The Annals of Statistics
1996, Vol. 24, No. 3, 1145–1160

ROBUST INFERENCE FOR VARIANCE COMPONENTS MODELS FOR SINGLE TREES OF CELL LINEAGE DATA

By R. M. Huggins

La Trobe University

Previously, Huggins and Staudte examined robust estimators for a variance components formulation of the bifurcating autoregressive model for cell lineage data. They gave asymptotic properties of the estimators if a large number of trees were observed. However, for single trees the derivation of these asymptotic properties is more complex. Here the asymptotic distributions of robust estimators of parameters associated with the stationary bifurcating autoregressive process as a single tree becomes large are obtained. These results follow from the formulation of the estimating functions as the product of a nonrandom matrix and the sum of vectors of functions of an infinite sequence of exchangeable random variables.

1. Introduction. The collection of trees of cell lineage data is a laborious process, involving time lapse photography and the inspection of a series of films of the descendants of an initial cell, from which characteristics of the individual cells and their relationship to one another can be determined. Data on cell lifetimes of *E.coli* was initially collected by E. O. Powell in the 1950s and 1960s [Powell (1955, 1956, 1958); Powell and Errington (1963)]. One of the aims of Powell's studies was to determine a biological basis for the high correlations between the lifetimes of sister cells. The biological background of the statistical analysis of cell lineages has been discussed in Huggins and Staudte (1994), who also gave references to cell lineage data collected from a range of cell lines. Scientific interest is largely in determining if there are nonzero correlations between mother and daughter cells or between the daughter cells. These correlations are an explicit part of the bifurcating autoregressive model of Cowan and Staudte (1986) which we examine here.

A typical tree of cell lifetimes, tree 41 of Staudte, Guiget and Collyn d'Hooge (1984) and Figure 1 of Huggins and Staudte (1994), is given in Figure 1. There is one extreme outlier, several other possible outliers and the initial observation and one further observation are missing. Missing initial observations are typical of trees of cell lifetimes as the birth time of the initial cell, which has been randomly selected from some population of cells, is rarely observed. However, the division of this cell is observed so that the relationships between its descendants are known. Typically other missing cells wander from the field of view so that even if they or their descendants reappear, their relationship with the cells whose lineage is observed is not known.

Received November 1992; revised January 1996.

AMS 1991 *subject classifications*. Primary 62F35; secondary 62F12, 62M99.

Key words and phrases. Robust estimation, asymptotic inference, variance components, cell lineage data.

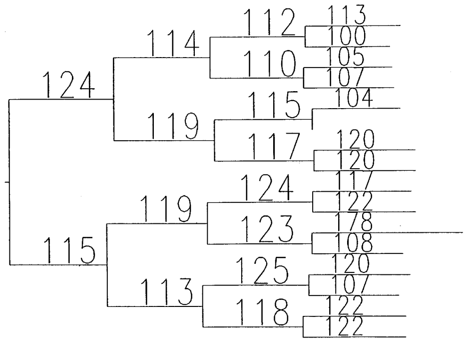

FIG. 1. *A tree of* EMT6 *cells* [*tree* 41 *of Staudte, Guiget and Collyn d'Hooge* (1984)].

The presence of outliers indicates that it is desirable to have an outlier-resistant estimation procedure as an alternative to maximum likelihood, if only to verify that the maximum likelihood estimates have not been unduly influenced by a few aberrant observations, or perhaps to conduct an initial screening of the data. The absence of some observations requires flexible statistical procedures which may be applied to any observed structure of a tree of cell lineage data. Huggins and Staudte (1994) considered robust inference for variants of the bifurcating autoregressive model of Cowan and Staudte (1986) based on observations on a large number of independent trees. However, in practice it is possible that a scientist may only be able to collect data on one large tree rather than many smaller trees. Here we show it is theoretically feasible to conduct robust inference for the stationary bifurcating autoregressive process using data from a single large tree, leaving the determination of any optimality properties of the estimators to a later date. The development of similar inference procedures for variants of this model, such as models with nonstationary means or variances or models which incorporate measurement error, remains an open problem. A further open problem concerns the estimation of the correlation between mothers and daughters and between daughters using data from a single tree if the means vary from generation to generation. This requires a separate parameter for the mean of individuals in each generation, and there will be too few observations on the early generations for the asymptotic approach presented here to be valid. Another unsolved problem is to determine how to incorporate the information from the left-censored initial observation and the cells which wander from the field of view, and are thus right censored, into the robust methodology.

Cowan and Staudte (1986) introduced the bifurcating autoregressive model for cell lineage trees. In their model, cells in a tree are labelled $1, 2, \ldots$ with cell n giving rise at division to daughter cells $2n$ and $2n+1$. Let x_n denote a characteristic, such as the lifetime, of the nth cell. In the bifurcating autoregression process,

$$x_{2n} = \theta x_n + w_{2n},$$
$$x_{2n+1} = \theta x_n + w_{2n+1}, \tag{1.1}$$

where $|\theta| < 1$ and the pairs (w_{2n}, w_{2n+1}) have a bivariate normal distribution with common mean and variance and correlation coefficient φ. The observations are generated by x_1 and the independent sequence of independent pairs (w_{2n}, w_{2n+1}), although as noted above for some characteristics such as cell lifetime, x_1 may not be observed. Of course for other characteristics such as cell diameter at division the initial individual will be readily observable. A balanced tree is a tree in which there are no missing observations, with the possible exception of the initial observation. In an unbalanced tree observations on at least one cell, apart from the initial cell, and its descendants are not observed. In the stationary bifurcating autoregressive process the joint distribution of (x_n, x_{2n}, x_{2n+1}) is the same for all n and the unconditional correlation coefficient for sister cells is $\rho = \mathrm{Corr}[x_{2n}, x_{2n+1}] = \theta^2 + (1-\theta^2)\varphi$.

We derive asymptotic properties of robust estimators for the parameters associated with a stationary bifurcating autoregressive process. However, our representation result, Theorem 2.2, is more general than this as it allows the means to be linear functions of some vector β of parameters, and it allows sisters to have different means. We also relax the assumption that the errors have a bivariate normal distribution and replace this with an assumption of elliptic symmetry.

Note that the bifurcating autoregressive process satisfies a "Markov" property in the sense that the joint distribution of x_{2n} and x_{2n+1} depends on $x_n, \dots, x_1$ only through x_n. The "Markov" nature of the bifurcating autoregressive process allows the use of simpler convergence results than those used by Miller (1977), which were based on results of Weiss (1973, 1975), to study maximum likelihood estimators in mixed linear models. The "Markov" property does not hold for the measurement error model of Huggins and Staudte (1994), so that the simplifications of our Corollary 2.2 are not available for that model, suggesting an extension of our results to this latter case may be quite difficult.

The maximum likelihood analysis of this model, under the assumption of multivariate normality, has been discussed by Cowan and Staudte (1986). Robust methods, based on a time series approach have been given by Huggins and Marschner (1991), but their methods required the use of estimated residuals to estimate the sister–sister correlation ρ. Here we consider robust inference based on a variance components model, developed by Huggins and Staudte (1994), which allows more flexible and tractable models for the mean and covariance structure and simultaneously estimates all the parameters in the model. Huggins (1993a, b) and Huggins and Staudte (1994) previously considered robust inference for pedigrees, repeated measures and cell lineage data using independent sampling units. In that setting the asymptotic properties of the estimators as the number of sampling units becomes large are straightforward and differ from the more complex situation involving only one tree considered here. A further difference between the results of Huggins and Staudte (1994) and the results presented here is that they regarded the structures of the trees, that is, which individuals were or were not observed, as being random quantities, whereas we condition on the observed structure of the cell lineage tree.

It is supposed that the vector of observations $X = (x_1, \dots, x_n)^t$ on some characteristic of cells in a tree has mean vector $\tilde{\mu}$ and covariance matrix Ω, where Ω is determined by the structure of the tree. The general model for the mean is of the form $\tilde{\mu} = Y\beta$ for some vector of parameters β and design matrix Y and this model is used in Theorem 2.2. However, our asymptotic results are given for the special case where Y is a $n \times 1$ vector of ones and β consists of the single parameter μ. Huggins and Staudte (1994) defined the covariance matrix Ω as follows: Let the matrix D give the generation distances separating cells in the tree. It consists of elements $d_{ij} = m+n$, where m and n are the respective numbers of generations from each of i and j to their nearest common ancestor. By definition $d_{ii} = 0$ and $d_{ij} = n$ if i is a direct ancestor of j living n generations earlier. Also introduce the direct lineage indicator matrix L having elements l_{ij}, which takes the value 1 if i and j are on the same line, that is, either i is an ancestor of j or vice versa; let $l_{ij} = 0$ otherwise. We define the diagonal elements to be $l_{ii} = 1$. Let J be a matrix of ones. Then for the bifurcating autoregression process with stationary variance, if we set $\mathrm{var}(x_n) = \sigma^2$, then

$$\Omega = \sigma^2 L \times \theta^D + \sigma^2 \rho (J - L) \times \theta^{D-2J},$$

where θ^D is the matrix with elements $\theta^{d_{ij}}$, and for matrices of the same dimensions, $\times$ denotes element by element matrix multiplication. Let $\eta = (\beta^t, \theta, \sigma^2, \rho)^t$ be the vector of parameters in the model.

To find robust estimates we suppose Ω is positive definite so that we may write $\Omega = AA^t$ and then let $Z = (z_1, \dots, z_n)^t = A^{-1}(X - \tilde{\mu})$. Note that the matrix A is not unique. In Section 2 a version of A, suitable for our purposes, is determined. In order to obtain the asymptotic properties of our estimators, we assume the distribution of X is elliptically symmetric in the sense that the distribution of Z is spherically symmetric. Let Ψ be a twice differentiable function, let $\psi = \Psi'$ be an odd function and let $\mathscr{K} = E(\psi(z_1)z_1)$. Huggins and Staudte (1994) proposed estimating η by minimizing the sum, over N independent trees, of

$$M = \sum_{j=1}^{n} \Psi(z_j) + \frac{\mathscr{K}}{2} \ln(|\Omega|),$$

with respect to η, which is a robust version of the likelihood under the assumption of multivariate normality. We retain the same form of the estimating functions for all spherically symmetric distributions, and the effect of changing distributional assumptions is only to change the quantity $\mathscr{K}$.

We consider estimating η by minimising M over observations $X = (x_1, \dots, x_n)^t$ on a single tree. This minimization approach results in the estimating functions for β,

$$V_{\beta, n} = -Y^t A^{-t} \psi(Z), \tag{1.2}$$

and for $\alpha = \theta$, σ^2 or ρ,

$$V_{\alpha, n} = \psi(Z)^t \frac{dA^{-1}}{d\alpha} A Z + \frac{\mathscr{K}}{2} \operatorname{tr}\left(\Omega^{-1} \frac{d\Omega}{d\alpha}\right). \tag{1.3a}$$

Note that (1.3a) is not a bounded function of the z_j, and in order to bound the effect of large z_j, we also consider replacing (1.3a) by

$$\bar{V}_{\alpha, n} = -\psi(Z)^t A^{-1} \frac{d\Omega}{d\alpha} A^{-t} \psi(Z) + \bar{\mathscr{K}} \operatorname{tr}\left(\Omega^{-1} \frac{d\Omega}{d\alpha}\right), \tag{1.3b}$$

where $\bar{\mathscr{K}} = E(\psi^2(z_1))$. We refer to estimators arising from minimizing M as robust I estimators and those arising from zeroes of (1.2) and (1.3b) are referred to as robust II estimators. Robust I estimators have computational advantages in that only the model for the mean vector and the covariance matrix need be specified, and the estimators may be found numerically using a minimisation procedure. On the other hand, the estimating functions for the robust II estimators need to be explicitly specified.

Let $V_n(\eta)$ or $\bar{V}_n(\eta)$ denote the respective vectors of estimating functions. Here we are concerned with the asymptotic behavior of the two robust estimators as the size of a single tree becomes large. Under the assumption of elliptic symmetry it is shown that the residuals form an exchangeable sequence of random variables so that de Finetti's theorem may be used to determine the asymptotic properties of the estimators. We do not require that the tree be balanced, although for notational convenience, much of the theory is derived in this setting.

Our approach is based on a matrix representation of the estimating functions for the parameters of the balanced stationary bifurcating autoregressive model in the form $C \sum_{j=1}^{n_0} P_j + C^* P^*$. Here n_0 is the number of observed mother-daughter triples, C and C^* are nonrandom matrices, P_j is a vector of functions of the residuals corresponding to the jth mother–daughters triple and P^* is a function of the residuals corresponding to the initial individual, if this individual has been observed. If the initial individual has not been observed, there will be an initial term arising from the joint distribution of the first two sisters that were observed. This term does not affect the asymptotic results and is not considered further. Thus asymptotic properties of the estimating functions may be derived from a consideration of $\sum_{j=1}^{n_0} P_j$, which is shown to be a vector-valued martingale.

The simple form of the matrix representation is due to the "Markov property" of the bifurcating autoregressive process and the stationarity assumption. If it is not assumed that the process is stationary, then the estimating functions take the more complex form $\sum_{j=1}^{n_0} C_j P_j + C^* P^*$ which is not considered here. Related representation results for general mixed linear models may be easily extracted from Theorem 2.2 as in Huggins (1996). In the general case this representation is of the form $E_n F_n$, where, given the structure of the data, E_n is a deterministic matrix and F_n is a vector of functions of the residuals. In the general representation the columns of E_n and the rows of F_n increase with n. This complicates its use in deriving the asymptotic properties of estimators in the general case. However, Huggins (1996) has applied this general result to address identifiability in the bifurcating autoregressive model with measurement error, and identifiability in models for repeated measures experiments and pedigree data may be similarly examined.

The choice of Ψ as twice differentiable or ψ as differentiable is no real practical restriction and allows relatively simple proofs, comparable to those required for maximum likelihood estimators. For example, we may take Ψ corresponding to Tukey's bisquare or, as an approximation to Huber's ψ, use $\psi(x) = 2\Phi(x/c) - 1$, where Φ is the normal cumulative distribution function and c is a constant. The extension to other less smooth ψ functions appears feasible, but would be far more difficult technically.

An important difference between the results presented here for a single tree and those of Huggins and Staudte (1994) for a sequence of trees is that in order to preserve the martingale property of the estimating equations, $\mathscr{K}$ is now an unobservable random variable rather than a fixed constant. An exception is the special case when the tree does in fact have a multivariate normal distribution. In practice, $\mathscr{K}$ is usually taken to be the appropriate quantity for the multivariate normal distribution so that the estimators are consistent for this model.

Note that the robust estimating functions are based on conditional residuals given the previous individuals in the tree and are thus sensitive to the order within the sister–sister pairs. While the problems with order within the sister–sister pairs could be overcome by choosing a symmetric decomposition of Ω, rather than the lower triangular decomposition considered here, this would result in a different, less convenient, interpretation of the residuals Z. Moreover, there is no simple characterisation corresponding to our Corollary 2.2, which would complicate the asymptotic theory.

2. Properties of the estimating functions. We state two results, Theorems 2.1 and 2.2, for mixed linear models in general. Corollaries 2.1–2.3 are specific to the stationary bifurcating autoregressive process. Outlines of the proofs are given in Section 4.

THEOREM 2.1. *Let the random vector X_n have mean vector $\tilde{\mu}_n = Y_n\beta = (\mu_1, \dots, \mu_n)^t$, for some design matrix Y_n and vector β of parameters. Let Ω_n denote the covariance matrix of X_n. Let A_n be the lower triangular Cholesky decomposition of Ω_n and let $Z_n = (z_{n1}, \dots, z_{nn})^t = A_n^{-1}(X_n - \tilde{\mu}_n)$. Then there exists a sequence of uncorrelated zero mean random variables $z_1, z_2, \dots$ such that $A_n^{-1}(X_n - \tilde{\mu}_n) = Z_n = (z_1, \dots, z_n)^t$. Further if the distribution of X_n is el-*

liptically symmetric, then the conditional distributions of z_{n+1} given $z_n, \ldots, z_1$ are symmetric about zero, and the sequence $z_1, z_2, \ldots$ is an exchangeable sequence of martingale differences.

Thus for our choice of A_n, the z_{nj} do not depend on n. In particular, $Z_n = (z_{n1}, \ldots, z_{nn})^t$ may be embedded in an infinite sequence $Z = (z_1, \ldots, z_n)^t$ of exchangeable random variables. As a consequence of the proof of Theorem 2.1, the z_n corresponding to the stationary bifurcating autoregressive process may be written in an accessible form.

COROLLARY 2.1. *Let $X_n = (x_1, \ldots, x_n)^t$ be the first n observations on a bifurcating autoregressive process. Let*

$$\Gamma_1 = \begin{pmatrix} \sigma^2 & \sigma^2\theta \\ \sigma^2\theta & \sigma^2 \end{pmatrix} \quad \textit{and} \quad \Gamma_{12} = \begin{pmatrix} \sigma^2\theta \\ \rho\sigma^2 \end{pmatrix}.$$

Denote the mean vector of the mother–daughters triple by $(\mu_n, \mu_{2n}, \mu_{2n+1})^t$, let $x_2 = (x_n, x_{2n})^t$, let $\mu_2 = (\mu_n, \mu_{2n})^t$ and define $\delta_e^2 = \sigma^2(1-\theta^2)$ and $\delta_o^2 = \sigma^2 - \Gamma_{12}^t\Gamma_1^{-1}\Gamma_{12}$. Then the z_n of Theorem 2.1 are of the form

$$z_{2n} = \delta_e^{-1}(x_{2n} - \mu_{2n} - \theta(x_n - \mu_n))$$

and

$$z_{2n+1} = \delta_o^{-1}(x_{2n+1} - \mu_{2n+1} - \Gamma_{12}^t\Gamma_1^{-1}(x_2 - \mu_2)).$$

In view of Theorem 2.1, we may apply de Finetti's theorem and suppose the existence of a σ-field $\mathscr{G}$ such that given $\mathscr{G}$, $z_1, z_2, \ldots$ is a sequence of independently and identically distributed random variables. For a random variable R, let $E_{\mathscr{G}}(R) = E(R|\mathscr{G})$. Following the argument of Lemma 2.1.1 of Taylor, Daffer and Patterson (1985), note that as the z_i [and also the $\psi(z_i)$ and $\psi'(z_i)z_i$] are uncorrelated, we have that $0 = E(z_1z_2) = E[E_{\mathscr{G}}(z_1z_2)] = E[E_{\mathscr{G}}^2(z_1)]$ so that $E_{\mathscr{G}}(z_1) = 0$ (and similarly $E_{\mathscr{G}}[\psi(z_1)] = 0$ and $E[\psi'(z_1)z_1] = 0$). Further note that, in general, $E[\psi(z_n)z_n - \mathscr{K}] = 0$, but this does not imply that $E_{\mathscr{G}}[\psi(z_n)z_n - \mathscr{K}] = 0$ or that $E_{\mathscr{G}}[\psi(z_n)z_n - \mathscr{K}|z_{n-1}, \ldots, z_1] = 0$. Let $\mathscr{K}_{\mathscr{G}} = E_{\mathscr{G}}[\psi(z_1)z_1]$ and $\bar{\mathscr{K}}_{\mathscr{G}} = E_{\mathscr{G}}[\psi^2(z_1)]$, so that using the conditional independence of $z_1, z_2, \ldots,$ given $\mathscr{G}$, we have $E_{\mathscr{G}}[\psi(z_n)z_n - \mathscr{K}_{\mathscr{G}}|z_{n-1}, \ldots, z_1] = E_{\mathscr{G}}[\psi(z_n)z_n - \mathscr{K}_{\mathscr{G}}] = 0$. Thus in the sequel we minimize $M_{\mathscr{G}, n} = \sum_{j=1}^n \rho(z_j) + \mathscr{K}_{\mathscr{G}}/2\ln(|\Omega_n|)$, so that (1.1) is now written as $V_{\beta, n} = -Y_n^tA_n^{-t}\psi(Z_n)$, (1.3a) is $V_{\alpha, n} = \psi(Z)^t(dA^{-1}/d\alpha)AZ + (\mathscr{K}_{\mathscr{G}}/2)\operatorname{tr}(\Omega^{-1}(d\Omega/d\alpha))$ and (1.3b) is now $\bar{V}_{\alpha, n} = -\psi(Z_n)A_n^{-1}(d\Omega_n/d\alpha)A_n^{-t}\psi(Z_n) + \bar{\mathscr{K}}_{\mathscr{G}}\operatorname{tr}(\Omega_n^{-1}(d\Omega_n/d\alpha))$.

To show that the estimating equations are in fact martingales, let y_1 denote the first row of Y, write $Y_{n+1} = (Y_n^t, y_{n+1}^t)^t$, $Z_{n+1} = (Z_n^t, z_{n+1})^t$, partition Ω_{n+1} and define δ_{n+1} and r_{n+1} as above. The martingale property is a consequence of the following theorem and the above properties of $(z_1, z_2, \ldots)$. Like Theorem 2.1, Theorem 2.2 applies to mixed linear models in general.

THEOREM 2.2. *Let X_n be as in Theorem* 2.1 *and let $V_{\beta,n}$, $V_{\alpha,n}$ and $\bar{V}_{\alpha,n}$ be defined by* (1.2), (1.3a) *and* (1.3b) *respectively. Then we may write*

$$\begin{aligned} V_{\beta,1} &= -\delta_1^{-1} y_1^t \psi(z_1),\\ V_{\alpha,1} &= -\delta_1^{-1} \frac{d\delta_1}{d\alpha}(\psi(z_1)z_1 - \mathscr{K}_{\mathscr{G}}),\\ V_{\beta,n+1} &= V_{\beta,n} - (Y_n^t r_{n+1}^t + \delta_{n+1}^{-1} y_{n+1}^t)\psi(z_{n+1}), \end{aligned} \tag{2.1}$$

and if α is one of the parameters in the model for Ω,

$$\begin{aligned} V_{\alpha,n+1} = V_{\alpha,n} &- (\psi(z_{n+1})z_{n+1} - \mathscr{K}_{\mathscr{G}})\delta_{n+1}^{-1}\frac{d\delta_{n+1}}{d\alpha}\\ &- \psi(z_{n+1})\delta_{n+1}^{-1}\frac{da_{n+1}^t\,\Omega_n^{-1}}{d\alpha}A_n Z_n, \end{aligned} \tag{2.2}$$

$$\begin{aligned} \bar{V}_{\alpha,n+1} = \bar{V}_{\alpha,n} &+ 2(\psi^2(z_{n+1}) - \bar{\mathscr{K}}_{\mathscr{G}})\delta_{n+1}^{-1}\frac{d\delta_{n+1}}{d\alpha}\\ &+ 2\psi(z_{n+1})\delta_{n+1}^{-1}\frac{da_{n+1}^t\,\Omega_n^{-1}}{d\alpha}A_n \psi(Z_n). \end{aligned} \tag{2.3}$$

The general expressions of Theorem 2.2 admit some simplification for the balanced stationary bifurcating autoregressive process.

COROLLARY 2.2. *Using the notation of Corollary* 2.1, *for the balanced stationary bifurcating autoregressive process and $n > 1$, we obtain the following simplifications of terms in Theorem* 2.2:

$$-(Y_{2n-1}^t r_{2n}^t + \delta_e^{-1} y_{2n}^t) = -\delta_e^{-1}(-y_n^t \theta + y_{2n}^t),$$

$$-(Y_{2n}^t r_{2n+1}^t + \delta_o^{-1} y_{2n+1}^t) = -\delta_o^{-1}(-(y_n^t\ y_{2n}^t)\Gamma_1^{-1}\Gamma_{12} + y_{2n+1}^t),$$

$$\frac{da_{2n}^t\,\Omega_{2n-1}^{-1}}{d\alpha}A_{2n-1}Z_{2n-1} = \frac{d\theta}{d\alpha}(x_n - \mu_n),$$

and

$$\frac{da_{2n+1}^t\,\Omega_{2n}^{-1}}{d\alpha}A_{2n}Z_{2n} = \frac{d\Gamma_{12}^t\,\Gamma_1^{-1}}{d\alpha}(x_2 - \mu_2).$$

The latter two expressions depend on the past of the process only through x_n and x_{2n} and x_n, respectively.

Theorem 2.2 and Corollary 2.2 allow the matrix representation for the stationary bifurcating process exploited in Theorem 3.1.

COROLLARY 2.3. *Let X_n denote a balanced stationary bifurcating autoregressive process. There exist nonrandom matrices $C(\eta)$ and $C^*(\eta)$ and random vectors $P_j(\eta)$ and $P_j^*(\eta)$ such that the resulting estimating functions $V_n(\eta)$ may be written in the form $C(\eta)\sum_{j=1}^{n_0} P_j(\eta) + C^*(\eta)P^*(\eta)$.*

To prove Corollary 2.3, note that for each j the additions to the estimating functions due to observation $j+1$ consist of a random and nonrandom component. To construct the matrix representation, the random components corresponding to observations x_{2j} and x_{2j+1} are collected into the vector

$$P_j(\eta) = \big(\psi(z_{2j}), \psi(z_{2j+1}), \psi(z_{2j})z_{2j} - \mathscr{K}_{\mathscr{G}}, \psi(z_{2j+1})z_{2j+1} - \mathscr{K}_{\mathscr{G}},\\ \psi(z_{2j})(x_j - \mu), \psi(z_{2j+1})(x_j - \mu), \psi(z_{2j+1})(x_{2j} - \mu)\big)^t.$$

To incorporate the initial observation, let $P^*(\eta) = (\psi(z_1), \psi(z_1)z_1 - \mathscr{K}_{\mathscr{G}})^t$. In the stationary case, where $y_n = 1$, the nonrandom components are then written in matrix form as follows. Define $K_e = -\delta_e^{-1}(1-\theta)$, $K_o = -\delta_o^{-1}(1-(11)\Gamma_1^{-1}\Gamma_{12})$ and for each of $\alpha = \theta$, σ^2 or ρ let $K_{1\alpha} = -\delta_e^{-1}(d\delta_e/d\alpha)$, $K_{2\alpha} = -\delta_o^{-1}(d\delta_o/d\alpha)$, $K_{3\alpha} = -\delta_e^{-1}(d\theta/d\alpha)$ and let $K_{4\alpha} = (K_{41\alpha}, K_{42\alpha}) = -\delta_o^{-1}\, d\Gamma_{12}^t\Gamma_1^{-1}/d\alpha$. Define $C(\eta)$ and $C^*(\eta)$ by

$$\begin{pmatrix} K_e & K_o & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K_{1\theta} & K_{2\theta} & K_{3\theta} & K_{41\theta} & K_{42\theta} \\ 0 & 0 & K_{1\sigma^2} & K_{2\sigma^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & K_{2\rho} & 0 & K_{41\rho} & K_{42\rho} \end{pmatrix},$$

and

$$\begin{pmatrix} -\delta_1^{-1}y_1^t & 0 \\ 0 & -\delta_1^{-1}d\delta_1/d\theta \\ 0 & -\delta_1^{-1}d\delta_1/d\sigma^2 \\ 0 & -\delta_1^{-1}d\delta_1/d\rho \end{pmatrix},$$

respectively, where $C^*(\eta)$ is needed to calculate the contribution from the initial individual.

3. Asymptotic properties of the estimators for the stationary bifurcating autoregressive process. The consistency and asymptotic normality of the estimators may be derived following Marschner (1991), who modified an earlier result of Huggins and Marschner (1991). See Crowder (1976, 1986) and Klimko and Nelson (1978) for related results. We say that a positive random variable W is bounded away from zero if for all $\varepsilon > 0$ there exists a $\Delta > 0$ such that $P(W > \Delta) > 1 - \varepsilon$. Let $V_n(\eta)$ be a set of differentiable estimating equations for $\eta \in \mathscr{R}^p$ and let $G_n(\eta) = n^{-1}\,dV_n(\theta)/d\eta$. The limit in probability of a sequence W_n of random variables is denoted by $\text{p-lim}_{n\to\infty} W_n$.

THEOREM A [Marschner (1991)]. *Suppose that $n^{-1}V_n(\eta_0) \to_p 0$, $G_n(\eta_0) \to_p G(\eta_0)$, where $G(\eta_0)$ is a symmetric and positive definite, $n^{-1/2}V_n(\eta_0) \to_d N(0, \Sigma(\eta_0))$ for some positive definite matrix $\Sigma(\eta_0)$ and*

$$\operatorname*{p-lim}_{n\to\infty} \sup_{||\eta-\eta_0||<\delta} ||G_n(\eta) - G_n(\eta_0)|| \le H(\delta),$$

where $H(\delta) > 0$ *a.s. and* $\text{p-lim}_{\delta\to 0} H(\delta) = 0$. *Then there exists a sequence* $\hat{\eta}_n$ *of solutions of* $V_n(\eta) = 0$ *such that* $\hat{\eta}_n \to_p \eta_0$ *and* $n^{1/2}(\hat{\eta}_n - \eta_0) \to_d N(0, G(\eta_0)^{-1}\Sigma(\eta_0)G(\eta_0)^{-1})$.

The first three conditions concern the behaviour of the vector estimating functions and its matrix of derivatives under the "true" model. The first two conditions and the fourth continuity condition are required to show the estimators are consistent using Brouwer's fixed point theorem as in Aitchison and Silvey (1958) and Crowder (1976, 1986). The central limit theorem for the estimating functions is used to show the estimators are asymptotically normal via the mean value theorem. In practice, consistent estimators of $G(\eta_0)$ and $\Sigma(\eta_0)$ are required. The consistency of $\hat{\eta}_n$ and the continuity conditions in Theorem A imply that $G_n(\hat{\eta}_n) \to_p G(\eta_0)$. In order to estimate Σ_0, we require an estimator $\Sigma_n(\eta_0)$ satisfying $\Sigma_n(\eta_0) \to_p \Sigma(\eta_0)$; continuity conditions on $\Sigma_n(\eta)$ equivalent to those on $G_n(\eta)$ are also required.

3.1. *The robust* I *estimators.* We use the matrix representation of $V_{n_0}(\eta)$ of Corollary 2.3 and Theorem A to derive the asymptotic properties of the estimators in the stationary case. The parameters of interest are $\eta = (\mu, \theta, \sigma^2, \rho)^t$. Initially suppose that the tree is balanced and let $n_0 = [n/2]$. Define, $\Sigma^*_{n_0}(\eta) = n_0^{-1}\sum_{j=1}^{n_0} P_j(\eta)P_j(\eta)^t$ and $\Sigma_{n_0}(\eta) = C(\eta)\Sigma^*_{n_0}(\eta)C^t(\eta) + C^*(\eta)P^*P^{*t}C^{*t}(\eta)$. Further, let $\Sigma(\eta_0) = E_{\mathscr{G}}(\psi^2(z_1))C(\eta_0)\Sigma^*(\eta_0)C^t(\eta_0)$ and $G(\eta_0) = E_{\mathscr{G}}(\psi'(z_1))C(\eta_0)G^*(\eta_0)C^t(\eta_0)$, where $\Sigma^*(\eta)$ and $G^*(\eta)$ are defined by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & C_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 & \theta\sigma^2 \\ 0 & 0 & 0 & 0 & 0 & \theta\sigma^2 & \sigma^2 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & C_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 & \theta\sigma^2 \\ 0 & 0 & 0 & 0 & 0 & \theta\sigma^2 & \sigma^2 \end{pmatrix},$$

respectively, and C_1 and C_2 are defined by $C_1 = E_{\mathscr{G}}(\psi(z_1)z_1 - \mathscr{K}_{\mathscr{G}})^2/E_{\mathscr{G}}(\psi^2(z_1))$ and $C_2 = [E_{\mathscr{G}}(\psi(z_1)z_1) + E_{\mathscr{G}}(\psi'(z_1)z_1^2)]/E_{\mathscr{G}}(\psi'(z_1))$.

THEOREM 3.1. *Let X_n be the balanced stationary bifurcating autoregressive process. Suppose that $E(z_1^2) < \infty$, ψ is bounded, continuous and odd, ψ' is even and bounded, $|\psi'(x) - \psi'(y)| \leq b|x - y|$ for some constant b and $E((x_j - \mu)^4) < \infty$. Then there exists a sequence $\hat{\eta}_{n_0}$ of solutions of $V_{n_0}(\eta) = 0$ such that $n_0^{1/2}(\hat{\eta}_{n_0} - \eta_0) \to_d N(0, G(\eta_0)^{-1}\Sigma(\eta_0)G(\eta_0)^{-1})$ and both $G_{n_0}(\hat{\eta}_{n_0}) \to_p G(\eta_0)$ and $\Sigma_{n_0}(\hat{\eta}_{n_0}) \to_p \Sigma(\eta_0)$, where $G(\eta_0)$ is positive definite.*

To prove Theorem 3.1, note that Corollary 2.3, the exchangeability of the z_i and the laws of large numbers of Huggins (1995) now allow a straightforward, if tedious, checking of the conditions of Theorem A by applying standard martingale theory to the martingale $\sum_{j=1}^{n_0} P_j(\eta_0)$. To do this, the convergence of various quantities is required and Huggins (1995) has shown that for the stationary bifurcating autoregressive process with $|\theta| < 1$, constant mean μ and variance $\sigma^2 < \infty$, $n^{-1}\sum_{j=1}^{n} x_j \to \mu$ a.s., $n^{-1}\sum_{j=1}^{n} x_j^2 \to_p \sigma^2 + \mu^2$ and for any $\varepsilon > 0$, for large enough n, with large probability, $n^{-1}\sum_{j=1}^{n} |x_j| \leq (E|w - \gamma| + \gamma)/(1 - \theta) + \varepsilon$, where $\gamma = \mu(1 - \theta)$ and w has the common distribution of the w_n. For example, to show that $n_0^{-1}\sum_{j=1}^{n_0} \psi(z_{2j})x_j \to 0$ a.s., we note that $\sum_{j=1}^{n_0} \psi(z_{2j})x_j$ is a martingale with conditional variance $E_{\mathscr{G}}(\psi^2(z_1))\sum_{j=1}^{n_0} x_j^2$ and so that the above convergence results and the law of large numbers for martingales [Shiryayev (1984), page 487] yield the desired result. The Lindeberg condition may be easily checked and an application of the corresponding central limit theorem for martingales may be applied to show $n_0^{-1/2}V_n(\eta_o) \to_d N(0, \Sigma(\eta_0))$. The checking of the continuity conditions on $G_n(\eta)$ is tedious but straightforward.

The more difficult part is to show that $G(\eta_0)$ is positive definite. However, using the formulation of $G(\eta_0)$ in terms of $C(\eta_0)$ and $G^*(\eta_0)$ above, it is only necessary to show that the rows of $C(\eta)$ are linearly independent. This can be easily done using a computer algebra package as in Huggins (1996).

The extension to the common case where the tree is unbalanced, in the sense that only one of a pair of sisters may be observed or even neither of the sisters are observed, is straightforward, the only real change being the use of laws of large numbers for the unbalanced rather than the balanced case. Let n denote the number of observed individuals, let n_1 denote the number of observed sister–sister pairs and let n_2 denote the number of single sisters observed. Then $n = 2n_1 + n_2$. Let $n_0 = n_1 + n_2$. Consider the pairs where both sisters were observed separately from those where only one sister was observed. Let $P_j(\eta)$, $j = 1, \ldots, n_1$, denote the vector $P_j(\eta)$ as previously defined for pairs where both sisters were observed and let $P_j^*(\eta)$, $j = 1, \ldots, n_2$, be the corresponding vector for those pairs where only one sister was observed. Let

$$D(\eta) = \begin{pmatrix} K_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K_{1\theta} & 0 & K_{3\theta} & 0 & 0 \\ 0 & 0 & K_{1\sigma^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

so that the estimating functions are now of the form $V_{n_0}(\eta)=C(\eta)\sum_{j=1}^{n_1} P_j(\eta)+D(\eta)\sum_{j=1}^{n_2} P^*_j(\eta)$.

THEOREM 3.2. *Let X_n be the unbalanced stationary bifurcating autoregressive process. Suppose that the conditions of Theorem* 3.1 *hold, the distribution of n_1 and n_2 are independent of the distribution of the full* (*unobserved*) *tree $\{x_j,\ j=1,\dots,n\}$ and $n_i/n_0\to_p p_i$, $i=1,2$, for some constants p_i, $i=1,2$, with $p_1>0$. Then there exists a sequence $\hat\eta_{n_0}$ of solutions of $V_{n_0}(\eta)=0$ such that $n_0^{1/2}(\hat\eta_{n_0}-\eta_0)\to_d N(0,G(\eta_0)^{-1}\Sigma(\eta_0)G(\eta_0)^{-1})$ and both $G_{n_0}(\hat\eta_{n_0})\to_p G(\eta_0)$ and $\Sigma_{n_0}(\hat\eta_{n_0})\to_p\Sigma(\eta_0)$, where $\Sigma(\eta_0)=p_1C(\eta_0)\Sigma^*(\eta_0)C^t(\eta_0)+p_2D(\eta_0)\Sigma^*(\eta_0)\cdot D^t(\eta_0)$ and $G(\eta_0)=p_1C(\eta_0)G^*(\eta_0)C^t(\eta_0)+p_2D(\eta_0)G^*(\eta_0)D^t(\eta_0)$, with $G(\eta_0)$ being positive definite.*

3.2. *The robust* II *estimators.* Let $\tilde X=\mu+A\psi(Z)$ and denote the jth element of $\tilde X$ by $\tilde x_j$. Let $\tilde P^t_j(\eta)$ be defined by

$$\big(\psi(z_{2j}),\psi(z_{2j+1}),(\psi^2(z_{2j})-K_{1\mathscr{G}}),(\psi^2(z_{2j+1})-K_{1\mathscr{G}}),\\ \psi(z_{2j})(\tilde x_j-\mu),\psi(z_{2j+1})(\tilde x_j-\mu),\psi(z_{2j+1})(\tilde x_{2j}-\mu)\big).$$

The robust II estimators are of the form $C(\eta)\sum_{j=1}^{n_0}\tilde P_j(\eta)+C^*(\eta)\tilde P^*(\eta)$.

To state the result we need to modify some of our previous definitions. Let $\tilde\Sigma^*(\eta)$ be $\Sigma^*(\eta)$ with C_1 and C_2 replaced by $C_1^*=E_{\mathscr{G}}(\psi^2(z_1)-K_{1,\mathscr{G}})^2/E_{\mathscr{G}}(\psi^2(z_1))$ and $C_2^*=2[E_{\mathscr{G}}(\psi(z_1)z_1\psi'(z_1))]/E_{\mathscr{G}}(\psi'(z_1))$, respectively, and σ^2 replaced by $\tilde\sigma^2=E_{\mathscr{G}}(\psi^2(z_1))\sigma^2$. Let

$$C_3=(1-\theta^2)^{-1}\big[(1/2)E_{\mathscr{G}}(z_1\psi(z_1))(\delta_o^2+\delta_e^2(1+(\rho-\theta^2)/(1-\theta^2)))\big],$$

$$C_3^*=E_{\mathscr{G}}(\psi'(z_1))C_3$$

and

$$C_4=2E_{\mathscr{G}}(\psi'(z_1)\psi(z_1)z_1),$$

and define $\tilde G^*(\eta_0)$ to be $G^*(\eta_0)$ with C_2 replaced by C_4 and σ^2 replaced by C_3^*. Further define $\tilde\Sigma(\eta_0)=C(\eta)\tilde\Sigma^*(\eta_0)C^t(\eta)$ and $\tilde G(\eta_0)=C(\eta)\tilde G^*(\eta_0)C^t(\eta)$, let $\tilde\Sigma^*_{n_0}(\eta)=n_0^{-1}\sum_{j=1}^{n_0}\tilde P_j(\eta)\tilde P^t_j(\eta)$ and let $\tilde\Sigma_{n_0}(\eta)=C(\eta)\tilde\Sigma^*_{n_0}(\eta)C^t(\eta)$.

THEOREM 3.3. *Let X_n be the balanced stationary bifurcating autoregressive process. Suppose that $E(z_1^2)<\infty$, ψ is bounded, continuous and odd and ψ' is even and bounded. Then there exists a sequence $\tilde\eta_{n_0}$ of solutions of $\tilde V_{n_0}(\eta)=0$ such that $n_0^{1/2}(\tilde\eta_n-\eta_0)\to_d N(0,\tilde G(\eta_0)^{-1}\tilde\Sigma(\eta_0)\tilde G(\eta_0)^{-1})$ and both $\tilde G_{n_0}(\tilde\eta_n)\to_p\tilde G(\eta_0)$ and $\tilde\Sigma_{n_0}(\tilde\eta_n)\to_p\tilde\Sigma(\eta_0)$, where $\tilde G(\eta_0)$ is positive definite.*

The derivation of the asymptotic behaviour of the robust II estimators is similar to that of the robust I estimators with some technical complications. If we note that $\tilde{X}$ is still a bifurcating autoregressive process, the required convergence results are easily checked. Checking the continuity condition is extremely tedious in this case. The unbalanced case is similar to Theorem 3.2 and is omitted.

4. Proofs. In this section we outline the proofs of some of the main results.

PROOF OF THEOREM 2.1. First, it is clear from its definition that Z_n is a vector of zero mean uncorrelated random variables for each n. Next, write

$$\Omega_{n+1} = \begin{pmatrix} \Omega_n & a_{n+1} \\ a_{n+1}^t & b_{n+1} \end{pmatrix},$$

$$A_{n+1} = \begin{pmatrix} A_n & 0 \\ l_{n+1} & \delta_{n+1} \end{pmatrix}$$

and

$$A_{n+1}^{-1} = \begin{pmatrix} A_n^{-1} & 0 \\ r_{n+1} & \delta_{n+1}^{-1} \end{pmatrix},$$

where $l_{n+1} = a_{n+1}^t A_n^{-t}$, $\delta_{n+1} = (b_{n+1} - l_{n+1} l_{n+1}^t)^{1/2}$ and $r_{n+1} = -\delta_{n+1}^{-1} a_{n+1}^t \Omega_n^{-1}$, so that, using the results of Stewart [(1973), page 141], $\Omega_{n+1} = A_{n+1} A_{n+1}^t$. Hence, $Z_{n+1} = A_{n+1}^{-1}(X_{n+1} - \tilde{\mu}_{n+1}) = (Z_n^t, z_{n+1,\, n+1})$ so that, in particular, $z_{n+1} = z_{n+1,\, n+1}$ is defined by the first $n+1$ members of the tree and $z_{n+1} = \delta_{n+1}^{-1}[(x_{n+1} - \mu_{n+1}) - a_{n+1}^t \Omega_n^{-1}(X_n - \tilde{\mu}_n)]$. The second part of the theorem follows from noting that the joint distribution of $z = (z_1, \dots, z_n)^t$ is spherically symmetric, that $z_1, \dots, z_n$ are exchangeable, for example, Kingman (1972), and that by symmetry the conditional density of z_{n+1} given $z_n, \dots, z_1$ depends on z_{n+1} only through z_{n+1}^2 [Kelker (1970)], and is thus symmetric about zero. □

PROOF OF COROLLARY 2.1. The results of Kelker (1970) directly show that the mean of x_{n+1} given $x_1, \dots, x_n$, is $\mu_{n+1} + a_{n+1}^t \Omega_n^{-1}(X_n - \tilde{\mu}_n)$ and the "Markov" property of the bifurcating autoregressive process implies that $E(x_{2n} | x_{2n-1}, \dots, x_1) = E(x_{2n} | x_n)$ and $E(x_{2n+1} | x_{2n}, \dots, x_1) = E(x_{2n+1} | x_{2n}, x_n)$, which may be identified as $\mu_{2n} + \theta(x_n - \mu_n)$ and $\mu_{2n+1} + \Gamma_{12}^t \Gamma_1^{-1}(x_2 - \mu_2)$, respectively. To identify δ_{2n}^2 and δ_{2n+1}^2 as δ_e^2 and δ_o^2, respectively, note that they correspond to the conditional variances in the multivariate normal case. □

PROOF OF THEOREM 2.2. First, (1.2) and the decomposition of A_{n+1}^{-1} in terms of A_n^{-1} of Theorem 2.1 yields (2.1). A similar decomposition of

$A_{n+1}^{-1}\, dA_{n+1}/d\alpha$ yields

$$\psi(Z_{n+1})^t A_{n+1}^{-1}\frac{dA_{n+1}}{d\alpha} Z_{n+1} = \psi(Z_n) A_n^{-1}\frac{dA_n}{d\alpha} Z_n + \psi(z_{n+1})\left[r_{n+1}\frac{dA_n}{d\alpha} + \delta_{n+1}^{-1}\frac{dl_{n+1}}{d\alpha}\right] Z_n + \psi(z_{n+1})\delta_{n+1}^{-1}\frac{d\delta_{n+1}}{d\alpha} z_{n+1}.$$

and $\mathrm{tr}(\Omega_{n+1}^{-1}\, d\Omega_{n+1}/d\alpha) = 2[\mathrm{tr}(A_n^{-1}\, dA_n/d\alpha) + \delta_{n+1}^{-1}\, d\delta_{n+1}/d\alpha]$. Elementary calculations reveal that

$$r_{n+1}\frac{dA_n}{d\alpha} + \delta_{n+1}^{-1}\frac{dl_{n+1}}{d\alpha} = \delta_{n+1}^{-1}\frac{da_{n+1}^t\,\Omega_n^{-1}}{d\alpha} A_n, \tag{4.1}$$

which is sufficient for (2.2). Next,

$$d\Omega_{n+1}/d\alpha = A_{n+1} dA_{n+1}^t/d\alpha + (dA_{n+1}/d\alpha)\, A_{n+1}^t,$$

which yields

$$A_{n+1}^{-1}\frac{d\Omega_{n+1}}{d\alpha} A_{n+1}^{-t} = \begin{pmatrix} A_n^{-1}\dfrac{d\Omega_n}{d\alpha} A_n^{-t} & \dfrac{dA_n^t}{d\alpha} r_{n+1}^t + \dfrac{dl_{n+1}^t}{d\alpha}\delta_{n+1}^{-1} \\ r_{n+1}\dfrac{dA_n}{d\alpha} + \delta_{n+1}^{-1}\dfrac{dl_{n+1}}{d\alpha} & 2\delta_{n+1}^{-1}\dfrac{d\delta_{n+1}}{d\alpha} \end{pmatrix} \tag{4.2}$$

and

$$\mathrm{tr}\left(\Omega_{n+1}^{-1}\frac{d\Omega_{n+1}}{d\alpha}\right) = \mathrm{tr}\left(\Omega_n^{-1}\frac{d\Omega_n}{d\alpha}\right) + 2\delta_{n+1}^{-1}\frac{d\delta_{n+1}}{d\alpha}. \tag{4.3}$$

Equations (4.1), (4.2) and (4.3) now yield (2.3). □

PROOF OF COROLLARY 2.2. First, $V_{\beta,\,n+1} = V_{\beta,\,n} + \psi(z_{n+1})(dz_{n+1}/d\beta)^t$, $dz_{2n}/d\beta = -\delta_e^{-1}(y_{2n} - \theta y_n)$ and $dz_{2n+1}/d\beta = -\delta_o^{-1}(y_{2n+1} - \Gamma_{12}^t\Gamma_1^{-1}(y_n^t, y_{2n}^t)^t)$, which is sufficient for the first two identities. We establish the latter two identities by exploiting the fact that our estimating functions are derived from the log-likelihood for the multivariate normal case. The "Markov" property of the bifurcating autoregressive process implies that the likelihood may be written as the product of conditional densities. Consider the contribution to the log-likelihood of x_{2n+1}, for example, with the contribution of x_{2n} being simpler. The contribution of x_{2n+1} is $-\ln(\delta_o) - z_{2n+1}^2/2$, and the derivative of this quantity with respect to the variance component α is $-[\delta_o^{-1}\, d\delta_o/d\alpha + z_{2n+1}\, dz_{2n+1}/d\alpha]$. Further, $dz_{2n+1}/d\alpha = -z_{2n+1}\delta_o^{-1}\, d\delta_o/d\alpha - z_{2n+1}\delta_o^{-1}(d(\Gamma_{12}^t\Gamma_1^{-1})/d\alpha)(x_2 - \mu_2)$, and an examination of Theorem 2.2 now yields the corollary. □

PROOF OF THEOREM A. Marschner (1991) only gave an outline of the proof of this result, so for completeness a fuller proof is given here. Let $\Delta > 0$ denote the smallest eigenvalue of $G(\eta_0)$. Choose $\varepsilon > 0$ so that $\Delta - 3\varepsilon > 0$ and then

$\delta > 0$ so that $H(\delta) < \varepsilon/2$. Then choose n so that, with probability larger than $1-\varepsilon$, $||n^{-1}V_n(\eta_0)|| < \varepsilon\delta$, $||G_n(\eta_0) - G(\eta_0)|| < \varepsilon$ and

$$\sup_{||\eta-\eta_0||<\delta} ||G_n(\eta) - G_n(\eta_0)|| < H(\delta) + \frac{\varepsilon}{2} < \varepsilon. \tag{4.4}$$

Thus, for $||\eta - \eta_0|| = \delta$ and any η^* such that $||\eta^* - \eta_0|| < \delta$,

$$\begin{aligned}
&(\eta-\eta_0)^t G_n(\eta^*)(\eta-\eta_0)\\
&\quad = (\eta-\eta_0)^t G(\eta_0)(\eta-\eta_0) + (\eta-\eta_0)^t (G_n(\eta_0) - G(\eta_0))(\eta-\eta^0)\\
&\qquad + (\eta-\eta_0)^t (G_n(\eta^*) - G_n(\eta_0))(\eta-\eta^0)\\
&\quad \geq \delta^2\Delta - \delta^2\varepsilon - \delta^2 \sup_{||\eta-\eta_0||<\delta} ||G_n(\eta) - G_n(\eta_0)||.
\end{aligned}$$

Hence, using (4.4), $(\eta-\eta_0)^t G_n(\eta^*)(\eta-\eta^0) > \delta^2(\Delta - 2\varepsilon) > 0$. Let $h_n(\eta) = n^{-1}V_n(\eta_0) + G_n(\eta^*)(\eta-\eta_0)$. Then $(\eta-\eta_0)h_n(\eta) > \delta^2(\Delta-3\varepsilon) > 0$. Hence we may apply Lemma 2 of Aitchison and Silvey (1958), as in Crowder (1976, 1986), to establish the existence of a sequence $\hat{\eta}_n$ of zeros of h_n. Now using the mean value theorem we may choose η^* so that $h_n(\eta) = n^{-1}V_n(\eta)$ and $\hat{\eta}_n$ is a zero of $n^{-1}V_n(\eta)$. Standard arguments now show that $P(|\hat{\eta}_n - \eta_0| < \delta) \to 1$ as required. Asymptotic normality is proven via the mean value theorem; that is, for η^* between $\hat{\eta}_n$ and η_0 we have $V_n(\hat{\eta}_n) = V_n(\eta_0) + G_n(\eta^*)(\hat{\eta}_n - \eta_0)$. In view of the above, we see that $||G_n(\eta^*) - G_n(\eta_0)|| \to_p 0$, and the result now follows from the convergence of $G_n(\hat{\eta}_n)$ and the central limit theorem for the estimating functions. $\square$

Acknowledgments. The author is grateful to R. G. Staudte for many helpful conversations concerning cell lineage trees and robust inference, to I. C. Marschner for conversations concerning asymptotic properties of estimators and to J. Zhang for producing Figure 1. The author is grateful to the Editor and referees for suggestions which greatly improved the presentation. The research was supported by an Australian Research Council Grant.

REFERENCES

AITCHISON, J. and SILVEY, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29** 813–828.

COWAN, R. and STAUDTE, R. G. (1986). The bifurcating autoregression model in cell lineage studies. *Biometrics* **42** 769–783.

CROWDER, M. J. (1976). Maximum likelihood estimation for dependent observations. *J. Roy. Statist. Soc. Ser. B* **38** 45–53.

CROWDER, M. J. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory* **2** 305–330.

HUGGINS, R. M. (1993a). On the robust analysis of variance components models for pedigree data. *Austral. J. Statist.* **35** 43–57.

HUGGINS, R. M. (1993b). A robust approach to the analysis of repeated measures. *Biometrics* **49** 715–720.

HUGGINS R. M. (1995). A law of large numbers for the bifurcating autoregressive process. *Comm. Statist. Stochastic Models* **11** 273–278.

HUGGINS, R. M. (1996). On the identifiability of measurement error in the bifurcating autoregressive model. *Statist. Probab. Lett.* **27** 17–23.

HUGGINS, R. M. and MARSCHNER, I. C. (1991). Robust analysis of the bifurcating autoregressive model in cell lineage studies. *Austral. J. Statist.* **33** 209–220.

HUGGINS, R. M. and STAUDTE, R. G. (1994). Variance components models for dependent cell populations. *J. Amer. Statist. Assoc.* **89** 19–29.

KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalisation. *Sankhyā Ser. A* **32** 419–430.

KINGMAN, J. F. C. (1972). On random sequences with spherical symmetry. *Biometrika* **59** 492–493.

KLIMKO, L A. and NELSON, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6** 629–642.

MARSCHNER, I. C. (1991). Robust estimation for epidemic models. *Austral. J. Statist.* **33** 221–240.

MILLER, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762.

POWELL, E. O. (1955). Some features of the generation times of individual bacteria. *Biometrika* **42** 16–44.

POWELL, E. O. (1956). An improved culture chamber for the study of living bacteria. *Journal of the Royal Microscopical Society* **75** 235.

POWELL, E. O. (1958). An outline of the pattern of bacterial generation times. *Journal of General Microbiology* **18** 382–417.

POWELL, E. O. and ERRINGTON, F. P. (1963). Generation times of individual bacteria: some corroborative measurements. *Journal of General Microbiology* **31** 315–327.

SHIRYAYEV, A. N. (1984). *Probability.* Springer, New York.

STAUDTE, R. G., GUIGET, M. and COLLYN D'HOOGE, M. (1984). Additive models for dependent cell populations. *Journal of Theoretical Biology* **109** 127–146.

STEWART, G. W. (1973). *Introduction to Matrix Computations.* Academic Press, New York.

TAYLOR, R. L., DAFFER, P. Z. and PATTERSON, R. F. (1985). *Limit Theorems for Sums of Exchangeable Random Variables.* Rowman and Allanheld, Totowa, NJ.

WEISS, L. (1973). Asymptotic properties of maximum likelihood estimators in some non-standard cases. II. *J. Amer. Statist. Assoc.* **68** 428–430.

WEISS, L. (1975). The asymptotic distribution of the likelihood ratio in some non-standard cases. *J. Amer. Statist. Assoc.* **70** 204–208.

DEPARTMENT OF STATISTICS
LA TROBE UNIVERSITY
BUNDOORA, 3083
AUSTRALIA