Note that if ψ is differentiable then analytic derivatives with respect to θ of the expression on the right are easily obtained.

4. Pharmacokinetics of phenobarbital

We illustrate the use of the proposed methods by application to routine clinical pharmacokinetic data collected from $n = 59$ newborn infants treated with phenobarbital during the first sixteen days after birth as reported by Grasela & Donn (1985). Each individual received an initial dose of phenobarbital (μg/kg) followed by one or more sustaining doses by intravenous administration. A total J_i of anywhere from 1 to 6 blood samples were obtained from each individual at times (hours) other than dosage times as part of routine monitoring. Phenobarbital concentration (μg/ml) was determined for each blood sample by high pressure liquid chromatographic assay. The total number of concentration measurements was $N = 155$. Information collected from each individual included birth weight (kg) and 5-minute Apgar score. These data are described in detail and analyzed by Grasela & Donn (1985); see also Boeckmann, Sheiner & Beal (1990).

The pharmacokinetics of phenobarbital may be described by a one-compartment open model with intravenous bolus administration and first-order elimination (Grasela & Donn, 1985). This model states that mean plasma concentration of phenobarbital in individual i at time t due to a dose D_{ij} administered at time t_{ij} is given by $(D_{ij}/\beta_{2i})\exp\{-(t-t_{ij})\beta_{1i}/\beta_{2i}\}$, $t > t_{ij}$, where β_{1i} is the total clearance of phenobarbital in (liter/hour)/kg and β_{2i} is the apparent volume of distribution of phenobarbital in liter/kg. Since individuals received several doses over the study period, concentration is a sum of such terms; one term for each dose with $t_{ij} < t$. Usually this model is written not as a sum but as a recursion (Grasela & Donn, 1985).

Blood concentration measurements are often approximately normally distributed and exhibit variability that increases with level (Beal & Sheiner, 1988). This phenomenon is attributable in part to the error in the high pressure liquid chromatographic assay used to process blood samples. Thus an assumption that

$$p_e(e_{i1}, \ldots, e_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, \sigma, \beta_i) = \prod_{j=1}^{J_i} n[e_{ij} | \{\sigma f(x_{ij}, \beta_i)\}^2]$$

is both reasonable and permits comparison of our results with Grasela & Donn (1985).

Grasela & Donn (1985) adopted the inter-individual regression model $\beta_{1i} = \gamma_1 w_i e^{z_{1i}}$, $\beta_{2i} = \gamma_2 w_i (1 + \gamma_3 \delta_i) e^{z_{2i}}$, after extensive model evaluation, where w_i is the birth weight of the ith individual and δ_i is a dummy variable which is 1 if the 5-minute Apgar score of individual i is less than 5 and is 0 otherwise. With this specification, if the z_i are symmetrically

distributed then clearance and volume will have skewed distributions, which accords well with experience with this type of data. We used this specification to permit a comparison with previous analyses of these data (Grasela & Donn, 1985; Boeckmann, Sheiner & Beal, 1990) but comment further below.

We fit this model subject to the constraint $\int zh(z)\,dz = 0$ for $K = 0, 2, 3, 4$ using the methods described in Section 3. The optimization results are displayed in Table 1. Also displayed in Table 1 are the Schwarz, Hannan-Quin, and Akaike criteria. The Schwarz and Hannan-Quin criteria select the normal density ($K = 0$) whereas the Akaike criterion selects the $K = 2$ seminonparametric density.

The estimates for both specifications $K = 0$ and $K = 2$ are displayed in Table 2 together with the estimates of Grasela & Donn (1985) who used the First Order linearization method due to Beal & Sheiner (1982) as implemented by Boeckmann, Sheiner & Beal (1990).

Graphics associated with the models selected by the Schwarz, Hannan-Quin, and Akaike criteria are shown in Figure 1. The most interesting feature is the bi-modality of the seminonparametric estimate $\hat{h}_2$ seen in panels (a) and (b) which divides the sample into the two groups seen in panels (c) and (d). In response to a query, Professor Grasela told us that the seven infants represented by diamonds in panels (c) and (d) had low measured concentrations after the loading dose. The initial concentration measurement is more influential for apparent volume of distribution than for clearance in this model, which is well known in the pharmacokinetics literature and explains the appearance of panel (d). These low concentrations did not seem to be associated with any attribute that was measured in the study. A relevant, unmeasured attribute or a misspecified inter-individual regression are possible explanations.

We took the specification of the inter-individual regression model above as a given in order to illustrate our proposed method by comparison with previously reported results. However, it is usually necessary to determine an appropriate inter-individual regression model from the data. Rather than using hypothesis tests to determine the model, Davidian & Gallant (1993) suggest a graphical strategy based on seminonparametric empirical Bayes estimates. Other procedures based on empirical Bayes estimates have been proposed; see Maitre, Buhrer, Thomson & Stanski (1991) and Mandema, Verotta & Sheiner (1992). We caution, however, that validity, reliability, and comparative performance of procedures based on empirical Bayes estimates is an open problem.

5. Simulation results

We applied the method to four simulated data sets in order to assess its ability to reveal modes or bumps in the random effects density under conditions likely to be encountered in practice: sparse and unequal numbers of observations per individual and coefficients of variation of the random parameters around 27%.

In the simulation, the intra-individual regression function was the unit-dose mono-exponential $f(x_{ij}, \beta_i) = (1/\beta_{2i}) \exp\{-(\beta_{1i}/\beta_{2i}) x_{ij}\}$, where x_{ij} is the time of observation j, $1 \leq j \leq J_i$, on individual i, $1 \leq i \leq 110$. Intra-individual errors were normal with standard deviation proportional to level

$$p_e(e_{i1}, \ldots, e_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, \sigma, \beta_i) = \prod_{j=1}^{J_i} n[e_{ij} | 0, \{\sigma f(x_{ij}, \beta_i)\}^2],$$

with $\sigma = 0.05$. The inter-individual random parameters for clearance and volume were $\beta_{1i} = e^{\gamma_1 + z_{1i}}$, $\beta_{2i} = e^{\gamma_2 + z_{2i}}$, respectively, where $\gamma = \log(0.1, 0.5)$.

For each individual, J_i was randomly selected from the uniform distribution on the integers 1 through 5. The times of observation, given J_i, were randomly selected from the U[0, 0.9] distribution on the scale of proportion of dose eliminated with respect to the rate constant $\int \beta_1/\beta_2 h(z) dz$.

Bi-modal random effects densities $h(z)$ were generated by mixing two normal distributions $N(\mu, RR')$ and $N(-\mu, RR')$ with mixing proportion α and $\mu = \{(sep/2)\sqrt{(r_{11}^2 + r_{12}^2)}, 0\}'$. For $sep > 2$, this density has a visually perceptible second mode or bump. The separation in the modes or bumps is with respect to the random effect for clearance z_1 with the modes separated by sep standard deviations of z_1. The four simulations correspond to values of $\alpha = 0.3, 0.5$ and $sep = 2.5, 4.0$.

The elements of the upper triangular matrix R determine the coefficients of variation for β_1 and β_2 and were chosen to achieve coefficients of variation between 25% and 28% for both β_1 and β_2 throughout. In each case $r_{22} = 0.24$, $r_{12} = 0.0$; whereas $r_{11} = 0.16$ for $sep = 2.5$ and $r_{11} = 0.12$ for $sep = 4.0$.

We generated one data set for each case using the same initial seed. Using the density $p(y_{i1}, \ldots, y_{iJ_i} | x_{i1}, \ldots, x_{iJ_i}, \tau, z_i)$ implied by the unit-dose mono-exponential model with constant intra-individual coefficient of variation that was used to generate the data sets, we applied the method of Section 3, increasing K until the Akaike criterion, the least stringent,

ceased to decline. The results of the optimizations are reported in Table 3. In all but one case, the three model selection criteria selected the $K = 2$ specification. For the case $sep = 4$, $\alpha = 0.3$, the criteria split between the $K = 1$ and $K = 3$ specifications; graphs of these two estimated densities differ very little and would lead to the same conclusions regarding the shape of the density in any application.

Figures 2 and 3 are graphical displays of the estimated densities with $K = 2$ for the two extreme cases $sep = 4$, $\alpha = 0.5$ and $sep = 2.5$, $\alpha = 0.3$. The estimated and true densities agree reasonably closely; more importantly, the estimated densities convey the correct qualitative impression. The same is true of the two omitted plots.

6. Discussion

In this paper we propose a method for maximum likelihood estimation of the fixed parameters of the nonlinear mixed effects model together with the density of the random effects. It is a truncation method based on a series representation of the density due to Gallant & Nychka (1987) that follows from an assumption that the density is smooth. Once estimates are obtained, empirical Bayes estimation of the random parameters is straightforward.

We applied the method to pharmacokinetic data reported by Grasela & Donn (1985) and discovered interesting features of the inter-individual random effects density unlikely to be revealed by the use of parametric methods.

Application of the method to simulated data suggests that it can produce reliable qualitative information regarding the possibility of bumps and modes in the random effects density.

A Fortran program implementing the method is in the public domain. It is available, together with a User's Guide as a PostScript file, either via ftp anonymous at ccvr1.cc.ncsu.edu (128.109.212.20) in directory pub/arg/nlmix or from the Carnegie-Mellon University e-mail server by sending the one-line e-mail message "send nlmix from general" to statlib@lib.stat.cmu.edu. The program computes parameter estimates, empirical Bayes estimates of the random effects, data for plotting, and simulations from the estimated density. Runtimes for the computations reported in Sections 4 and 5 were less than 15 minutes on a Sun SparcStation 2 from every start value we tried; the time required to generate plots, empirical Bayes estimates, and simulations is much less.

Acknowledgements

This material is based upon work supported by the National Science Foundation and the North Carolina Agricultural Experiment Station. We should like to thank Thaddeus H. Grasela Jr. and Lewis B. Sheiner for use of data and Thomas M. Ludden for helpful discussions. We should also like to thank the referees for a careful reading and for recommending substantial improvements.

References

Abramowitz, M. & Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover.

Andrews, D. W. K. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric models. *Econometrica* 59, 307–46.

Beal, S. L. & Sheiner, L. B. (1982). Estimating population kinetics. *CRC Crit. Rev. in Biomed. Eng.* 8, 195–222.

Beal, S. L. & Sheiner, L. B. (1985). Methodology of population pharmacokinetics. In *Drug Fate and Metabolism-Methods and Techniques,* Ed. E. R. Garrett & J. L. Hirtz., pp. 135–83. New York: Marcel Dekker.

Beal, S. L. & Sheiner, L. B. (1988). Heteroskedastic nonlinear regression. *Technometrics* 30, 327-38.

Boeckmann, A. J., Sheiner, L. B. & Beal, S. L. (1990). *NONMEM User's Guide, Part V, Introductory Guide.* San Francisco: University of California at San Francisco.

Davidian, M. & Gallant, A. R. (1992). Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. *J. Pharmacokin. Biopharm.* 20, 531–58.

Davidian, M. & Giltinan, D. M. (1993). Some simple methods for estimating intra-individual variability in nonlinear mixed effects model. *Biometrics* 49. To appear.

Davis, P. J. & Rabinowitz, P. (1975). *Methods of Numerical Integration.* New York: Academic Press.

Eastwood, B. J. (1991). Asymptotic normality and consistency of semi-nonparametric regression estimators using an upward F test truncation rule. *J. Econometrics* 48, 151–82.

Eastwood, B. J. & Gallant, A. R. (1991). Adaptive truncation rules for seminonparametric estimators that achieve asymptotic normality. *Econometric Theory* 7, 307–40.

Gallant, A. R. (1987). *Nonlinear Statistical Models.* New York: Wiley.

Gallant, A. R. & Nychka, D. W. (1987). Seminonparametric maximum likelihood estimation. *Econometrica* 55, 363–90.

Gallant, A. R., Rossi, P. E. & Tauchen, G. E. (1992). Stock prices and volume. *Rev. Financ. Stud.* 5, 199–242.

Gallant, A. R. & Souza, G. (1991). On the asymptotic normality of Fourier flexible form estimates. *J. Econometrics.* 50, 329–53.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. & Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Am. Statist. Assoc.* 85, 972–85.

Gill, P. E., Murray, W., Saunders, M. A. & White, M. H. (1983). *User's Guide for SOL/NPSOL: A Fortran Package for Nonlinear Programming.* Palo Alto, California: Stanford University.

Golub, G. H., (1973). Some modified matrix eigenvalue problems. *SIAM Review* 15, 318–34.

Golub, G. H. & Welsch, J. H., (1969). Calculation of Gaussian quadrature rules. *Math. Comp.* 23, 221–30.

Grasela, T. H., Jr. & Donn, S. M. (1985). Neonatal population pharmacokinetics of phenobarbital derived from routine clinical data. *Dev. Pharmacol. Ther.* 8, 374–83.

Hannan, E. J. (1987). Rational transfer function approximation. *Statist. Sci.* 2, 1029–54.

Lindstrom, M. L. & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* 46, 673–87.

Maitre, P. O., Buhrer, M., Thomson, D. & Stanski, D. R. (1991). A three-step approach combining Bayesian regression and NONMEM population analysis: application to midazolam. *J. Pharamacokin. Biopharm.* 19, 377–84 .

Mallet, A. (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika* 73, 645–56.

Mallet, A., Mentré, F., Steimer, J-L & Lokiec, F. (1988). Nonparametric maximum likelihood estimation for population pharmacokinetics, with application to cyclosporine. *J. Pharmacokin. Biopharm.* 16, 311–27.

Mandema, J. W., Verotta, D. & Sheiner, L. B. (1992). Building population pharmacokinetic-pharmacodynamic models. *J. Pharamacokin. Biopharm.* 20. To appear.

Potscher, B. M. (1989). Model selection under nonstationarity: autoregressive models and stochastic linear models. *Ann. Statist.* 17, 347–70.

Racine-Poon, A. (1985). A Bayesian approach to nonlinear random effects models. *Biometrics* 41, 1015–23.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Steimer, J. L., Mallet, A., Golmard, J. L. & Boisvieux, J. F. (1984). Alternative approaches to estimation of population pharmacokinetic parameters: comparison with the nonlinear mixed effect model. *Drug Metab. Rev.* 15, 265–92.

Vonesh, E. F. & Carter, R. L. (1992). Mixed effects nonlinear regression for unbalanced repeated measures. *Biometrics* 48, 1–18.

Table 1. *Optimization results for phenobarbital data, $n = 59$ individuals: degree of the polynomial part of h_K (K); effective number of parameters (p_{net}); the negative of the optimized log-likelihood divided by the total number ($N = 155$) of measured concentrations ($s_n(\tau, h)$); the Schwarz (BIC), Hannan-Quinn (HQ), and Akaike (AIC) model selection criteria.*

K	p_{net}	$s_n(\hat{\tau}, \hat{h}_K)$	BIC	HQ	AIC
0	7	2.79914	2.91302	2.87221	2.84430
2	10	2.77116	2.93385	2.87555	2.83567
3	14	2.76630	2.99406	2.91244	2.85662
4	19	2.75613	3.06524	2.95447	2.87871

Table 2. *Parameter estimates for phenobarbital data, $n = 59$ individuals: as reported by Grasela & Donn (1985) (Grasela & Donn), maximum likelihood estimates with normal ($K = 0$) random effects density (Normal Effects), maximum likelihood estimates with seminonparametric ($K = 2$) random effects density (SNP Effects).*

	Grasela & Donn		Normal Effects		SNP Effects	
Parameter	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.
γ_1	0.0047	0.0002	0.0048	0.0002	0.0047	0.0003
γ_2	0.96	0.024	0.9780	0.0300	0.9808	0.0362
γ_3	0.135	–	0.1449	0.0638	0.1054	0.0596
σ	0.107	–	0.1129	0.0126	0.1096	0.0131
$\mathrm{var}(u_1)$	–	–	0.0471	0.0189	0.0450	0.0359
$\mathrm{var}(u_2)$	–	–	0.0224	0.0092	0.0271	0.0143
$\mathrm{cov}(u_1, u_2)$	0.0	constr.	0.0179	0.0122	0.0144	0.0176

Table 3. *Optimization results for four simulated data sets, $n = 110$ individuals sampled from a mixture of two normal distributions: degree of the polynomial part of h_K (K); effective number of parameters (p_{net}); the negative of the optimized log-likelihood divided by the total number ($N = 336$) of measured concentrations ($s_n(\tau, h)$); the Schwarz (BIC), Hannan-Quinn (HQ), and Akaike (AIC) model selection criteria; separation of the clearance random effect in standard deviations (sep) and the mixture proportion (α).*

K	p_{net}	$s_n(\hat{\tau}, \hat{h}_K)$	BIC	HQ	AIC	K	p_{net}	$s_n(\hat{\tau}, \hat{h}_K)$	BIC	HQ	AIC
		sep $= 4$, $\alpha = 0.5$						sep $= 2.5$, $\alpha = 0.5$			
0	5	-0.4847	-0.4414	-0.4585	-0.4698	0	5	-0.4751	-0.4318	-0.4489	-0.4602
1	7	-0.5923	-0.5317	-0.5556	-0.5715	1	7	-0.5284	-0.4678	-0.4917	-0.5075
2	10	-0.6256	-0.5391	-0.5732	-0.5959	2	10	-0.5883	-0.5018	-0.5359	-0.5586
3	14	-0.6346	-0.5134	-0.5612	-0.5929	3	14	-0.5916	-0.4704	-0.5183	-0.5500
		sep $= 4$, $\alpha = 0.3$						sep $= 2.5$, $\alpha = 0.3$			
0	5	-0.4974	-0.4542	-0.4712	-0.4826	0	5	-0.5357	-0.4924	-0.5095	-0.5208
1	7	-0.6397	-0.5791	-0.6030	-0.6188	1	7	-0.5702	-0.5096	-0.5335	-0.5493
2	10	-0.6512	-0.5647	-0.5988	-0.6215	2	10	-0.6319	-0.5453	-0.5795	-0.6021
3	14	-0.6743	-0.5531	-0.6009	-0.6326	3	14	-0.6361	-0.5149	-0.5627	-0.5944
4	20	-0.6890	-0.5158	-0.5841	-0.6294						

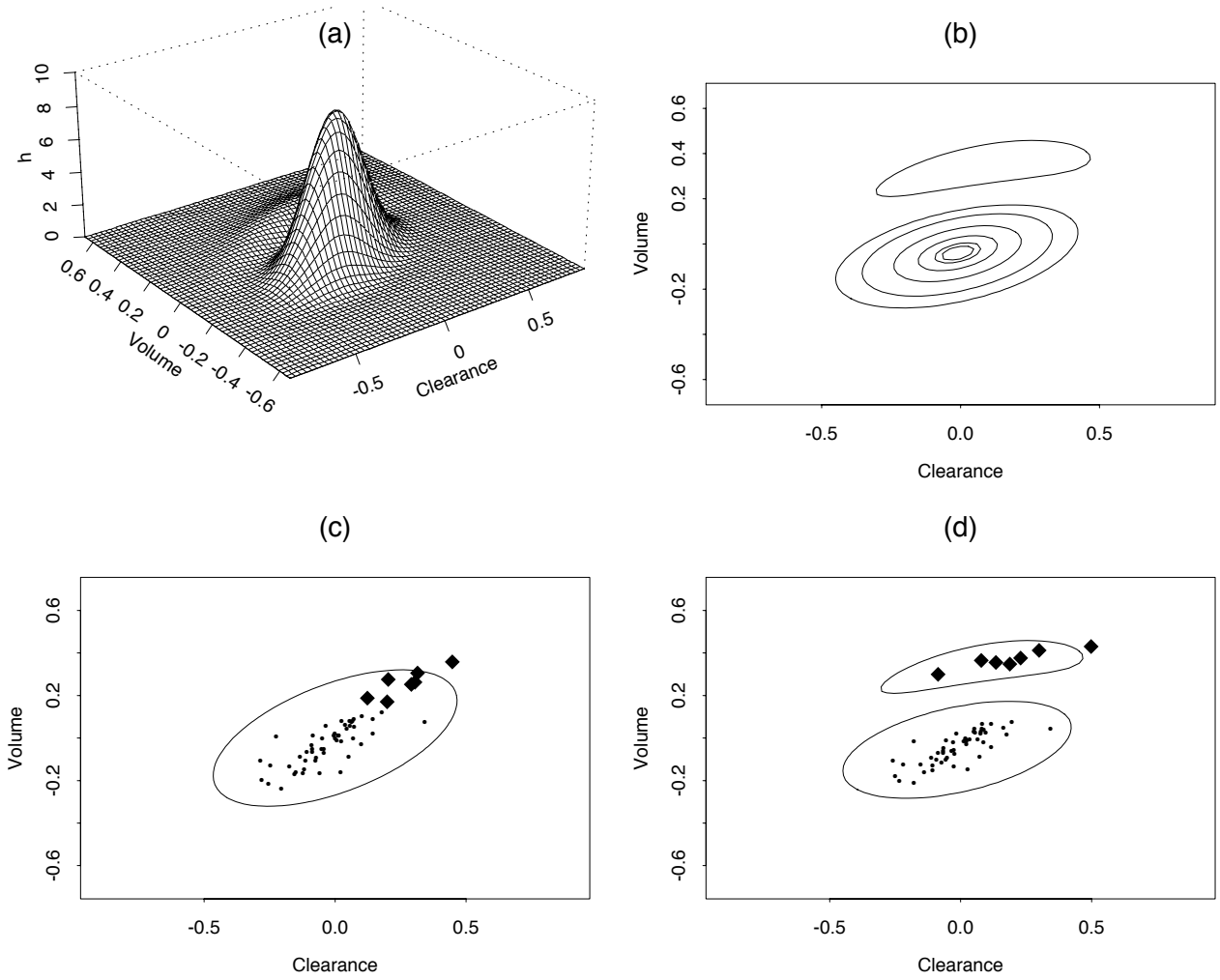

Fig. 1. *Estimated inter-individual random effects densities and empirical Bayes estimates of inter-individual random effects for the phenobarbital data: (a) perspective plot of the estimated seminonparametric* $(K = 2)$ *density; (b) contour plot of the estimated seminonparametric* $(K = 2)$ *density at quantiles 10%, 25%, 50%, 75%, 90%, and 95%; (c) 10% quantile contour of the estimated normal density* $(K = 0)$ *and corresponding empirical Bayes estimates of the inter-individual effects (dots and diamonds); (d) 10% quantile contour of estimated seminonparametric density* $(K = 2)$ *and corresponding empirical Bayes estimates of the inter-individual effects (dots and diamonds). Diamonds flag the same individuals in both panels.*

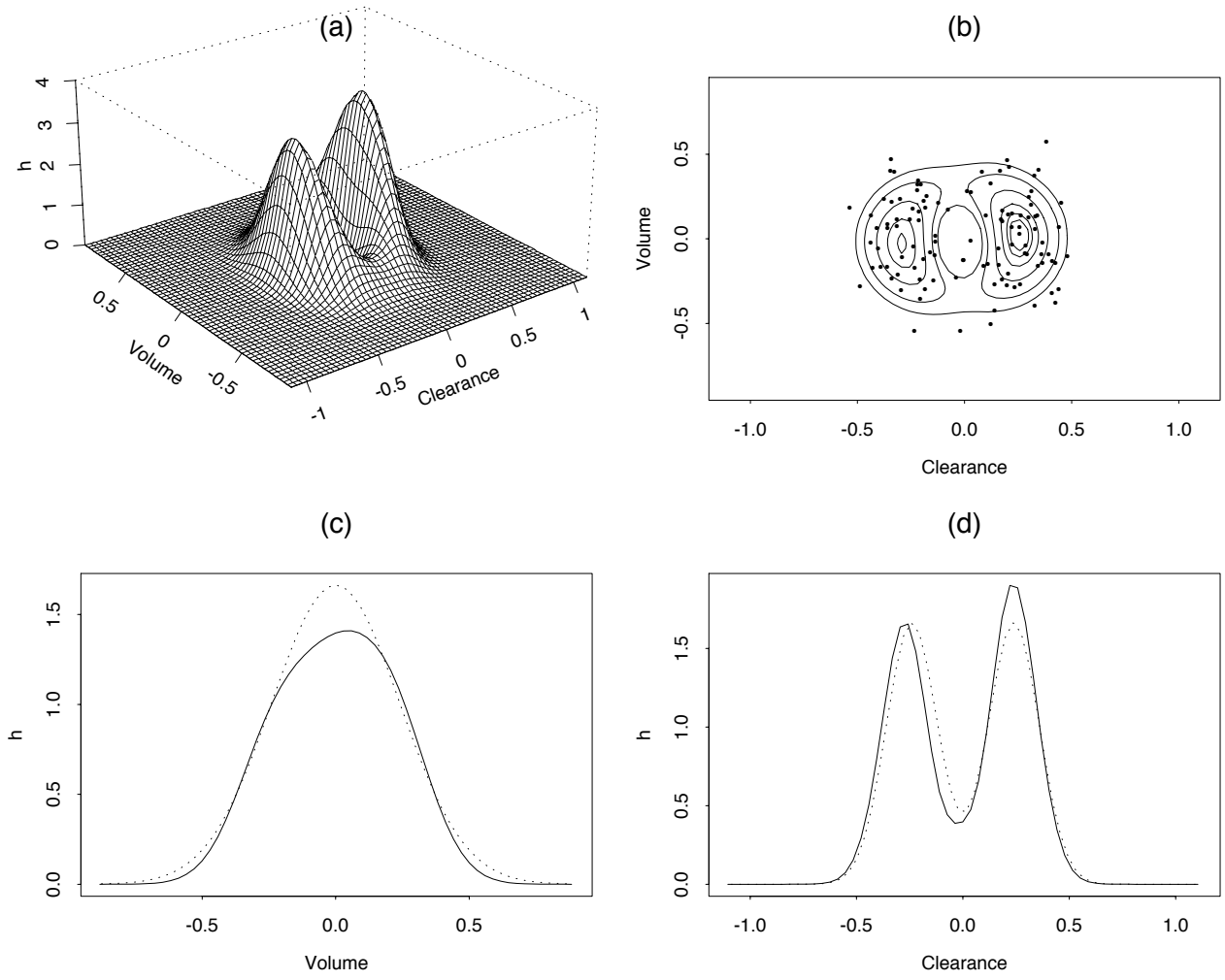

Fig. 2. *Estimated seminonparametric inter-individual random effects density* ($K = 2$) *and true density for simulated data,* $sep = 4.0$ *and* $\alpha = 0.5$ *: (a) perspective plot of the estimated joint seminonparametric density; (b) contour plot of the estimated joint seminonparametric density at quantiles 10%, 25%, 50%, 75%, 90%, and 95% and the simulated random effects (dots), the middle contour is 10%; (c) marginal density for volume, integral of the joint seminonparametric density (solid line) and true density (dotted line); (d) marginal density for clearance, integral of the joint seminonparametric density (solid line) and true density (dotted line).*

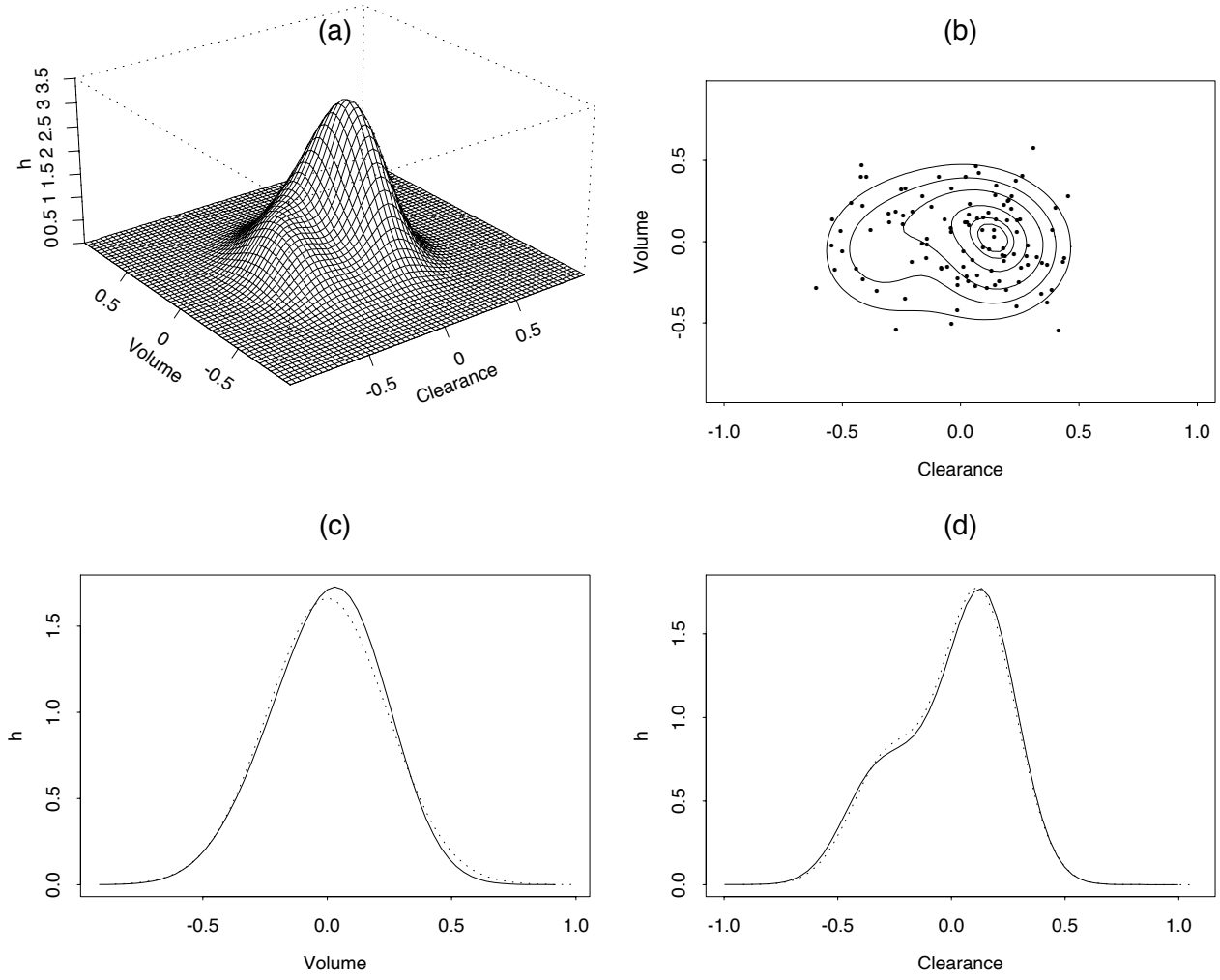

Fig. 3. *Estimated seminonparametric inter-individual random effects density* ($K = 2$) *and true density for simulated data,* $sep = 2.5$ *and* $\alpha = 0.3$: *(a) perspective plot of the estimated joint seminonparametric density; (b) contour plot of the estimated joint seminonparametric density at quantiles 10%, 25%, 50%, 75%, 90%, and 95% and the simulated random effects (dots); (c) marginal density for volume, integral of the joint seminonparametric density (solid line) and true density (dotted line); (d) marginal density for clearance, integral of the joint seminonparametric density (solid line) and true density (dotted line).*