The Annals of Statistics
1990, Vol. 18, No. 4, 1757–1773

ESTIMATION OF THE MEANS IN THE BRANCHING PROCESS WITH IMMIGRATION

BY C. Z. WEI[1] AND J. WINNICKI

University of Maryland and Columbia University

Let $\{X_n\}$ be the branching process with immigration and let m and λ be the means of the offspring and immigration distributions, respectively. Estimation results for m and λ were obtained in the literature for the subcritical ($m < 1$) and supercritical ($m > 1$) cases, but no unified estimation procedure was developed, which would allow inference without knowledge of the range of m. The goal of this paper is to investigate this problem.

1. Introduction. In this paper we attempt to solve a long-standing estimation problem for the branching process with immigration, raised by Heyde and Seneta in 1974.

The branching process with immigration can be defined recursively by

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n,i} + I_n, \qquad n = 1, 2, \dots . \tag{1.1}$$

We can interpret X_n as the size of the n-th generation of a population, where $Y_{n,i}$ is the offspring size of the i-th individual in the $(n-1)$-st generation and I_n is the number of immigrants contributing to the population's n-th generation. Throughout this paper we will assume that $\{Y_{n,i}\}$ and $\{I_n\}$ are independent sequences of i.i.d., nonnegative, integer-valued random variables with finite means m and λ and finite variances σ^2 and b^2, respectively. We also assume that the initial value X_0 is a nonnegative, integer-valued, square-integrable random variable which is independent of $\{Y_{n,i}\}$ and $\{I_n\}$. [See Athreya and Ney (1972) for the basic properties of $\{X_n\}$.]

The study of estimation problems for the parameters m and λ on the basis of observing a single realization $\{X_0, \dots, X_n\}$ dates back to Smoluchowski (1916). Bartlett and Patankar [see Bartlett (1955)] began investigating this problem by the maximum likelihood approach. As noted by Heyde and Seneta (1972), the expressions for the maximum likelihood estimates of m in the parametric models considered were, in general, too complicated to be useful. [However, if the numbers of immigrants I_n, or even all offspring sizes $Y_{n,i}$ as well, are included in the sample, then the maximum likelihood approach yields useful results. These are given in Bhat and Adke (1981), Venkataraman (1982) and Venkataraman and Nanthi (1982).]

Received November 1987; revised October 1989.
[1]Research supported by NSF Grant DMS-86-03084.
AMS 1980 *subject classifications*. Primary 62M99; secondary 62M05.
Key words and phrases. Branching process with immigration, conditional least squares, weighted conditional least squares, inconsistency, limiting distribution.

The same authors [Heyde and Seneta (1972, 1974)] were the first to obtain estimation results for m and λ without imposing specific distribution assumptions on $\{Y_{n,i}\}$ and $\{I_n\}$.

For the case $m > 1$, they pointed out that the results of Heyde and Seneta (1971) and Heyde (1970) show that the ratio estimators X_n/X_{n-1} and $\sum_{i=1}^n X_i/\sum_{i=1}^n X_{i-1}$ can be used to estimate m.

For the case $m < 1$, they used a formal analogy between the branching process with immigration and the first order autoregressive process to derive strongly consistent and asymptotically normal estimators for m and λ. [Their moment assumptions were subsequently weakened by Quine (1976).] Their estimators are closely related and asymptotically equivalent to the conditional least squares estimators, first obtained by Klimko and Nelson (1978). Rewrite (1.1) as

$$X_n = mX_{n-1} + \lambda + \varepsilon_n, \tag{1.2}$$

where $\varepsilon_n = X_n - mX_{n-1} - \lambda$. Then ε_n is a martingale difference with respect to $\mathscr{F}_n$, where $\mathscr{F}_n = \sigma(X_0, X_1, \dots, X_n)$ and (1.2) is a stochastic regression equation. The conditional least squares estimators of m and λ resulting from (1.2) are

$$\hat{m}_n = \left[\sum_{i=1}^n X_i \sum_{i=1}^n X_{i-1} - n\sum_{i=1}^n X_i X_{i-1}\right] \Big/ \left\{\left[\sum_{i=1}^n X_{i-1}\right]^2 - n\sum_{i=1}^n X_{i-1}^2\right\}$$

and

$$\hat{\lambda}_n = \left[\sum_{i=1}^n X_{i-1}X_i \sum_{i=1}^n X_{i-1} - \sum_{i=1}^n X_{i-1}^2 \sum_{i=1}^n X_i\right] \Big/ \left\{\left[\sum_{i=1}^n X_{i-1}\right]^2 - n\sum_{i=1}^n X_{i-1}^2\right\}.$$

In the subcritical case ($m < 1$) the asymptotic properties of these estimators were extensively studied by Venkataraman (1982).

The previously mentioned results do not solve the problem of how to estimate m and λ if we do not know whether $m < 1$, $m = 1$ or $m > 1$ [cf. Heyde and Seneta (1974), pages 576 and 577].

In an attempt to obtain such a unified theory, Wei and Winnicki (1987) considered the estimators $\hat{m}_n$ and $\hat{\lambda}_n$ in the cases $m \geq 1$. Under the assumption that $\sigma^2 < \infty$ and $b^2 < \infty$, they showed the following:

1. If $m = 1$, then $\hat{m}_n \to_P m$ and

$$\begin{aligned} &\left(\sum_{i=1}^n (X_{i-1} + 1)\right)^{1/2} (\hat{m}_n - m) \\ &\quad \to_d \left\{\tfrac{1}{2}Y^2(1)\left(\int_0^1 Y(t)\,dt\right)^{1/2} - \left(Y(1) + \tfrac{1}{2}\sigma^2\right)\left(\int_0^1 Y(t)\,dt\right)^{3/2}\right\} \\ &\qquad \times \left\{\int_0^1 Y^2(t)\,dt - \left(\int_0^1 Y(t)\,dt\right)^2\right\}^{-1}, \end{aligned} \tag{1.3}$$

where $Y(t) = \lim_{n\to\infty} X_{[nt]}/n$ [weakly in $D[0,\infty)$] is a limiting diffusion process [Wei and Winnicki (1989)].

2. If $m > 1$, then $\hat{m}_n \to m$ a.s. and

$$\left(\sum_{i=1}^{n}(X_{i-1}+1)\right)^{1/2}(\hat{m}_n - m) \to_d N\left(0,(m+1)^2(m^2+m+1)^{-1}\sigma^2\right). \tag{1.4}$$

They also proved that $\hat{\lambda}_n$ is not a consistent estimator if $m \geq 1$.

This naturally leads to the question whether there exist other estimators which are always consistent.

In Section 2, based on the idea of the weighted conditional least squares [Nelson (1980)], we propose new estimators $\tilde{m}_n$ and $\tilde{\lambda}_n$. The estimator $\tilde{m}_n$ is shown to be consistent in all cases and to have a normal limit law if $m \neq 1$. If $m = 1$, then its asymptotic distribution is expressed in terms of the limiting process $Y(t)$ [cf. (1.3)]. In the case $m \neq 1$, the asymptotic variances of $\tilde{m}_n$ and $\hat{m}_n$ can be compared to show that $\tilde{m}_n$ is a more efficient estimator (cf. Section 4.5).

Even more convincing evidence for using the weighted conditional least squares estimators are the results for $\tilde{\lambda}_n$, which is shown to be consistent if $m < 1$ as well as $m = 1$. However, the limiting distribution of $\tilde{\lambda}_n$ is known only if $m < 1$ or $m = 1$ and $2\lambda > \sigma^2$.

In Section 3 we will show that the inconsistency of $\hat{\lambda}_n$ or $\tilde{\lambda}_n$ in the supercritical case is inevitable. More precisely, we will prove that there does not exist a consistent estimator for λ when $m > 1$.

2. The weighted conditional least squares estimators. Let us rewrite (1.2) as

$$\frac{X_n}{(X_{n-1}+1)^{1/2}} = m(X_{n-1}+1)^{1/2} + (\lambda - m)(X_{n-1}+1)^{-1/2} + \delta_n, \tag{2.1}$$

where $\delta_n = \varepsilon_n/(X_{n-1}+1)^{1/2}$.

Note that $E(\delta_n | \mathscr{F}_{n-1}) = 0$ and

$$E(\delta_n^2 | \mathscr{F}_{n-1}) = \frac{\sigma^2 X_{n-1} + b^2}{X_{n-1}+1}. \tag{2.2}$$

Since $E(\delta_n^2 | \mathscr{F}_{n-1}) \leq \sigma^2 + b^2$, the conditional variance of the "error" terms δ_n in the stochastic regression equation (2.1) would not fluctuate too much even when X_n is unbounded . It is also obvious that $E(\delta_n^2 | \mathscr{F}_{n-1}) \to \sigma^2$ a.s. as $X_n \to \infty$. Hence, when $\{X_n\}$ is transient, the error terms δ_n in δ (2.1) would be asymptotically homogeneous. Furthermore, from the definition of ε_n and the central limit theorem, it is not difficult to see that the asymptotic conditional distribution of δ_n given $\mathscr{F}_{n-1}$ is normal if $X_n \to \infty$. If δ_n were conditionally normal, then the least squares estimators based on (2.1) would be the maximum likelihood estimators. Some optimality properties in this case would then be expected.

These considerations lead us to study the least squares estimators based on (2.1), which we will call the weighted conditional least squares estimators,

$$\tilde{m}_n = \left\{ \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \frac{1}{X_{i-1}+1} - n \sum_{i=1}^{n} \frac{X_i}{X_{i-1}+1} \right\} \times \left\{ \sum_{i=1}^{n} (X_{i-1}+1) \sum_{i=1}^{n} \frac{1}{X_{i-1}+1} - n^2 \right\}^{-1}$$

and

$$\tilde{\lambda}_n = \left\{ \sum_{i=1}^{n} X_{i-1} \sum_{i=1}^{n} \frac{X_i}{X_{i-1}+1} - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} \frac{X_{i-1}}{X_{i-1}+1} \right\} \times \left\{ \sum_{i=1}^{n} (X_{i-1}+1) \sum_{i=1}^{n} \frac{1}{X_{i-1}+1} - n^2 \right\}^{-1}.$$

Here, $\tilde{\lambda}_n = \tilde{m}_n + \tilde{p}_n$, where $\tilde{p}_n$ is the least squares estimator of $p = \lambda - m$ in (2.1).

We will now investigate the asymptotic properties of $\tilde{m}_n$ and $\tilde{\lambda}_n$. We start with the subcritical case. It is well known that under the assumption that $m < 1$ and $E(\log^+ I_n) < \infty$, the process $\{X_n\}$ has a unique stationary distribution. If the distribution of X_0 is the stationary distribution, then $\{X_0\}$ is stationary and ergodic. By the coupling property of irreducible, aperiodic, positive recurrent Markov chains, we can assume that $\{X_n\}$ is stationary and ergodic, regardless of the distribution of X_0 [cf. Wei and Winnicki (1989), Remark 2.9].

THEOREM 2.1. *Assume that $m < 1$. Then $\tilde{m}_n$ and $\tilde{\lambda}_n$ are strongly consistent and*

$$\left(\left(\sum_{i=1}^{n} (X_{i-1}+1) \right)^{1/2} (\tilde{m}_n - m), \left(\sum_{i=1}^{n} (X_{i-1}+1)^{-1} \right)^{1/2} (\tilde{\lambda}_n - \lambda) \right)' \to_d N(\mathbf{0}, V^{-1} W V'^{-1}), \tag{2.3}$$

where

$$V = \begin{bmatrix} \dfrac{EX}{[E(X+1)]^{1/2}} & \left[E\dfrac{1}{X+1} \right]^{-1/2} \\ \dfrac{E[X/(X+1)]}{[E(X+1)]^{1/2}} & \left[E\dfrac{1}{X+1} \right]^{1/2} \end{bmatrix},$$

$$W = \begin{bmatrix} E(\sigma^2 X + b^2) & E\dfrac{\sigma^2 X + b^2}{X+1} \\ E\dfrac{\sigma^2 X + b^2}{X+1} & E\dfrac{\sigma^2 X + b^2}{(X+1)^2} \end{bmatrix}$$

and X is a random variable with the stationary distribution of the process $\{X_n\}$.

PROOF. The proof of strong consistency of $\tilde{m}_n$ and $\tilde{\lambda}_n$ is accomplished using the ergodic theorem, which implies that

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to EX = \frac{\lambda}{1-m} \quad \text{a.s.},$$

$$\text{(2.4)} \qquad \frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_{i-1}+1} \to E\frac{1}{X+1} \quad \text{a.s.},$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{X_{i-1}+1} \to E\frac{mX+\lambda}{X+1} = m + (\lambda - m)E\frac{1}{X+1} \quad \text{a.s.}$$

To prove the joint asymptotic normality, observe that

$$\text{(2.5)} \qquad \left(\left(\sum_{i=1}^{n}(X_{i-1}+1)\right)^{1/2}(\tilde{m}_n - m), \left(\sum_{i=1}^{n}(X_{i-1}+1)^{-1}\right)^{1/2}(\tilde{\lambda}_n - \lambda)\right)' = V_n^{-1}Z_n,$$

where

$$V_n = \begin{bmatrix} \left(\frac{1}{n}\sum_{i=1}^{n}(X_{i-1}+1)\right)^{-1/2}\left(\frac{1}{n}\sum_{i=1}^{n}X_{i-1}\right) & \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_{i-1}+1}\right)^{-1/2} \\ \left(\frac{1}{n}\sum_{i=1}^{n}(X_{i-1}+1)\right)^{-1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{X_{i-1}}{X_{i-1}+1}\right) & \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_{i-1}+1}\right)^{1/2} \end{bmatrix}$$

and

$$Z_n = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\varepsilon_i}{X_{i-1}+1} \end{pmatrix}.$$

As in (2.4), $V_n \to V$ a.s. Since by Jensen's inequality and nondegeneracy of X,

$$\det(V) = \left\{E(X+1)E\left[\frac{1}{X+1}\right]\right\}^{-1/2}\left\{E\left[\frac{1}{X+1}\right]E(X+1) - 1\right\} > 0,$$

V^{-1} exists. In order to show (2.3), it is now sufficient to prove that

$$Z_n \to_d N(\mathbf{0}, W).$$

By the Cramér–Wold theorem, we only have to show that for $\mathbf{c} = (c_1, c_2)' \in \mathbf{R}^2$ such that $\mathbf{c} \neq \mathbf{0}$,

$$\text{(2.6)} \qquad \mathbf{c}'Z_n \to_d N(\mathbf{0}, \mathbf{c}'W\mathbf{c}).$$

Since $\sum_{i=1}^{n}\{c_1\varepsilon_i + c_2\varepsilon_i/(X_{i-1}+1)\}$ is a martingale with stationary increments, (2.6) follows from the martingale central limit theorem [cf. Hall and Heyde

(1980), page 58] and the fact that

$$E\left(\left[c_1\varepsilon_i + \frac{c_2\varepsilon_i}{X_{i-1}+1}\right]^2\right) = \mathbf{c}'W\mathbf{c}. \qquad \square$$

THEOREM 2.2. *Assume that $m > 1$. Then $\tilde{m}_n \to m$ a.s., while $\tilde{\lambda}_n$ is not weakly consistent. Furthermore,*

$$\left(\sum_{i=1}^{n}(X_{i-1}+1)\right)^{1/2}(\tilde{m}_n - m) \to_d N(0,\sigma^2). \tag{2.7}$$

PROOF. It is well known that under the assumption that $m > 1$,

$$m^{-n}X_n \to L \text{ a.s.,} \quad \text{where } 0 < L < \infty. \tag{2.8}$$

So

$$\sum_{i=1}^{\infty}\frac{1}{X_{i-1}+1} < \infty \quad \text{a.s.,} \tag{2.9}$$

$$n^{-1}\sum_{i=1}^{n}\frac{X_i}{X_{i-1}+1} \to m \quad \text{a.s.} \tag{2.10}$$

and

$$m^{-n}\sum_{i=1}^{n}X_i \to (m-1)^{-1}mL \quad \text{a.s.,} \tag{2.11}$$

which easily yields $\tilde{m}_n \to m$ a.s.

Inconsistency of $\tilde{\lambda}_n$ is a simple corollary of Proposition 3.3.

Let us show (2.7). We have

$$\left(\sum_{i=1}^{n}(X_{i-1}+1)\right)^{1/2}(\tilde{m}_n - m) = \frac{A_n - B_n}{1 - C_n}, \tag{2.12}$$

where

$$A_n = \frac{\sum_{i=1}^{n}\varepsilon_i}{\left[\sum_{i=1}^{n}(X_{i-1}+1)\right]^{1/2}},$$

$$B_n = \left[n\sum_{i=1}^{n}\frac{\varepsilon_i}{X_{i-1}+1}\right]\Bigg/\left\{\sum_{i=1}^{n}\frac{1}{X_{i-1}+1}\left[\sum_{i=1}^{n}(X_{i-1}+1)\right]^{1/2}\right\},$$

and

$$C_n = n^2\left[\sum_{i=1}^{n}(X_{i-1}+1)\sum_{i=1}^{n}\frac{1}{X_{i-1}+1}\right]^{-1}.$$

Note that $C_n \to 0$ a.s. by (2.9) and (2.11).

Since $\varepsilon_n/(X_{n-1}+1)$ is a martingale difference with respect to $\mathscr{F}_n$ and

$$\sum_{i=1}^{\infty} E\left(\frac{\varepsilon_i^2}{(X_{i-1}+1)^2}\middle|\mathscr{F}_{i-1}\right) = \sum_{i=1}^{\infty} \frac{\sigma^2 X_{i-1}+b^2}{(X_{i-1}+1)^2} < \infty \quad \text{a.s.},$$

the martingale convergence theorem [cf. Hall and Heyde (1980), Theorem 2.17] implies that $\sum_{i=1}^n \varepsilon_i/(X_{i-1}+1)$ converges almost surely. Hence, using (2.11), we have that $B_n \to 0$ a.s. It remains to show that $A_n \to_d N(0,\sigma^2)$. But this follows from Theorem 3.5 of Wei and Winnicki (1989). □

REMARK 2.3. Using (2.12) it is not difficult to show that if $m > 1$, then

$$\tilde{m}_n - \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_{i-1}} = O\Big(n^2/\textstyle\sum_{i=1}^n (X_{i-1}+1)\Big) \quad \text{a.s.}$$

Thus, $\tilde{m}_n$ is asymptotically equivalent to the estimator $\sum_{i=1}^n X_i/\sum_{i=1}^n X_{i-1}$, which is known to be the nonparametric maximum likelihood estimator for m in the nonimmigration branching process [Feigin (1977) and Keiding and Lauritzen (1978)].

In order to investigate the critical case we need the following lemma:

LEMMA 2.4. *Assume that* $m = 1$. *Then*

$$\sum_{i=1}^n \frac{1}{X_{i-1}+1} \to \infty \quad a.s. \tag{2.13}$$

Furthermore, if $2\lambda > \sigma^2$ *and* $E|Y_{n,i}|^{2+\delta} < \infty$ *for some* $\delta > 0$, *then*

$$\sum_{i=1}^{\infty} \frac{1}{(X_{i-1}+1)^{1+\alpha}} < \infty \quad a.s. \tag{2.14}$$

for any $\alpha > 0$ *and*

$$(\log n)^{-1} \sum_{i=1}^n \frac{1}{X_{i-1}+1} \to_P (\lambda - \sigma^2/2)^{-1}. \tag{2.15}$$

The proof of Lemma 2.4 is in Wei and Winnicki [(1989), Theorem 2.12, Lemma 2.13 and Theorem 2.16].

THEOREM 2.5. *Assume that* $m = 1$. *Then* $\tilde{m}_n \to_P m$, $\tilde{\lambda}_n \to_P \lambda$ *and*

$$\left(\sum_{i=1}^n (X_{i-1}+1)\right)^{1/2} (\tilde{m}_n - m) \to_d (Y(1)-\lambda)\Big/\left(\int_0^1 Y(t)\,dt\right)^{1/2}, \tag{2.16}$$

where the process $Y(t)$ *is defined after* (1.3). *Furthermore, under the additional assumptions that* $2\lambda > \sigma^2$, $E|Y_{n,i}|^{2+\delta} < \infty$ *and* $E|I_n|^{2+\delta} < \infty$ *for some* $\delta > 0$, *we have that*

$$\left(\sum_{i=1}^n (X_{i-1}+1)^{-1}\right)^{1/2} (\tilde{\lambda}_n - \lambda) \to_d N(0,\sigma^2). \tag{2.17}$$

PROOF. First, we claim that

$$\sum_{i=1}^{n} \frac{\varepsilon_i}{X_{i-1}+1} = o\left(\sum_{i=1}^{n} \frac{1}{X_{i-1}+1}\right) \quad \text{a.s.} \tag{2.18}$$

We know that $\sum_{i=1}^{n} \varepsilon_i/(X_{i-1}+1)$ is a martingale. Notice also that

$$\begin{aligned}
&\sum_{i=2}^{\infty} \left(\sum_{j=1}^{i-1} \frac{1}{X_{j-1}+1}\right)^{-2} E\left[\left(\frac{\varepsilon_i}{X_{i-1}+1}\right)^2 \middle| \mathscr{F}_{i-1}\right] \\
&\quad = \sum_{i=2}^{\infty} \left(\sum_{j=1}^{i-1} \frac{1}{X_{j-1}+1}\right)^{-2} \frac{\sigma^2 X_{i-1}+b^2}{(X_{i-1}+1)^2} \\
&\quad \le (\sigma^2+b^2) \sum_{i=2}^{\infty} \left(\frac{1}{X_{i-1}+1}\right)\left(\sum_{j=1}^{i-1} \frac{1}{X_{j-1}+1}\right)^{-2} < \infty \quad \text{a.s.},
\end{aligned}$$

where convergence follows from (2.13). Hence, (2.18) follows by an application of the strong law for martingales [Hall and Heyde (1980), Theorem 2.18].

Now we write

$$\tilde{\lambda}_n - \lambda = \frac{D_n - E_n}{1 - C_n}, \tag{2.19}$$

where C_n is defined after (2.12),

$$D_n = \left[\frac{\sum_{i=1}^{n} X_{i-1}}{\sum_{i=1}^{n}(X_{i-1}+1)}\right]\left[\sum_{i=1}^{n} \frac{1}{X_{i-1}+1}\right]^{-1} \sum_{i=1}^{n} \frac{\varepsilon_i}{X_{i-1}+1}$$

and

$$E_n = \left[\sum_{i=1}^{n} \frac{X_{i-1}}{X_{i-1}+1}\right]\left[\sum_{i=1}^{n}(X_{i-1}+1)\sum_{i=1}^{n} \frac{1}{X_{i-1}+1}\right]^{-1} \sum_{i=1}^{n} \varepsilon_i.$$

By (2.18) and the fact that $\sum_{i=1}^{n} X_{i-1}/\sum_{i=1}^{n}(X_{i-1}+1) \le 1$,

$$D_n \to 0 \quad \text{a.s.} \tag{2.20}$$

By Corollary 2.3 (cf. Remark 2.4) of Wei and Winnicki (1989) we know that

$$\left(n^{-1}X_n, n^{-2}\sum_{i=1}^{n}(X_{i-1}+1)\right) \to_d \left(Y(1), \int_0^1 Y(t)\,dt\right). \tag{2.21}$$

Since

$$\int_0^1 Y(t)\,dt > 0 \quad \text{a.s.}, \tag{2.22}$$

(2.13) and (2.21) imply

$$C_n \to_P 0. \tag{2.23}$$

Further,

$$\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i = \frac{1}{n}(X_n - X_0 - n\lambda) \to_d Y(1) - \lambda. \tag{2.24}$$

By (2.21), (2.13), (2.24) and the fact that $\sum_{i=1}^{n} X_{i-1}/(X_{i-1}+1) \le n$,

$$E_n \to_P 0. \tag{2.25}$$

Now (2.20), (2.25) and (2.23) give $\tilde{\lambda}_n \to_P \lambda$. To prove (2.16) we will use (2.12). By (2.21) and (2.24),

$$A_n \to_d (Y(1) - \lambda) \Big/ \left(\int_0^1 Y(t)\, dt \right)^{1/2}. \tag{2.26}$$

In view of (2.18), (2.21) and (2.22),

$$B_n \to_P 0. \tag{2.27}$$

Consequently, (2.16) follows from (2.25), (2.27) and (2.23). Note that (2.16) implies that $\tilde{m}_n \to_P m$. Finally, let us show (2.17). By (2.19) and (2.23), it suffices to prove that

$$\left(\sum_{i=1}^{n} \frac{1}{X_{i-1}+1} \right)^{1/2} D_n \to_d N(0, \sigma^2) \tag{2.28}$$

and

$$\left(\sum_{i=1}^{n} \frac{1}{X_{i-1}+1} \right)^{1/2} E_n \to_P 0. \tag{2.29}$$

Relation (2.29) follows easily from (2.15), (2.24), (2.21) and (2.22). For (2.28), observe that (2.21) and (2.22) imply that

$$\frac{\sum_{i=1}^{n} X_{i-1}}{\sum_{i=1}^{n} (X_{i-1}+1)} \to_P 1.$$

Thus, by (2.15) it is sufficient to show that

$$(\log n)^{-1/2} \sum_{i=1}^{n} \frac{\varepsilon_i}{X_{i-1}+1} \to_d N\left(0, \sigma^2 \left(\lambda - \frac{\sigma^2}{2}\right)^{-1}\right). \tag{2.30}$$

We will apply the martingale central limit theorem [Hall and Heyde (1980), Corollary 3.1]. For the conditional variance, by (2.14) and (2.15),

$$\begin{aligned}
&(\log n)^{-1} \sum_{i=1}^{n} E\left[\frac{\varepsilon_i^2}{(X_{i-1}+1)^2} \middle| \mathscr{F}_{i-1} \right] \\
&\quad = (\log n)^{-1} \sum_{i=1}^{n} \frac{\sigma^2 X_{i-1} + b^2}{(X_{i-1}+1)^2} \\
&\quad = (\log n)^{-1} \left\{ \sigma^2 \sum_{i=1}^{n} \frac{1}{X_{i-1}+1} + (b^2 - \sigma^2) \sum_{i=1}^{n} \frac{1}{(X_{i-1}+1)^2} \right\} \\
&\quad \to_P \sigma^2 \left(\lambda - \frac{\sigma^2}{2}\right)^{-1}.
\end{aligned} \tag{2.31}$$

For the conditional Lyapunov's condition, by Lemma 2.1 of Lai and Wei (1983), there is a constant C_δ such that

$$(\log n)^{-1-\delta/2} \sum_{i=1}^{n} E\left\{\left[\frac{|\varepsilon_i|}{X_{i-1}+1}\right]^{2+\delta}\middle|\mathscr{F}_{i-1}\right\}$$

$$\le (\log n)^{-1-\delta/2} C_\delta \left\{\sum_{i=1}^{n} \frac{\left(E\left[\varepsilon_i^2|\mathscr{F}_{i-1}\right]\right)^{1+\delta/2}}{(X_{i-1}+1)^{2+\delta}}\right\} \tag{2.32}$$

$$= (\log n)^{-1-\delta/2} C_\delta \left\{\sum_{i=1}^{n} \frac{\left(\sigma^2 X_{i-1}+b^2\right)^{1+\delta/2}}{(X_{i-1}+1)^{2+\delta}}\right\} \to 0 \quad \text{a.s.},$$

where convergence follows from (2.14).

Relations (2.31) and (2.32) show that the conditions of the martingale central limit theorem are satisfied and the proof of Theorem 2.5 is complete. □

REMARK 2.6. The condition $2\lambda > \sigma^2$, required to establish asymptotic normality of $\tilde{\lambda}_n$ in Theorem 2.5, implies that the process $\{X_n\}$ is transient. The remaining case $2\lambda \le \sigma^2$ is the case when $\{X_n\}$ is null-recurrent. This dichotomy for the critical branching process with immigration was discovered by Pakes (1971). It is not known what is the limiting distribution of $\tilde{\lambda}_n$ when $m = 1$ and $2\lambda \le \sigma^2$.

3. Nonexistence of consistent estimators in the supercritical case.

In this section we will show that except for the offspring distribution's mean and variance, no parameters of the supercritical branching process with immigration have consistent estimators.

We will first establish a necessary condition for existence of consistent estimators. Let $\{X_0, X_1, \dots\}$ be a time-homogeneous, denumerable Markov chain whose probability measure P is determined by an initial distribution ρ and a transition function p. We will identify $P = (\rho, p)$. Suppose that $P \in \mathscr{P}$, where $\mathscr{P}$ is a family of probability measures of time-homogeneous, denumerable Markov chains.

For any P and $Q = (\delta, q)$ in $\mathscr{P}$ define the likelihood ratio

$$\Lambda_n(P,Q) = \frac{\delta(X_0)\prod_{k=0}^{n-1} q(X_{k+1}|X_k)}{\rho(X_0)\prod_{k=0}^{n-1} p(X_{k+1}|X_k)} \quad \text{a.s.-}P.$$

A parameter is a function $\theta: \mathscr{P} \to \Theta$, where Θ is a metric space with metric d. Let $\mathscr{F}_n = \sigma(X_0, \dots, X_n)$. A sequence $\{\hat{\theta}_n\}$ of $\mathscr{F}_n$-measurable, Θ-valued functions is said to be a consistent estimator for θ if $d(\hat{\theta}_n, \theta(P)) \to_P 0$ for all $P \in \mathscr{P}$.

Clearly, if there exists a consistent estimator for θ, then for any P, Q such that $\theta(P) \ne \theta(Q)$, $P \perp Q$. It follows by the properties of the likelihood ratio [cf. Chow, Robbins and Siegmund (1972), pages 12 and 19] that

$$\Lambda_n(P,Q) \to 0 \quad \text{a.s.-}P. \tag{3.1}$$

Relation (3.1) provides a criterion for proving nonexistence of a consistent estimator.

We will be concerned with the case when the parameter θ depends only on the transition function p, i.e., for any initial distributions ρ and δ,

$$\theta(\rho, p) = \theta(\delta, p) \equiv \theta(p). \tag{3.2}$$

Suppose that there exists an estimator $\hat{\theta}_n$ which is consistent for any initial distribution. Then (3.1) implies

$$\prod_{k=n_0}^{n_0+n} \frac{q(X_{k+1}|X_k)}{p(X_{k+1}|X_k)} \to_{n\to\infty} 0 \quad \text{a.s.-}(\rho, p) \tag{3.3}$$

for any integer $n_0 \geq 0$ and p, q such that $\theta(p) \neq \theta(q)$.

Criterion (3.3) can be applied if we know the asymptotic behavior of the transition functions p and q. We will now develop a relevant asymptotic result for the case when $\{X_n\}$ is the branching process with immigration.

Recall that the span of the distribution of an $\mathbf{N}_0$-valued random variable X is the largest positive integer h such that for some $u \in \{0, \dots, h-1\}$, $X = u$ (mod h) a.s. Then u is called the offset of the distribution of X. We also denote the support of X, $\{k: P(X = k) > 0\}$, by supp(X). Subscripts supp_P, etc., will denote the corresponding quantities under the measure P.

Given nonnegative numbers m and σ^2, let $\mathscr{C} = \mathscr{C}(m, \sigma^2)$ denote the class of all transition functions of branching processes with immigration such that the offspring distribution has mean m, variance σ^2 and a finite third moment $E(Y_{n,i}^3)$ and the immigration distribution has a finite mean $E(I_n)$. For simplicity, we will also assume that the offspring distribution has offset 0 and span 1. Notice that a transition function of a branching process with immigration can be identified with a pair of offspring and immigration distributions.

For arbitrary initial distributions ρ and δ, let $P = (\rho, p)$ and $Q = (\delta, q)$, where $p, q \in \mathscr{C}$.

Lemma 3.1. *Assume that $m > 1$, $\sigma^2 < \infty$. If*

$$\mathrm{supp}_P(X_n) \subseteq \mathrm{supp}_Q(X_n) \tag{3.4}$$

for all sufficiently large n, then

$$q(X_{n+1}|X_n) = \frac{1}{(2\pi\sigma^2 X_n)^{1/2}} \exp\left[-\frac{(X_{n+1} - mX_n)^2}{2(\sigma\sqrt{X_n})^2}\right]\left[1 + O\left(\frac{n}{m^{n/2}}\right)\right] \quad a.s.\text{-}P. \tag{3.5}$$

Proof. For any $k, l \in \mathbf{N}_0$,

$$q(k|l) = Q\left(\sum_{j=1}^{l} Y_{1,j} + I_1 = k\right) = \sum_{i=0}^{k} Q\left(\sum_{j=1}^{l} Y_{1,j} = k - i\right) Q(I_1 = i).$$

By the local limit theorem [Petrov (1975)],

$$Q\left(\sum_{j=1}^{l} Y_{1,j} = i\right) = \frac{1}{\sigma\sqrt{l}}\varphi(x_i) + O\left(\frac{1}{l}\right),$$

where $\varphi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, $x_i = (i - ml)/\sigma\sqrt{l}$, $|O(1/l)| \le c_1/l$ and c_1 does not depend on l and i. Hence,

$$q(k|l) = \sum_{i=0}^{k} \frac{1}{\sigma\sqrt{l}}\varphi(x_{k-i})Q(I_1 = i) + O\left(\frac{1}{l}\right).$$

By the mean value theorem

$$\left|\varphi(x_{k-i}) - \varphi\left(\frac{k - ml}{\sigma\sqrt{l}}\right)\right| \le c_2\frac{i}{\sigma\sqrt{l}}, \qquad i = 0,\dots,k,$$

where $c_2 = \sup_{-\infty<x<\infty}|\varphi'(x)| < \infty$. Since $E_Q(I_1) = \lambda_Q < \infty$, it follows that

$$\sum_{i=0}^{k}\left|\varphi(x_{k-i}) - \varphi\left(\frac{k-ml}{\sigma\sqrt{l}}\right)\right|Q(I_1 = i) \le \frac{c_2}{\sigma\sqrt{l}}\sum_{i=0}^{\infty} iQ(I_1 = i) = O\left(\frac{1}{\sqrt{l}}\right). \tag{3.6}$$

Also, $\sum_{i=k+1}^{\infty} Q(I_1 = i) = o(1/k)$, so

$$q(k|l) = \frac{1}{\sigma\sqrt{l}}\varphi\left(\frac{k-ml}{\sigma\sqrt{l}}\right) + o\left(\frac{1}{k}\right) + O\left(\frac{1}{l}\right).$$

Hence, using (3.4)

$$q(X_{n+1}|X_n) = \frac{1}{\sigma\sqrt{X_n}}\varphi\left(\frac{X_{n+1} - mX_n}{\sigma\sqrt{X_k}}\right) + o\left(\frac{1}{X_{n+1}}\right) + O\left(\frac{1}{X_n}\right) \quad \text{a.s.-}P.$$

But by (2.8), $X_{n+1} = O(X_n)$ a.s.-P, so

$$q(X_{n+1}|X_n) = \frac{1}{\sqrt{2\pi\sigma^2 X_n}}\exp\left[-\frac{(X_{n+1} - mX_n)^2}{2\left(\sigma\sqrt{X_n}\right)^2}\right] + O\left(\frac{1}{X_n}\right) \quad \text{a.s.-}P. \tag{3.7}$$

Finally, by an analogue of the law of the iterated logarithm for $\{X_n\}$ [Heyde and Leslie (1971)]

$$\exp\left[\frac{(X_{n+1} - mX_n)^2}{2\left(\sigma\sqrt{X_n}\right)^2}\right] = O(e^{\log n}) = O(n) \quad \text{a.s.-}P. \tag{3.8}$$

Applying (3.8) and (2.8) to (3.7) we obtain (3.5). □

COROLLARY 3.2. *Under the assumptions of Lemma* 3.1, *there exists an integer-valued random variable N such that*

$$\lim_{n\to\infty} \prod_{k=N}^{N+n} \frac{q(X_{k+1}|X_k)}{p(X_{k+1}|X_k)} > 0 \quad a.s.\text{-}P. \tag{3.9}$$

PROOF. Note that Lemma 3.1 holds also if $Q = P$. This implies that

$$N = \inf\{n: p(X_{k+1}|X_k)q(X_{k+1}|X_k) > 0 \text{ for all } k \geq n\}$$

is finite a.s.-P.

Now consider the product

$$\prod_{k=N}^{N+n} \frac{q(X_{k+1}|X_k)}{p(X_{k+1}|X_k)}.$$

By Lemma 3.1 it is enough to show that

$$\lim_{n\to\infty} \prod_{k=N}^{N+n} \frac{1 - O_Q(nm^{-n/2})}{1 - O_P(nm^{-n/2})} > 0 \quad \text{a.s.-}P. \tag{3.10}$$

[Notice that the limit in (3.9) may be infinity.] But (3.10) clearly holds since

$$\sum_{k=N}^{\infty} \left| \frac{1 - O_Q(nm^{-n/2})}{1 - O_P(nm^{-n/2})} - 1 \right| < \infty \quad \text{a.s.-}P. \qquad \square$$

Criterion (3.2) together with Corollary 3.2 yield the following proposition.

PROPOSITION 3.3. *Let θ be a parameter defined on a class $\mathscr{P}$ of transition functions of the branching process with immigration. If for some $m > 1$, $\sigma^2 < \infty$, θ takes at least two values on $\mathscr{P} \cap \mathscr{C}(m, \sigma^2)$, then there is no consistent estimator of θ.*

EXAMPLE 3.4. Consider the class of branching processes with immigration whose offspring distribution is Poisson with parameter m and the immigration distribution is Poisson with parameter λ. By Proposition 3.3, if $m > 1$, then there is no consistent estimator of λ. Notice that in this example all parameters are identifiable. Hence, as opposed to the classical situation of i.i.d. observations, nonexistence of consistent estimators is not due to nonidentifiability. The essential problem is that for certain ranges of parameters the probability measures of the process are not mutually singular for distinct parameter values.

REMARK 3.5. Consider the nonparametric situation assuming that $\mathscr{P}$ is the class of all transition functions of supercritical branching process with immigration whose offspring and immigration distributions have finite second moments. It is well known that the parameters m and σ^2 admit consistent estimators [Heyde (1974)]. Since the parametric class of Example 3.4 is contained in $\mathscr{P}$, we conclude that λ does not have a consistent estimator in $\mathscr{P}$.

REMARK 3.6. In this section we investigated the problem of nonexistence of consistent estimators by examining the asymptotic behavior of the likelihood ratio. Lockhart (1982) established nonexistence of consistent estimators for parameters of the nonimmigration branching process. His method used the total variation distance between the probability measures of the process to prove that they are not mutually singular. Our results can be viewed as a generalization of his, since they remain valid when the immigration distribution is concentrated at zero, provided that the "a.s." statements are understood to hold "on the set $\{X_n \to \infty\}$." Conversely, using Lemma 3.1 we could use Lockhart's approach to prove our results. An advantage of the approach of the present paper is that it allows a generalization to the critical case [Winnicki (1990)].

4. Concluding remarks.

4.1. The limiting distributions of the estimators for m and λ discussed in this paper depend on the parameters σ^2 and b^2. Hence, to use these results in practice would require estimating the variances. A unified estimation theory for the variances, analogous to the one for the means presented in this paper, has been developed by Winnicki (1990) and we refer the reader to his paper for detailed results.

4.2. Of particular interest, also considering the historical development of the subject, would be statistically testing the $m < 1$, $m = 1$, $m > 1$ trichotomy. In principle, we can construct such tests using our asymptotic distribution results for the estimators of m. However, the calculation of the critical points is complicated. Since the limiting distributions depend on the unknown parameters they would have to be estimated. The variance of the limiting distribution of $\tilde{m}_n$ in the subcritical case can be written as

$$\begin{aligned}\sigma_m^2 = {} & \frac{\sigma^2\big[(E(X+1)E[1/(X+1)])^2 - E(X+1)E[1/(X+1)]\big]}{[E(X+1)E[1/(X+1)] - 1]^2} \\ & + \frac{(b^2-\sigma^2)E(X+1)E\big(1/(X+1)^2\big)^2}{[E(X+1)E[1/(X+1)] - 1]^2} \\ & + \frac{-(b^2-\sigma^2)E(X+1)\big(E\big[1/(X+1)^2\big]\big)^2}{[E(X+1)E[1/(X+1)] - 1]^2}\end{aligned} \tag{4.1}$$

(cf. Theorem 2.1). Suppose that σ^2 and b^2 are known or have been estimated (cf. Section 4.1). Since $m < 1$, the ergodic properties of $\{X_n\}$ imply that

$$\hat{\sigma}^2_{m,n} = \frac{\sigma^2(\Sigma_{i=1}^n (X_{i-1}+1)\Sigma_{i=1}^n[1/(X_{i-1}+1)])^2}{[\Sigma_{i=1}^n (X_{i-1}+1)\Sigma_{i=1}^n[1/(X_{i-1}+1)] - n^2]^2}$$

$$+ \frac{-\sigma^2 n^2 \Sigma_{i=1}^n (X_{i-1}+1)\Sigma_{i=1}^n[1/(X_{i-1}+1)]}{[\Sigma_{i=1}^n (X_{i-1}+1)\Sigma_{i=1}^n[1/(X_{i-1}+1)] - n^2]^2}$$

$$+ \frac{(b^2-\sigma^2) n^2 \Sigma_{i=1}^n (X_{i-1}+1)\Sigma_{i=1}^n[1/(X_{i-1}+1)]}{[\Sigma_{i=1}^n (X_{i-1}+1)\Sigma_{i=1}^n[1/(X_{i-1}+1)] - n^2]^2}$$

$$+ \frac{-(b^2-\sigma^2) n \Sigma_{i=1}^n (X_{i-1}+1)(\Sigma_{i=1}^n[1/(X_{i-1}+1)])^2}{[\Sigma_{i=1}^n (X_{i-1}+1)\Sigma_{i=1}^n[1/(X_{i-1}+1)] - n^2]^2}$$

is a (strongly) consistent estimator for σ_m^2. It also follows from (2.9) and (2.11) that $\hat{\sigma}^2_{m,n} \to \sigma_m^2$ a.s. if $m > 1$. Hence, if we are testing the hypothesis $m = 1$ versus the alternative $m \neq 1$, then under the alternative the error probabilities can be estimated. However, under the hypothesis $m = 1$, the limiting distribution $(Y(1) - \lambda)/\int_0^1 Y(t)\,dt)^{1/2}$ is a complicated functional of the diffusion process $\{Y(t)\}$ and the only approach we can suggest is numerical tabulation. Notice that the distribution of the process $\{Y(t)\}$ depends only on the parameters λ and σ^2 which can be estimated. A possible clue to obtaining analytical results is the fact that the joint Laplace transform of $Y(1)$ and $\int_0^1 Y(t)\,dt$ is known to be

$$E\left(\exp\left(-sY(1) - z\int_0^1 Y(t)\,dt\right)\right)$$
$$= \left[s\sqrt{\frac{2}{\sigma^2 z}}\sinh\left(\sqrt{\frac{\sigma^2 z}{2}}\right) + \cosh\left(\sqrt{\frac{\sigma^2 z}{2}}\right)\right]^{-2\lambda/\sigma^2},$$

where $s \geq 0$, $z \geq 0$ [Mellein (1983)].

4.3. Notice that if m is far from 1, then it will be obvious from the behavior of the process whether we are in the recurrent or the explosive case. Thus, of greatest interest are the questions of inference when $m = 1$ or m is close to 1. This, in particular, because of the nonnormal limit law when $m = 1$, which suggests a nonnormal approximation in the near-critical case. A similar problem has been studied by Chen and Wei (1987) for the first-order autoregressive time series, but analogous results for the branching process with immigration have yet to be developed.

4.4. If $m > 1$, the asymptotic variance of $\tilde{m}_n$ equals σ^2, which is smaller than the asymptotic variance of $\hat{m}_n$ [cf. Theorem 2.2 and (1.4)]. If $m < 1$, it is possible to compare the asymptotic variances of $\tilde{m}_n$ and $\hat{m}_n$ for m approaching 1. We will restrict attention to the "continuously subcritical" class of offspring distributions considered in Quine and Seneta (1969), with the immigration distribution fixed. In addition, we will suppose that $2\lambda \le \sigma^2$, where σ^2 is the variance of the offspring distribution as $m \uparrow 1$. Then, using (4.1) and the theorem in Quine and Seneta (1969), it is not difficult to show that $\sigma_m^2 \to \sigma^2$ as $m \uparrow 1$. On the other hand, if $m < 1$, $\gamma = E(Y_{n,i} - m)^3 < \infty$ and $d = E(I_n - \lambda)^3 < \infty$, then

$$\left(\sum_{i=1}^{n} (X_{i-1} + 1)\right)^{1/2} (\hat{m}_n - m) \to_d N(0, \Sigma_m),$$

where

$$\Sigma_m = \left(\frac{\lambda}{1-m} + 1\right)\Bigg[\sigma^2\left(\frac{\lambda\gamma}{1-m} + d\right)(1-m^2)^2\left(\frac{\lambda\sigma^2}{1-m} + b^2\right)^{-2}(1-m^3)^{-1}$$
$$+1 - m^2 + 3m\sigma^4(1-m^2)\left(\frac{\lambda\sigma^2}{1-m} + b^2\right)^{-1}(1-m^3)^{-1}\Bigg],$$

[cf. Wei and Winnicki (1987)]. A simple calculation shows that $\Sigma_m \to 2\sigma^2 + 2\lambda$ as $m \uparrow 1$. Hence, also in the case $m < 1$, $m \to 1$ the asymptotic variance of $\tilde{m}_n$ is smaller than the asymptotic variance of $\hat{m}_n$.

4.5. Although our estimators $\tilde{m}_n$ and $\tilde{\lambda}_n$ perform better than $\hat{m}_n$ and $\hat{\lambda}_n$, nothing is known about their efficiency in general. For a possible approach to this question using estimating equations, see Godambe and Heyde (1987) and Heyde (1987). Another possible approach to the question of optimality is to use the asymptotic efficiency concepts as outlined in Hall and Heyde [(1980), Chapter 6]. It would also be of interest to compare our estimators with the estimators obtained based on observing all immigration sizes and/or all offspring sizes (e.g., maximum likelihood estimators). Related questions of loss of information with an application to the nonimmigration branching process are treated in Le Cam and Yang (1988).

REFERENCES

ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. Springer, New York.

BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge Univ. Press, Cambridge.

BHAT, B. R. and ADKE, S. R. (1981). Maximum likelihood estimation for branching process with immigration. *Adv. in Appl. Probab.* **13** 498–509.

CHAN, N. H. and WEI, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.* **15** 1050–1063.

CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1972). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston.

FEIGIN, P. D. (1977). A note on maximum likelihood estimation for simple branching processes. *Austral. J. Statist.* **19** 152–154.

GODAMBE, V. P. and HEYDE, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55** 231–244.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Wiley, New York.
HEYDE, C. C. (1970). Extension of a result of Seneta for the supercritical Galton–Watson process. *Ann. Math. Statist.* **41** 739–742.
HEYDE, C. C. (1974). On estimating the variance of the offspring distribution in a simple branching process. *Adv. in Appl. Probab.* **6** 421–433.
HEYDE, C. C. (1987). On combining quasi-likelihood estimating functions. *Stochastic Process. Appl.* **25** 281–288.
HEYDE, C. C. and LESLIE, J. R. (1971). Improved classical limit analogues for Galton–Watson process with or without immigration. *Bull. Austral. Math. Soc.* **5** 145–155.
HEYDE, C. C. and SENETA, E. (1971). Analogues of classical limit theorems for the supercritical Galton–Watson process with immigration. *Math. Biosci.* **11** 249–259.
HEYDE, C. C and SENETA, E. (1972). Estimation theory for growth and immigration rates in a multiplicative process. *J. Appl. Probab.* **9** 235–258.
HEYDE, C. C. and SENETA, E. (1974). Notes on "Estimation theory for growth and immigration rates in a multiplicative process." *J. Appl. Probab.* **11** 572–577.
KEIDING, N. and LAURITZEN, S. F. (1978). Marginal maximum likelihood estimates and estimation of the offspring mean in a branching process. *Scand. J. Statist.* **5** 106–110.
KLIMKO, L. A. and NELSON, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6** 629–642.
LAI, T. L. and WEI, C. Z. (1983). Lacunary and generalized linear processes. *Stochastic Process. Appl.* **14** 187–199.
LE CAM, L. and YANG, G. L. (1988). Distinguished statistics, loss of information and a theorem of Robert B. Davies. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **2** 163–175. Springer, New York.
LOCKHART, R. (1982). On nonexistence of consistent estimates in Galton–Watson process. *J. Appl. Probab.* **19** 842–846.
MELLEIN, B. (1983). Kac functionals of diffusion processes approximating critical branching processes. In *Stochastic Processes Applied to Physics and Other Related Fields* (B. Gómez, S. M. Moore, A. M. Rodriguez-Vargas and A. Reuda, eds.) 184–198. World Science, Singapore.
NELSON, P. I. (1980). A note on strong consistency of least squares estimators in regression models with martingale difference errors. *Ann. Statist.* **8** 1057–1064.
PAKES, A. G. (1971). On the critical Galton–Watson process with immigration. *J. Austral. Math. Soc.* **12** 476–482.
PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York.
QUINE, M. P. (1976). Asymptotic results for estimators in a subcritical branching process with immigration. *Ann. Probab.* **4** 319–325. [Correction note (1977) **5** 318.]
QUINE, M. P. and SENETA, E. (1969). A limit theorem for the Galton–Watson process with immigration. *Austral. J. Statist.* **11** 166–173.
SMOLUCHOWSKI, M. (1916). Drei Vorträge über Diffusion, Brownsche Bewegung und Wagulation von Kolloidteilchen. *Physik. Z.* **17** 557–585.
VENKATARAMAN, K. N. (1982). A time series approach to the study of the simple subcritical Galton–Watson process with immigration. *Adv. in. Appl. Probab.* **14** 1–20.
VENKATARAMAN, K. N. and NANTHI, K. (1982). A limit theorem on a subcritical Galton–Watson process with immigration. *Ann. Probab.* **10** 1069–1074.
WEI, C. Z. and WINNICKI, J. (1987). A unified estimation theory for the branching process with immigration. Technical Report, Univ. Maryland.
WEI, C. Z. and WINNICKI, J. (1989). Some asymptotic results for the branching process with immigration. *Stochastic Process. Appl.* **31** 261–282.
WINNICKI, J. (1990). Estimation of the variances in the branching process with immigration. *Probab. Theory Related Fields*. To appear.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742

DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027