Ann. Inst. Statist. Math.
Vol. 48, No. 1, 75–88 (1996)

ASYMPTOTICS AND BOOTSTRAP FOR INVERSE GAUSSIAN REGRESSION

GUTTI JOGESH BABU[1*] AND YOGENDRA P. CHAUBEY[2**]

[1] *Department of Statistics, 319 Classroom Building, Pennsylvania State University, University Park, PA 16802, U.S.A.*
[2] *Department of Mathematics and Statistics, Concordia University, Loyola Campus, 7141 Sherbrooks Street West, Montreal, Quebec, Canada H4B 1R6*

(Received January 13, 1995; revised May 24, 1995)

Abstract. This paper studies regression, where the reciprocal of the mean of a dependent variable is considered to be a linear function of the regressor variables, and the observations on the dependent variable are assumed to have an inverse Gaussian distribution. The large sample theory for the *pseudo maximum likelihood* estimators is available in the literature, only when the number of replications increase at a fixed rate. This is inadequate for many practical applications. This paper establishes consistency and derives the asymptotic distribution for the *pseudo maximum likelihood* estimators under very general conditions on the design points. This includes the case where the number of replications do not grow large, as well as the one where there are no replications. The bootstrap procedure for inference on the regression parameters is also investigated.

Key words and phrases: Chi-square distribution, inverse Gaussian distribution, *pseudo maximum likelihood* estimator, strong consistency, weak convergence.

1. Introduction

The inverse Gaussian distribution has received considerable attention as a model for describing positively skewed data after the pioneering work of Tweedie (1957*a*, 1957*b*) and the subsequent review paper by Folks and Chhikara (1978). The latter paper portrayed many similarities between this distribution and the normal distribution. It also discussed many other provocative departures from the normal distribution. See the review paper by Iyengar and Patwardhan (1988) and the monograph by Chhikara and Folks (1989) for various aspects of this distribu-

* Research supported in part by NSF Grant DMS-9208066.
** Research supported in part by NSERC of Canada.

tion and recent research accounts. Its density function is given by

$$(1.1)\qquad f(y;\delta,\nu)=(2\pi y^3\nu)^{-1/2}\exp\left(-\frac{(1-\delta y)^2}{2\nu y}\right),\qquad 0<y<\infty,$$

where δ is known as the drift parameter and ν is known as the volatility parameter. These parameters take strictly positive values. This two-parameter family of distributions will be referred to as the IG(δ,ν). The mean and variance of density (1.1) are given by $1/\delta$ and ν/δ^3. The above distribution can be viewed as the first passage time distribution of a Wiener process with an absorbing barrier.

The inverse Gaussian distribution has found applications in diverse fields including, life testing (Bhattacharyya and Fries (1982*a*) and Chhikara and Folks (1977)), reliability (Bhattacharyya and Fries (1982*b*) and Whitmore (1979)) to name a few. Several regression models for the inverse Gaussian distribution have been studied by Davis (1977), Whitmore and Yalovsky (1978), Bhattacharyya and Fries (1982*b*) and Whitmore (1983). Chaubey *et al.* (1993) have used such models in survey sampling.

To formulate the inverse Gaussian regression model, let y_i, $i=1,\ldots,n$, be n independent observations distributed respectively as IG(δ_i,ν), where $\delta_i=x_i'\beta>0$. Here $\beta=(\beta_1,\ldots,\beta_p)'$ is a vector of regression parameters and $x_i=(x_{i1},\ldots,x_{ip})'$ is a vector of explanatory variables. We can write $y_i^{-1}=x_i'\beta+\epsilon_i$, where $\nu^{-1}y_i\epsilon_i^2$ are *i.i.d.* χ_1^2 variables. The *pseudo maximum likelihood* estimators of β and ν are derived by Whitmore (1983) and Bhattacharyya and Fries (1986) and are given by

$$(1.2)\qquad \hat\beta=(X'YX)^{-1}X'\mathbf{1},$$

$$(1.3)\qquad \hat\nu=(\mathbf{1}'Y^{-1}\mathbf{1}-\mathbf{1}'X\hat\beta)/n$$

where Y is the diagonal matrix with i-th diagonal element being y_i, $\mathbf{1}$ is the n-vector of all ones and $X=(x_1,\ldots,x_n)'$. These are called *pseudo maximum likelihood* estimators because the condition $x_i'\hat\beta>0$ for all i may not be satisfied. Whitmore (1983) avoids the problem by introducing the notion of a *defective* inverse Gaussian distribution in case the drift parameter is negative but further assuming that resulting estimators are non-negative. It may be observed that (1.2) is also the weighted least squares estimator when the weight associated with response variable $1/y$ is y. Bhattacharyya and Fries (1986) and Fries and Bhattacharyya (1983) study the asymptotic properties of the estimators. This asymptotic theory is based on letting the number of replications go to infinity at a fixed rate. The case of small number of replications and large number of distinct design points, that is of common use in practice, has obtained considerable attention for usual linear regression model (see, e.g., Jacquez *et al.* (1968), Fuller and Rao (1978)). However, such a case has not been considered for the inverse Gaussian regression. In this paper we study the asymptotic properties of the resulting estimators $\hat\beta$ and $\hat\nu$ under very general assumptions on the design points.

The asymptotic distributions of the estimators in (1.2) and (1.3) are derived in Section 2. Section 3 establishes the strong consistency of the estimators. The bootstrap procedure is described and its properties are analyzed in Section 4. Finally, an application of these techniques to a real data set is presented in Section 5.

2. Asymptotic distribution

Recall that the parameter space Θ for the regression coefficients β satisfies the condition $x_i'\beta > 0$ for all i. Instead of this condition, we shall consider slightly stronger assumption $\inf_i x_i'\beta > 0$ for all i and $\beta \in \Theta$. The following theorem shows that the asymptotic distributions of $\hat{\nu}$ and $\hat{\beta}$, when properly normalized, are normal.

THEOREM 2.1. *Suppose for each $\beta \in \Theta$, there exists a constant $c_\beta > 0$ satisfying*

(i) $x_i'\beta \geq c_\beta$ *for all* i

and

(ii) $\max_{1\leq i\leq n} \|x_i\|^3 \operatorname{tr}((X'X)^{-1}) \to 0$ *as* $n \to \infty$.

Further let $U_1 = (n/2)^{1/2}((\hat{\nu}/\nu) - 1)$ *and* $U_2 = (X'MX/\nu)^{1/2}(\hat{\beta} - \beta)$ *then the random vector*

$$(U_1, U_2') \xrightarrow{\mathcal{L}} \mathcal{N}_{(p+1)}(0, I_{(p+1)}),$$

where for any symmetric positive definite matrix G, the unique positive definite square root F of G (i.e. $G = F^2$) is denoted by $G^{1/2}$, and M is the diagonal matrix with i-th diagonal element being $m_i = \delta_i^{-1}$.

Note that as $m_i > 0$ for all i, the matrix $X'MX$ is positive definite, whenever $X'X$ is a positive definite matrix.

Remark 1. Condition (i) basically implies that the reciprocal of the means of y_i are bounded away from zero which is a reasonable assumption in practice and has been considered by several authors (see Chhikara and Folks (1989) for details). Condition (ii) on the design matrix is not too restrictive as it may be implied by a set of conditions which are similar to those assumed in the case of usual multiple linear regression models (see Bunke and Bunke (1986), Section 2.4.2). In particular, condition (ii) holds if for some sequence of positive real numbers $q_n \to \infty$,

(iii) $\max_{1\leq i\leq n} \|x_i\| = o(q_n^{1/3})$

and

(iv) $q_n^{-1}(X'X) \to S$, a positive definite matrix.

The lemmas needed in establishing Theorem 2.1 are presented below.

LEMMA 2.1. *Let $d_{in} = x_i'(X'MX)^{-1}x_i$ and $d_n = \max_{1\leq i\leq n} d_{in}$. Condition* (ii) *of Theorem* 2.1 *implies that $d_n \to 0$ as $n \to \infty$.*

PROOF. Using the spectral decomposition and Cauchy-Schwarz inequality, it is trivial to establish

$$x'Jx \leq \|x\|^2 \operatorname{tr}(J) \quad \text{and} \quad x'J^2x \leq \|x\|^2(\operatorname{tr}(J))^2, \tag{2.1}$$

for any symmetric non-negative definite matrix J and vector x. Hence

$$d_{in} = x_i'(X'MX)^{-1}x_i \leq \|x_i\|^2 \operatorname{tr}((X'MX)^{-1}), \tag{2.2}$$

and

$$\begin{aligned}
(\operatorname{tr}((X'MX)^{-1}))^2 &= \left(\sum_{i=1}^{n} \operatorname{tr}((X'MX)^{-1}(X'X)^{-1}x_i x_i')\right)^2 \\
&= \left(\sum_{i=1}^{n} x_i'(X'MX)^{-1}(X'X)^{-1}x_i\right)^2 \\
&\leq \left(\sum_{i=1}^{n} \|x_i'(X'MX)^{-1}\| \|x_i'(X'X)^{-1}\|\right)^2 \\
&\leq \left(\sum_{i=1}^{n} x_i'(X'MX)^{-2}x_i\right)\left(\sum_{i=1}^{n} x_i'(X'X)^{-2}x_i\right) \\
&\leq \left(\max_{1\leq j\leq n} m_j^{-1}\right)\left(\sum_{i=1}^{n} m_i x_i'(X'MX)^{-2}x_i\right) \operatorname{tr}((X'X)^{-1}) \\
&\leq \left(\max_{1\leq j\leq n} m_j^{-1}\right) \operatorname{tr}((X'MX)^{-1}) \operatorname{tr}((X'X)^{-1}).
\end{aligned}$$

This implies

$$\operatorname{tr}((X'MX)^{-1}) \leq \left(\max_{1\leq j\leq n} m_j^{-1}\right) \operatorname{tr}((X'X)^{-1}). \tag{2.3}$$

Hence, from (2.2) and (2.3) we have

$$d_n \leq \|\beta\| \max_{1\leq j\leq n} \|x_j\|^3 \operatorname{tr}((X'X)^{-1}), \tag{2.4}$$

which tends to zero by the assumptions of the lemma. This completes the proof.

Throughout this paper we use the notation $A = (X'MX)^{-1/2}$. We also use the expectations

$$E(y_i) = \delta_i^{-1}, \qquad E(y_i^{-1}) = \delta_i + \nu, \qquad \text{and} \qquad E(y_i^2) = \nu\delta_i^{-3} + \delta_i^{-2}. \tag{2.5}$$

Note that $\operatorname{Var}(y_i) = \nu\delta_i^{-3}$.

LEMMA 2.2. *Under the conditions of Theorem 2.1, the positive definite matrix*

$$C = A(X'YX)A \xrightarrow{P} I_p.$$

PROOF. We need to prove that for each, $1 \le k, l \le p$, the (k,l)-th element b_{kl} of the matrix $AX'(Y-M)XA$ tends to zero in probability.

Since $E(b_{kl}) = 0$, it is enough to prove that $E(b_{kl})^2 \to 0$. Note that $E(b_{kl})^2 = \nu \sum_{i=1}^n m_i^3 z_{ikl}^2$ where z_{ikl} is the (k,l)-th element of $Z_i = Ax_i x_i' A$. By condition (i), we have

$$\sum_{k,l} E(b_{kl})^2 \le \nu c_\beta^{-2} \sum_{i=1}^n \sum_{k,l} m_i z_{ikl}^2 = \nu c_\beta^{-2} \sum_{i=1}^n \operatorname{tr}(m_i Z_i^2). \tag{2.6}$$

Noting further that

$$\begin{aligned}\operatorname{tr}(Z_i^2) &= \operatorname{tr}(Ax_i x_i'(X'MX)^{-1} x_i x_i' A)\\ &= d_{in} \operatorname{tr}(x_i'(X'MX)^{-1} x_i) \le d_n \operatorname{tr}(x_i x_i'(X'MX)^{-1})\end{aligned}$$

and $\sum_{i=1}^n m_i x_i x_i'(X'MX)^{-1} = I_p$, the inequality in (2.6) implies that

$$\sum_{k,l} E(b_{kl})^2 \le d_n p \nu c_\beta^{-2},$$

which goes to zero by Lemma 2.1. This proves Lemma 2.2.

LEMMA 2.3. *Let $\mathbf{y} = (y_1, \dots, y_n)'$ and $V = X'(\mathbf{1} - M^{-1}\mathbf{y})$. Then under the conditions of Theorem* 2.1,

$$((n/2)^{1/2}((\tilde{\nu}/\nu) - 1), V'A/\sqrt{\nu}) \xrightarrow{\mathcal{L}} \mathcal{N}_{(p+1)}(0, I_{(p+1)}),$$

where $\tilde{\nu} = n^{-1} \sum_{i=1}^n (1 - \delta_i y_i)^2 y_i^{-1}$.

PROOF. First we show that $\operatorname{Cov}(\tilde{\nu}, AV) = 0$. By letting r_i denote the i-th element of the vector $\mathbf{1} - M^{-1}\mathbf{y}$, we have

$$\begin{aligned}\operatorname{Cov}(\tilde{\nu}, AV) &= \operatorname{Cov}\left(n^{-1} \sum_{i=1}^n ((1-\delta_i y_i)^2 y_i^{-1} - \nu), \sum_{i=1}^n A x_i r_i\right)\\ &= n^{-1} \sum_{i=1}^n \operatorname{Cov}(v_i, A x_i r_i),\end{aligned}$$

where $v_i = (1 - \delta_i y_i)^2 y_i^{-1} - \nu$. Since

$$\begin{aligned}E(v_i r_i) &= E(((\delta_i^2 y_i + y_i^{-1} - 2\delta_i) - \nu)(1 - \delta_i y_i))\\ &= E(\delta_i^2 y_i + y_i^{-1}) - E(\delta_i^3 y_i^2 + \delta_i),\end{aligned} \tag{2.7}$$

substituting for the expectations (2.5) in (2.7), we find that $E(v_i r_i) = 0$ which proves $\operatorname{Cov}(\tilde{\nu}, AV) = 0$. Consider now the linear combination $W_{a,b} = a\sqrt{n}(\tilde{\nu} - \nu) + b'AV = \sum_{i=1}^n w_i$, where (a, b') is a non-zero vector and $w_i = (a/\sqrt{n})v_i + b'Ax_i r_i$. We will show that $\sum_{i=1}^n E(w_i^2) = 2a^2\nu^2 + \nu\|b\|^2$ and $\sum_{i=1}^n E(w_i^4) \to 0$ as $n \to \infty$

and hence by Lyapounov's central limit theorem it will follow that for the vector $(a, b') \neq 0$, $W_{a,b}$ will converge to the normal distribution with mean zero and variance $2\nu^2 a^2 + \nu\|b\|^2$. Therefore lemma will follow by the Cramér-Wold device.

Shuster (1968) has shown that the distribution of $(v_i + \nu)/\nu = (1 - \delta_i y_i)^2/(\nu y_i)$ is χ_1^2. Hence using the following cumulants of the inverse Gaussian distribution (see Chhikara and Folks (1989), p. 12)

$$\kappa_2(IG(\delta, \nu)) = \nu/\delta^3, \qquad \kappa_4(IG(\delta, \nu)) = 15\nu^3/\delta_i^7$$

and the χ_1^2 cumulants $\kappa_2(\chi_1^2) = 2$, $\kappa_4(\chi_1^2) = 48$, we have

$$\sum_{i=1}^{n} E(w_i^2) = \sum_{i=1}^{n} (2a^2\nu^2 n^{-1} + m_i(b'Ax_i)^2\nu) \tag{2.8}$$

and

$$\begin{aligned} \sum_{i=1}^{n} E(w_i^4) = &\sum_{i=1}^{n} ((48a^4\nu^4 n^{-2} + 15(b'Ax_i)^4 m_i^3 \nu^3) \\ &+ 3(2a^2\nu^2 n^{-1} + m_i\nu(b'Ax_i)^2)^2) \\ = &\, 48\nu^4 a^4 n^{-1} + 3\nu^2 \sum_{i=1}^{n} (m_i + 5m_i^2\nu) m_i (b'Ax_i)^4 \\ &+ 12a^4\nu^4 n^{-1} + 12a^2\nu^2 n^{-1} \sum_{i=1}^{n} m_i (b'Ax_i)^2. \end{aligned} \tag{2.9}$$

Since

$$\sum_{i=1}^{n} m_i (b'Ax_i)^2 = \operatorname{tr}\left(b'A \left(\sum_{i=1}^{n} m_i x_i x_i' \right) Ab \right) = b'b,$$

and $A \sum_{i=1}^{n} m_i x_i' x_i A = I_p$, (2.8) becomes

$$\sum_{i=1}^{n} E(w_i^2) = 2a^2\nu^2 + \nu\|b\|^2,$$

and by (2.1),

$$\sum_{i=1}^{n} m_i (b'Ax_i)^4 \leq \left(\max_{1 \leq i \leq n} x_i' A b b' A x_i \right) \sum_{i=1}^{n} m_i (b'Ax_i)^2 \leq d_n \|b\|^4.$$

Hence the expression in (2.9) converges to zero by Lemma 2.1 as $n \to \infty$. This completes the proof of Lemma 2.3.

PROOF OF THEOREM 2.1. Since $\hat{\beta} - \beta = (X'YX)^{-1}V$, we have by Lemma 2.3,

$$((n/2\nu)^{1/2}(\tilde{\nu} - \nu), (\hat{\beta} - \beta)'(X'YX)A) \xrightarrow{\mathcal{L}} \mathcal{N}_{(p+1)}(0, \nu I_{(p+1)}). \tag{2.10}$$

As $A(X'YX)(\hat{\beta}-\beta) = C(X'MX)^{1/2}(\hat{\beta}-\beta)$, we have by Lemma 2.2, and (2.10) that $(X'MX)^{1/2}(\hat{\beta}-\beta)$, C^{-1} and C are bounded in probability. Consequently,

$$(X'MX)^{1/2}(\hat{\beta}-\beta) - A(X'YX)(\hat{\beta}-\beta) = (I-C)(X'MX)^{1/2}(\hat{\beta}-\beta) \xrightarrow{P} 0,$$

and hence by (2.10),

$$(X'MX)^{1/2}(\hat{\beta}-\beta) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \nu I_p). \tag{2.11}$$

Further note that

$$n\tilde{\nu}\nu^{-1} = \sum_{i=1}^{n} \xi_i, \tag{2.12}$$

where $\xi_i = (1-\delta_i y_i)^2/(\nu y_i)$, $i = 1, \ldots, n$ are *i.i.d.* χ_1^2 random variables and

$$n\tilde{\nu} = (YX\beta - \mathbf{1})'Y^{-1}(YX\beta - \mathbf{1}) = n\hat{\nu} + (\hat{\beta}-\beta)'(X'YX)(\hat{\beta}-\beta) \tag{2.13}$$

(see Chhikara and Folks (1989), pp. 128–129). By (2.11) and Lemma 2.2,

$$\begin{aligned}&\nu^{-1}(\hat{\beta}-\beta)'(X'YX)(\hat{\beta}-\beta)\\ &\quad = \nu^{-1}(\hat{\beta}-\beta)'(X'MX)^{1/2}C(X'MX)^{1/2}(\hat{\beta}-\beta) \xrightarrow{\mathcal{L}} \chi_1^2,\end{aligned}$$

consequently $\sqrt{n}(\hat{\nu}-\tilde{\nu}) \xrightarrow{P} 0$. This proves Theorem 2.1.

Remark 2. Under conditions of Theorem 2.1, the right hand side of (2.3) converges to zero. Hence by (2.1) and (2.11),

$$\|(\hat{\beta}-\beta)\| \le \|(X'MX)^{1/2}(\hat{\beta}-\beta)\|(\mathrm{tr}((X'MX)^{-1}))^{1/2} \xrightarrow{P} 0,$$

giving the weak consistency of $\hat{\beta}$. The weak consistency of $\hat{\nu}$ obviously follows from Theorem 2.1.

Remark 3. If $n^{-1}(X'MX) \to T$, where T is a positive definite matrix, then under the conditions of Theorem 2.1

$$(n/\hat{\nu})^{1/2}(\hat{\beta}-\beta) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, T^{-1}).$$

Since, M is generally unknown, the practical use of Theorem 2.1 is limited. However, as the following theorem shows, we can get the same limiting distribution as in Theorem 2.1, when we replace M by Y.

THEOREM 2.2. *Under the conditions of Theorem* 2.1

$$(U_1, (\hat{\beta}-\beta)'((X'YX)/\nu)^{1/2}) \xrightarrow{\mathcal{L}} \mathcal{N}_{(p+1)}(0, I_{(p+1)}).$$

PROOF. To prove this theorem it is enough to establish that

$$((X'MX)^{1/2})^{-1}(X'YX)^{1/2} \xrightarrow{P} I_p. \tag{2.14}$$

But by (ii) of Theorem 2.1,

$$(X'MX)^{-1}X'YX - I = (X'MX)^{-1}(X'(Y-M)X) \xrightarrow{P} 0. \tag{2.15}$$

And, (2.14) follows from (2.15) and the following lemma.

LEMMA 2.4. *Let C and D be symmetric non-negative definite matrices. If $C^2D^2 \to I$ then $CD \to I$.*

PROOF. Let P and Q be orthogonal matrices such that $C = P\Lambda P'$ and $D = Q\Gamma Q'$, where $PP' = P'P = I$, $QQ' = Q'Q = I$, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ and $\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_p)$, λ_i, $i = 1, \ldots, p$ being the eigen values of C and γ_i, $i = 1, \ldots, p$ being those of D. Denoting P_i to be the i-th column of P and Q_i to be that of Q, we have

$$C^2D^2 = P\Lambda^2P'Q\Gamma^2Q' = \sum_{i,j} \lambda_i^2\gamma_j^2 P_iP_i'Q_iQ_j' \to I.$$

Since P_i and Q_j are bounded, and $\lambda_i\gamma_j \geq 0$,

$$|(\lambda_i\gamma_j - 1)P_i'Q_j| \leq |(\lambda_i^2\gamma_j^2 - 1)P_i'Q_j| = |P_i'(C^2D^2 - I)Q_j| \to 0.$$

Consequently, $\sum_{i,j}(\lambda_i\gamma_j - 1)P_i'Q_j \to 0$, i.e. $CD - I \to 0$, which completes the proof of the above lemma.

3. Strong consistency

The bootstrap methodology studied in the next section, requires the strong consistency of $\hat{\beta}$ and $\hat{\nu}$. This is established in the next theorem under slightly stronger conditions than those assumed in Theorem 2.1.

THEOREM 3.1. *Assume condition* (i) *of Theorem* 2.1 *and*

$$(\log n) \max_{1 \leq j \leq n} \|x_j\|^3 \operatorname{tr}((X'X)^{-1}) \to 0 \quad \text{as} \quad n \to \infty. \tag{3.1}$$

Then $\hat{\beta} - \beta \to 0$ and $\hat{\nu} - \nu \to 0$ a.s.

The next lemma is useful in proving Theorem 3.1.

LEMMA 3.1. *Under the conditions of Theorem* 3.1, *we have*

$$C = A(X'YX)A \to I_p \quad a.s., \tag{3.2}$$

$$D = (X'MX)^{-1}(X'YX) \to I_p \quad a.s, \tag{3.3}$$

$C^{-1} \to I_p$ *a.s., and* $D^{-1} \to I_p$ *a.s.*

PROOF. Using the inequality

$$1 - (1-u)^{1/2} - (u/2) \le (u^2/4) \quad \text{for} \quad |u| \le 1/2,$$

we get that the moment generating function M of $\mathrm{IG}(\delta, \nu)$ satisfies

$$M(t) = \exp\{(\delta/\nu)(1 - (1 - 2t\nu\delta^{-2})^{1/2})\} \le \exp\{\nu t^2 \delta^{-3} + t\delta^{-1}\},$$

whenever $|t\nu\delta^{-2}| < 1/4$. Hence for any $\epsilon > 0$, $\theta > 0$ and for any set of constants a_{in}, $i = 1, \ldots, n$, satisfying $|\theta a_{in} \nu \delta_i^{-2}| < 1/4$, we get by Markov's inequality, that

$$\begin{aligned} e^{\epsilon\theta} P\left(\sum_{i=1}^n (y_i - m_i) a_{in} > \epsilon\right) &\le \prod_{i=1}^n E(\exp\{\theta a_{in}(y_i - m_i)\}) \\ &\le \exp\left\{\nu\theta^2 \sum_{i=1}^n m_i^3 a_{in}^2\right\}. \end{aligned}$$

As a result, by taking $\epsilon\theta = 3\log n$, we get $P(|\sum(y_i - m_i)a_{in}| > \epsilon) = O(n^{-2})$, provided

$$\text{(3.4)} \qquad \left(\max_{1 \le i \le n} m_i |a_{in}|\right) \log n \to 0 \quad \text{and} \quad (\log n) \sum_{i=1}^n m_i^3 a_{in}^2 \to 0.$$

By the Borel-Cantelli lemma, this yields almost sure convergence of $\sum_{i=1}^n (y_i - m_i)a_{in}$ to zero.

To establish (3.2) we let a_{in} denote the (k,l)-th element of $Ax_i x_i' A$ and verify the convergence in (3.4). For this choice of a_{in}, we have

$$m_i^2 a_{in}^2 \le c_\beta^{-2} \operatorname{tr}(Ax_i x_i'(X'MX)^{-1} x_i x_i' A) \le c_\beta^{-2} d_n^2,$$

and

$$\begin{aligned} \sum_{i=1}^n m_i^3 a_{in}^2 &\le c_\beta^{-2} \sum_{i=1}^n m_i \operatorname{tr}(Ax_i x_i'(X'MX)^{-1} x_i x_i' A) \\ &\le c_\beta^{-2} \left(\max_{1 \le j \le n} x_j'(X'MX)^{-1} x_j\right) \operatorname{tr}(I_p) \\ &\le c_\beta^{-2} p d_n. \end{aligned}$$

These inequalities imply (3.4), since $d_n \log n \to 0$ by (2.4) and (3.1). This proves (3.2).

To prove (3.3), let a_{in} denote the (k,l)-th element of $(X'MX)^{-1} x_i x_i'$. Then we have

$$\text{(3.5)} \qquad a_{in}^2 \le \operatorname{tr}((X'MX)^{-1} x_i x_i' x_i x_i' (X'MX)^{-1}) = \|x_i\|^2 x_i'(X'MX)^{-2} x_i,$$

and hence by (2.1), (2.3) and (3.5),

$$a_{in}^2 \le \|\beta\|^2 \left(\max_{1\le j\le n} \|x_j\|^6\right) (\mathrm{tr}((X'X)^{-1}))^2.$$

This implies the first convergence in (3.4).

Since m_i are bounded by c_β^{-1}, we have by (2.3) and (3.5) that,

$$\begin{aligned}\sum_{i=1}^n m_i^3 a_{in}^2 &\le c_\beta^{-2}\left(\max_{1\le j\le n}\|x_j\|^2\right)\mathrm{tr}\left(\sum_{i=1}^n m_i x_i'(X'MX)^{-2}x_i\right)\\ &\le c_\beta^{-2}\left(\max_{1\le j\le n}\|x_j\|^2\right)\mathrm{tr}((X'MX)^{-1})\\ &\le \|\beta\| c_\beta^{-2}\left(\max_{1\le i\le n}\|x_i\|^3\right)\mathrm{tr}((X'X)^{-1}),\end{aligned}$$

which implies the second convergence in (3.4). This proves (3.3). Convergence of inverses to the identity matrix follows because the determinants of C and D converge to 1 a.s., and the (i,j)-th cofactor tends to 1 or 0 a.s. according as $i=j$ or not.

PROOF OF THEOREM 3.1. Strong consistency of $\hat\beta$ follows by Lemma 3.1, since

$$\begin{aligned}\hat\beta-\beta &= ((X'YX)^{-1}-(X'MX)^{-1})X'\mathbf{1}\\ &= (D^{-1}-I_p)\beta \to 0 \quad \text{a.s.}\end{aligned}$$

To prove the consistency of $\hat\nu$, in view of (2.12), (2.13) and the the strong law of large numbers, it is enough to show that

$$(\hat\beta-\beta)'(X'YX)(\hat\beta-\beta)(\log n)^{-2}\to 0 \quad \text{a.s.} \tag{3.6}$$

Clearly by (2.1), we have

$$\begin{aligned}(\hat\beta-\beta)'(X'YX)(\hat\beta-\beta) &= V'(X'YX)^{-1}V\\ &= V'AC^{-1}AV\\ &\le V'(X'MX)^{-1}V\,\mathrm{tr}((X'YX)^{-1}(X'MX))\\ &\le V'(X'MX)^{-1}V\,\mathrm{tr}((A(X'YX)A)^{-1}).\end{aligned}\tag{3.7}$$

By Lemma 3.1, we have $\mathrm{tr}((A(X'YX)A)^{-1})\to p$ a.s. To complete the proof we show that

$$V'A(\log n)^{-1}\to 0 \quad \text{a.s.,} \tag{3.8}$$

which implies, by (3.6) and (3.7), the strong consistency of $\hat\nu$. Note that for each $1\le j\le p$, with $a_{in}=j$-th coordinate of $x_i'A(m_i\log n)^{-1}$, we have

$$(m_i \log n a_{in})^2 \le x_i'(X'MX)^{-1}x_i \le d_n$$

and

$$(\log n)^2 \sum_{i=1}^{n} m_i^3 a_{in}^2 \le \sum_{i=1}^{n} m_i x_i'(X'MX)^{-1} x_i$$
$$= \operatorname{tr}\left(\sum_{i=1}^{n} m_i x_i x_i'(X'MX)^{-1}\right) = p.$$

Therefore the conditions in (3.4) are verified. This establishes (3.8) completing the proof of Theorem 3.1.

4. Bootstrap

In this section we describe the bootstrap procedure to estimate the sampling distribution of the estimators and study the asymptotic properties. Let Y_i^*, $i = 1, \ldots, n$ be independent inverse Gaussian random variables with means $1/x_i'\hat{\beta}$ and the dispersion parameter $\hat{\nu}$. Let $Y^* = \operatorname{diag}(Y_1^*, \ldots, Y_n^*)$. Given the original sample, the *pseudo maximum likelihood* estimator corresponding to $\hat{\beta}$ is given by $\beta^* = (X'Y^*X)^{-1}X'\mathbf{1}$.

Since

$$x_i'(\hat{\beta} - \beta) = x_i'A(X'MX)^{1/2}(\hat{\beta} - \beta)$$

we have

$$\|x_i'(\hat{\beta} - \beta)\| \le \{x_i'(X'MX)^{-1}x_i\}^{1/2}\|(X'MX)^{1/2}(\hat{\beta} - \beta)\|.$$

Hence under the conditions of Theorem 2.1, we have

$$P\{x_i'\hat{\beta} > (1/2)c_\beta \text{ for all } i\} \to 1 \quad \text{as} \quad n \to \infty,$$

and under the conditions of Theorem 3.1, we have $\inf_{1 \le i \le n} x_i'\hat{\beta} > 0$, for all large n and for almost all samples sequences.

THEOREM 4.1. *Under the conditions of Theorem* 3.1, *we have for almost all sample sequences,*

$$(X'\hat{M}X)^{1/2}(\beta^* - \hat{\beta}) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \nu I_p),$$

where $\hat{M} = \operatorname{diag}(x_1'\hat{\beta}, \ldots, x_n'\hat{\beta})$. *Consequently, for almost all sample sequences,*

$$\sup_{\boldsymbol{z} \in R^p} |P^*((X'\hat{M}X)^{1/2}(\beta^* - \hat{\beta}) \le \boldsymbol{z}) - P((X'MX)^{1/2}(\hat{\beta} - \beta) \le \boldsymbol{z})| \to 0,$$

where P^* *denotes the probability measure induced by the bootstrap sampling scheme, given the original data.*

Proof of Theorem 4.1 is similar to that of Theorem 2.1, as $\hat{\nu} \to \nu$ a.s. Instead of Lemma 2.2, one uses Lemma 3.1.

Remark 4. Using the ideas, especially Lemma 2, of Babu and Bai (1992), it can be shown under very general conditions on x_i that for any $\theta > 0$, and for almost all sample sequences,

$$P(AX'YX(\hat{\beta}-\beta) \in \nu H) - P^*((X'\hat{M}X)^{-1/2}X'Y^*X(\beta^*-\hat{\beta}) \in \hat{\nu}H) \\ = o(n^{-1/2}) + O(\Phi_p((\partial H)^{\theta/\sqrt{n}})),$$

holds for any p-dimensional borel set H.

5. Numerical illustration

Nelson (1971) presents the failure times of two batches of insulation material in a motorette test performed at elevated temperature settings. The first batch was tested at temperature settings of 190°C, 220°C and 240°C, while the second batch was tested only at 260°C. Bhattacharyya and Fries (1982*a*) find the IG reciprocal linear model to be adequate for this data; where they choose the x values given by $x = 10^{-8}(t^3 - 180^3)$, t denoting the temperature in centigrade. In this section we use this data to illustrate the results obtained in this paper. Denoting the failure times (in thousands of hours) by $y_1, y_2, \ldots, y_n$, we are thus fitting the model given by IG(δ_i, ν) where

$$\delta_i = \beta_0 + \beta_1 x_i. \tag{5.1}$$

The *pseudo maximum likelihood* estimates of β_0, β_1 and ν are respectively given by

$$\begin{aligned} \hat{\beta}_0 &= .037310, \\ \hat{\beta}_1 &= 7.317285 \quad \text{and} \\ \hat{\nu} &= .040233. \end{aligned} \tag{5.2}$$

These estimates differ from those reported in Bhattacharyya and Fries (1982*a*), because we have used the data for both batches in our computation whereas they used the data only for batch I.

A crude estimate of the variance covariance matrix is provided by $\hat{\nu}(X'YX)^{-1}$ because of Theorem 2.2 and 3.1. The asymptotic normality of the estimators can be used to approximate the confidence intervals for β_0, β_1 and/or related parameters. For example, an approximation to $100(1-\alpha)\%$ confidence interval for β_i is $\hat{\beta} \pm z_{\alpha/2}\sqrt{\hat{v}(\hat{\beta}_i)}$, where z_q the $100(1-q)$-th percentile of the standard normal variate and $\hat{v}(\hat{\beta}_i)$ is the estimate of the var$(\hat{\beta}_i)$ obtained from the above estimate of the variance-covariance matrix of $\hat{\beta}$. The estimates of variances of the *pseudo maximum likelihood* estimates of β_0, β_1 for the above data are given as

$$\begin{aligned} \hat{v}(\hat{\beta}_0) &= 5.593833 \times 10^{-4}, \\ \hat{v}(\hat{\beta}_1) &= .242709. \end{aligned} \tag{5.3}$$

We also used the Bootstrap method described in Section 4. This provides direct estimates of the bias as well as those of the variances of the estimators of the parameters. We generated 500 bootstrap samples, thus providing $(\hat{\beta}^*_{0(i)}, \hat{\beta}^*_{1(i)}, \hat{\nu}^*_i)$, $i = 1, 2, \ldots, 500$. Taking averages of these yields the Bootstrap estimates,

$$\begin{aligned} &\hat{\beta}_0^* = .039949, \\ &\hat{\beta}_0^* = 7.326207 \quad \text{and} \\ &\hat{\nu}^* = .038258. \end{aligned} \tag{5.4}$$

Thus the estimates of biases of the *pseudo maximum likelihood* estimators are given by

$$\begin{aligned} &\text{estimate of bias of } \hat{\beta}_0 = \hat{\beta}_0^* - \hat{\beta}_0 = .002639, \\ &\text{estimate of bias of } \hat{\beta}_1 = \hat{\beta}_1^* - \hat{\beta}_1 = .008922 \quad \text{and} \\ &\text{estimate of bias of } \hat{\nu} = \hat{\nu}^* - \hat{\nu} = -.001987. \end{aligned} \tag{5.5}$$

Furthermore, the sample variances computed from the 500 bootstrap sample values provide the following estimates of the $\text{var}(\hat{\beta}_0)$ and that of $\text{var}(\hat{\beta}_1)$,

$$\begin{aligned} &\hat{v}(\hat{\beta}_0^*) = 5.214227 \times 10^{-4}, \\ &\hat{v}(\hat{\beta}_1^*) = .245392. \end{aligned} \tag{5.6}$$

We note that for this data the Bootstrap estimates are close to the crude estimates but the usefulness of the Bootstrap is clearly demonstrated in computing the estimate of the bias and the variance of *pseudo maximum likelihood* estimators.

Acknowledgements

A major part of this research was completed while the second author was on sabbatical leave at Pennsylvania State University which generously provided the necessary facilities.

References

Babu, G. J. and Bai, Z. D. (1992). Edgeworth expansions for errors-in-variables models, *J. Multivariate Anal.*, **42**, 226–244.

Bhattacharyya, G. K. and Fries, A. (1982*a*). Inverse Gaussian regression and accelerated life tests, *Survival Analysis* (eds. J. Crowly and R. A. Johnson), IMS Lecture Notes-Monograph Series, **2**, 101–118, Institute of Mathematical Statistics, Hayward, California.

Bhattacharyya, G. K. and Fries, A. (1982*b*). Fatigue failure models—the Birnbaum Saunders versus the inverse Gaussian, *IEEE Transactions on Reliability*, **R-31**, 439–441.

Bhattacharyya, G. K. and Fries, A. (1986). On the inverse Gaussian multiple regression and model checking procedures, *Reliability and Quality Control*, (ed. A. P. Basu), 86–100, Elsevier, North-Holland.

Bunke, H. and Bunke O. (1986). *Statistical Inference in Linear Models*; *Statistical Methods of Model Building*, Vol. 1, Wiley, New York.

Chaubey, Y. P., Nebebe, F. and Chen, P. (1993). Small area estimation under an inverse Gaussian model in finite population sampling (submitted).

Chhikara, R. S. and Folks, J. L. (1977). The inverse Gaussian distribution as a lifetime model, *Technometrics*, **19**, 461–468.

Chhikara, R. S. and Folks, J. L. (1989). *The Inverse Gaussian Distribution*; *Theory*, *Methodology and Applications*, Marcel Dekker, New York.

Davis, A. S. (1977). Linear statistical inference as related to the inverse Gaussian distribution, Ph.D. Thesis, Department of Statistics, Oklahoma State University.

Folks, J. L. and Chhikara, R. S. (1978). The inverse Gaussian distribution and its statistical application—a review, *J. Roy. Statist. Soc.*, **B40**, 263–289.

Fries, A. and Bhattacharyya, G. K. (1983). Analysis of two-factor experiments under an inverse Gaussian model, *J. Amer. Statist. Assoc.*, **78**, 820–826.

Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix, *Ann. Statist.*, **6**, 1149–1158.

Iyengar, S. and Patwardhan, G. (1988). Recent developments in the inverse Gaussian distribution, *Handbook of Statistics* 7, (eds. P. R. Krishnaiah and C. R. Rao), 479–490, Elsevier, North-Holland, Amsterdam.

Jacquez, J. A., Mather, F. J. and Crawford, C. R. (1968). Linear regression with nonconstant, unknown error variances: Sampling experiments with least squares, weighted least squares and maximum likelihood estimators, *Biometrics*, **24**, 607–626.

Nelson, W. B. (1971). Analysis of accelerated life test data, *IEEE Transactions on Electrical Insulation*. **EI-6**, 165–181.

Shuster, J. J. (1968). On the inverse Gaussian distribution function, *J. Amer. Statist. Assoc.*, **63**, 1514–1516.

Tweedie, M. C. K. (1957*a*). Statistical properties of inverse Gaussian distributions—I, *Ann. Math. Statist.*, **28**, 362–377.

Tweedie, M. C. K. (1957*b*). Statistical properties of inverse Gaussian distributions—II, *Ann. Math. Statist.*, **28**, 696–705.

Whitmore, G. A. (1979). An inverse Gaussian model for labour turnover, *J. Roy. Statist. Soc. Ser. A*, **142**, 468–478.

Whitmore, G. A. (1983). A regression method for censored inverse Gaussian data, *Canad. J. Statist.*, **11**, 305–315.

Whitmore, G. A. and Yalovsky, M. (1978). A normalizing logarithmic transformation for inverse Gaussian random variables, *Technometrics*, **20**, 207–208.