

observed data (y_i, \mathbf{x}_i, z_i) $i = 1, \dots, n$, where n is the sample size, can be expressed as

$$g(y_i) = \mu (1 + \beta' \mathbf{x}_i + \tau_i) e^{\gamma' z_i + \varepsilon_i}, \quad (i = 1, \dots, n), \quad (1)$$

where μ is the general mean parameter and β and γ are the $p \times 1$ and $q \times 1$ vectors, respectively, of unknown parameters. The covariates \mathbf{x}_i and z_i are assumed standardized beforehand, i.e.,

$$\begin{aligned} \bar{x}_{.j} &= \frac{1}{n} \sum_{i=1}^n x_{ij} = 0, & s_{\bar{x}_{.j}}^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_{.j})^2 = 1, \\ \bar{z}_{.j} &= \frac{1}{n} \sum_{i=1}^n z_{ij} = 0, & s_{\bar{z}_{.j}}^2 &= \frac{1}{n-1} \sum_{i=1}^n (z_{ij} - \bar{z}_{.j})^2 = 1, \end{aligned}$$

where x_{ij} and z_{ij} are the j th elements of \mathbf{x}_i and z_i , respectively. Moreover, it is assumed that an additive error τ_i and a multiplicative error ε_i are independent, and identically and independently distributed with following means and variances.

$$\begin{aligned} E[\tau_i] &= 0, & E[\varepsilon_i] &= 0, \\ \text{Var}[\tau_i] &= \zeta^2, & \text{Var}[\varepsilon_i] &= \sigma^2. \end{aligned}$$

Each error is not only a measurement error but also a structural inexpressible one by covariates in an each structure. However, in this model, even if each error is assumed to be distributed as the normal distribution, it is difficult to obtain an explicit form of the distribution of $g(y_i)$. Therefore, we consider that the mixed distribution of two errors can be expressed in the one distribution. If we assume that $\varepsilon_i \sim N(0, \sigma^2)$ and $\tau_i \xrightarrow{D} N(0, \zeta^2)$ ($\zeta^2 / (1 + \beta' \mathbf{x}_i)^2 \rightarrow 0$), the following normal approximation can be obtained from the Taylor expansion.

$$\log g(y_i) \approx \log \mu + \log(1 + \beta' \mathbf{x}_i) + \gamma' z_i + \varepsilon_i + \frac{\tau_i}{1 + \beta' \mathbf{x}_i}.$$

Note that $g(y_i)$ has always positive value. Therefore, we assume that $g(y_i)$ is distributed as the log-normal distribution with location parameter η_i and dispersion parameter ψ_i^2 , where

$$\begin{aligned} \eta_i &= \log \mu + \log(1 + \beta' \mathbf{x}_i) + \gamma' z_i - \frac{1}{2} \log \left\{ 1 + \frac{\zeta^2}{(1 + \beta' \mathbf{x}_i)^2} \right\}, \\ \psi_i^2 &= \sigma^2 + \log \left\{ 1 + \frac{\zeta^2}{(1 + \beta' \mathbf{x}_i)^2} \right\}. \end{aligned} \quad (2)$$

The model (1) includes not only the ordinary linear regression model and the logarithmic linear model with normal error, but also the generalized linear model with the normal error as special cases. So, we call the model (1) the *Hybrid Linear Regression Model* (or simply HLRM). The best model assigning covariates to additive and multiplicative structure can be selected, using an information criterion based on the estimators obtained. In the following, we regard as $g(y_i) \equiv y_i$, that is the identity function, to simplify notation.

This paper is organized in the following way. In Section 2, we discuss a log-normal distribution as the distribution of y_i . In Section 3, we consider the method for estimating the unknown parameters. In Section 4, we give some illustrative applications of HLRM to actual data sets such as data on housing price in Hiroshima city and its vicinity (Takahashi et al. (2000)), the cost of constructing nuclear power plants (Mooz (1978)) and crime rates in 47 states of the US (Vandaele (1978)).

2. The Log-normal Distribution

2.1. Validity of Distribution

Here, we discuss the validity of the log-normal distribution as the distribution of y_i , especially, the ordinary linear model and logarithmic linear model are included in HLRM as the special cases.

From a great amount of literature on the distribution theory (see Johnson, Kotz and Balakrishnan (1994), pp. 207-258, etc.), if y_i is distributed as $\text{LN}(\eta_i, \psi_i^2)$, that is the log-normal distribution with location parameter η_i and dispersion parameter ψ_i^2 as in (2), the density function of y_i is written as

$$f(y_i, \eta_i, \psi_i^2) = \frac{1}{\sqrt{2\pi}\psi_i y_i} \exp \left\{ -\frac{(\log y_i - \eta_i)^2}{2\psi_i^2} \right\},$$

and the mean and variance of y_i are given by

$$\begin{aligned} E[y_i] &= \mu(1 + \beta' \mathbf{x}_i) e^{\gamma' \mathbf{z}_i + \sigma^2/2}, \\ \text{Var}[y_i] &= \left\{ \mu(1 + \beta' \mathbf{x}_i) e^{\gamma' \mathbf{z}_i + \sigma^2/2} \right\}^2 \left\{ \left(1 + \frac{\zeta^2}{(1 + \beta' \mathbf{x}_i)^2} \right) e^{\sigma^2} - 1 \right\}. \end{aligned} \quad (3)$$

Further, using the coefficients $a_j(\cdot)$ given in Barakat (1976) and $h_j(\cdot)$ the quasi-Hermite polynomials, the characteristic function of y_i is obtained as

$$C_{y_i}(t) = e^{i\lambda_i t} e^{-\lambda_i^2 \psi_i^2 t^2/2} \sum_{j=0}^{\infty} \frac{(i\psi_i)^j}{j!} a_j(i\lambda_j t) h_j(\lambda_j \psi_j t),$$

where $\lambda_i = e^{\eta_i}$.

If $\zeta^2 = 0$, $\mu = 1$ and $\beta = \mathbf{0}$, i.e., covariates do not have an additive effect on the response, y_i is distributed as the log-normal distribution with location parameter $\eta_i = \gamma' \mathbf{z}_i$ and dispersion parameter $\psi_i^2 = \sigma^2$. Therefore, the mean and variance of y_i is rewritten as

$$E[y_i] = e^{\gamma' \mathbf{z}_i + \sigma^2/2}, \quad \text{Var}[y_i] = \left(e^{\gamma' \mathbf{z}_i + \sigma^2/2} \right)^2 \left(e^{\sigma^2} - 1 \right).$$

These results mean that in the case $\zeta^2 = 0$, $\mu = 1$ and $\beta = \mathbf{0}$, y_i becomes the ordinary logarithmic linear model with normal error.

Moreover, if $\sigma^2 = 0$, $\mu = 1$ and $\gamma = \mathbf{0}$, i.e., the covariates do not have a multiplicative effect on the response y_i , the mean and variance of y_i are written as

$$E[y_i] = 1 + \beta' \mathbf{x}_i, \quad \text{Var}[y_i] = \zeta^2.$$

Let $\Delta_i^2 = \zeta^2 / (1 + \beta' \mathbf{x}_i)^2$, then

$$\begin{aligned} \lambda_i &= \exp \left\{ (1 + \beta' \mathbf{x}_i) \left(1 + \frac{\zeta^2}{(1 + \beta' \mathbf{x}_i)^2} \right)^{-1/2} \right\} = 1 + \beta' \mathbf{x}_i + O(\Delta_i^2), \\ \psi_i^2 &= \log \left\{ 1 + \frac{\zeta^2}{(1 + \beta' \mathbf{x}_i)^2} \right\} = \Delta_i^2 + O(\Delta_i^4). \end{aligned}$$

Therefore,

$$\lambda_i^2 \psi_i^2 = \zeta^2 + O(\Delta_i^4).$$

Using these results and $a_0(s) = 1$, $a_1(s) = 0$, $h_0(s) = 1$ and $h_1(s) = s$, the characteristic function of y_i can be given by

$$C_{y_i}(t) = \exp \left\{ i(1 + \beta' \mathbf{x}_i)t - \frac{1}{2} \zeta^2 t^2 \right\} + O(\Delta_i^2).$$

This means that y_i converges to the normal distribution with mean $1 + \beta' \mathbf{x}_i$ and variance ζ^2 as Δ_i^2 tends to 0. Namely, if $\sigma^2 = 0$, $\mu = 1$, $\gamma = \mathbf{0}$ and ζ^2 is sufficiently small compared with $(1 + \beta' \mathbf{x}_i)^2$, then we regard the distribution of y_i as $N(1 + \beta' \mathbf{x}_i, \zeta^2)$. Because y_i is always positive, it is the valid condition to be the normal distribution that ζ^2 is very small compared to $(1 + \beta' \mathbf{x}_i)^2$.

From the results mentioned above, it seems that this hybrid linear regression model includes the ordinary linear model and logarithmic linear model with normal error as special cases.

2.2. Outline of Derivation

Here, we outline of derivation on the log-normal distribution in HLRM. For the purpose of simplicity, we assume that $\varepsilon_i \sim N(0, \sigma^2)$ or $e^{\varepsilon_i} \sim \text{LN}(0, \sigma^2)$. Then r -th moment of e^{ε_i} can be written as

$$E[(e^{\varepsilon_i})^r] = e^{r^2 \sigma^2 / 2}.$$

Under the condition that τ_i and ε_i are independent, the mean and variance of y_i are given by (3). Using these equations, we consider the distribution of y_i as a log-normal distribution with location parameter η_i and dispersion parameter ψ_i^2 . Each parameter can be defined as follows. If y_i is distributed as $\text{LN}(\eta_i, \psi_i^2)$, the mean and variance of y_i satisfy as

$$E[y_i] = m_i \omega_i^{1/2}, \quad \text{Var}[y_i] = (m_i \omega_i^{1/2})^2 (\omega_i - 1),$$

where $m_i = \log \eta_i$ and $\omega_i = \log \psi_i^2$. Therefore the parameters can be expressed as

$$\eta_i = \log \left\{ \frac{(E[y_i])^2}{\sqrt{(E[y_i])^2 + \text{Var}[y_i]}} \right\}, \quad \psi_i^2 = \log \left\{ \frac{(E[y_i])^2 + \text{Var}[y_i]}{(E[y_i])^2} \right\},$$

and we can obtain the log-normal distribution with parameters (2)

3. Estimation of HLRM

3.1. Estimation of θ

When y_i is distributed as the log-normal distribution having a location and dispersion parameters given by (2), the log-likelihood for $\theta = (\alpha, \beta', \gamma', \zeta^2, \sigma^2)'$ is given by

$$\ell(\mathbf{y}; \theta) = -\frac{n}{2} \log(2\pi) - \sum_{i=1}^n \log y_i - \frac{1}{2} \sum_{i=1}^n \log \psi_i^2 - \frac{1}{2} \sum_{i=1}^n \frac{(\log y_i - \eta_i)^2}{\psi_i^2}, \quad (4)$$

where $\alpha = \log \mu$. The maximum likelihood estimator (MLE) of $\theta, \hat{\theta}$, can be obtained by maximizing this log-likelihood. We use the SPIDER algorithm (Ohtaki and Izumi (1999)) to obtain the value of $\hat{\theta}$.

We must pay attention to the case where the true parameter values that are on the boundary of the parameter space. On this issue, see, Moran (1971), Lehmann (1983) and Self and Liang (1987), etc. So, we calculate $\hat{\theta}$ in three cases:

- (i) $\zeta^2 \neq 0$ and $\sigma^2 \neq 0$, (Dual),
- (ii) $\zeta^2 \neq 0$ and $\sigma^2 = 0$, (Additive),
- (iii) $\zeta^2 = 0$ and $\sigma^2 \neq 0$, (Multiplicative),

where we call the case (i) Dual, case (ii) Additive and case (iii) Multiplicative variance structures, respectively. We choose the optimized $\hat{\theta}$ whose *AIC* (Akaike (1973)) is the smallest among the three cases. *AIC* is defined by

$$AIC = -2\ell(\mathbf{y}; \hat{\theta}) + 2 \times (\text{the number of independent parameters}), \quad (5)$$

where $\ell(\mathbf{y}; \hat{\theta})$ is given by substituting $\hat{\theta}$ into (4). Moreover, *AIC* is used to determine if the covariates have an additive contribution on their response or multiplicative. The candidate model with the smallest *AIC* is regarded as the best model among all candidate models.

3.2. Confidence Intervals of θ

Next, we consider confidence intervals for the unknown parameters of the HLRM. Let θ_0 denote the true parameter. Namely, θ_0 is defined by

$$\theta_0 = \arg \max_{\theta} \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} f_i(y_i; \theta) \log f_i(y_i; \theta) dy_i.$$

Then

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N_{p+q+3}(\mathbf{0}, \mathbf{J}(\theta_0)^{-1}), \quad (n \rightarrow \infty)$$

where

$$\mathbf{J}(\theta_0) = -\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\frac{\partial^2}{\partial \theta \partial \theta'} \log f_i(y_i; \theta) \right] \Big|_{\theta=\theta_0}.$$

In this case, the maximum likelihood estimator is not distributed identically but independently. Takeuchi (1976) described the asymptotic normality of the maximum likelihood estimator in such a case. Modifying this result, we can prove the asymptotic normality of $\hat{\theta}$. The partial derivatives with respect to θ up to the second order are shown in the Appendix. Substituting $\mathbb{E}[\log y_i - \eta_i] = 0$ and $\mathbb{E}[(\log y_i - \eta_i)^2] = \psi_i^2$ into the equations in the Appendix yields

$$\mathbf{J}(\theta_0) = [J_{ab}]|_{\theta=\theta_0}, \quad (a = 1, \dots, 5, b = 1, \dots, 4),$$

where $J_{ab} = J'_{ab}$ and

$$J_{11} = \sum_{i=1}^n \frac{1}{\psi_i^2}, \quad J_{12} = \sum_{i=1}^n \frac{1}{\psi_i^2 \rho_i} \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right) x'_i,$$

$$\begin{aligned}
J_{13} &= \sum_{i=1}^n \frac{1}{\psi_i^2} z'_i, & J_{14} &= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\psi_i^2 (\rho_i^2 + \zeta^2)}, & J_{15} &= 0, \\
J_{22} &= \sum_{i=1}^n \frac{1}{\psi_i^2 \rho_i^2} \left\{ \frac{2\zeta^4}{\psi_i^2 (\rho_i^2 + \zeta^2)^2} + \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right)^2 \right\} \mathbf{x}_i \mathbf{x}'_i, \\
J_{23} &= \sum_{i=1}^n \frac{1}{\psi_i^2 \rho_i} \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right) \mathbf{x}_i z'_i, \\
J_{24} &= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\psi_i^2 \rho_i (\rho_i^2 + \zeta^2)} \left\{ \frac{2\zeta^2}{\psi_i^2 (\rho_i^2 + \zeta^2)} + \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right) \right\} \mathbf{x}_i, \\
J_{25} &= -\sum_{i=1}^n \frac{\zeta^2}{\psi_i^4 \rho_i (\rho_i^2 + \zeta^2)} \mathbf{x}_i, & J_{33} &= \sum_{i=1}^n \frac{1}{\psi_i^2} z_i z'_i, \\
J_{34} &= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\psi_i^2 (\rho_i^2 + \zeta^2)} z_i, & J_{35} &= \mathbf{0}_q, \\
J_{44} &= \frac{1}{4} \sum_{i=1}^n \frac{1}{\psi_i^2 (\rho_i^2 + \zeta^2)^2} \left(\frac{2}{\psi_i^2} + 1 \right), \\
J_{45} &= \frac{1}{2} \sum_{i=1}^n \frac{1}{\psi_i^4 (\rho_i^2 + \zeta^2)}, & J_{55} &= \frac{1}{2} \sum_{i=1}^n \frac{1}{\psi_i^4},
\end{aligned}$$

where $\mathbf{0}_q$ is a q -dimensional vector all of whose elements are 0. Let u_α be the α -level standard normal quantile given by $\Phi(u_\alpha) = \alpha$, where $\Phi(x)$ is the distribution function of $N(0, 1)$. Then, one and two sided α -level confidence intervals for θ_0 can be expressed respectively as

$$\begin{aligned}
\mathcal{I}_1 &= \left[-\infty, \hat{\theta} - \frac{1}{\sqrt{n}} u_{(1-\alpha)} \mathbf{J}(\hat{\theta})^{-1/2} \mathbf{1}_{p+q+3} \right], \\
\mathcal{I}_2 &= \left[\hat{\theta} - \frac{1}{\sqrt{n}} u_{(1+\alpha)/2} \mathbf{J}(\hat{\theta})^{-1/2} \mathbf{1}_{p+q+3}, \hat{\theta} - \frac{1}{\sqrt{n}} u_{(1-\alpha)/2} \mathbf{J}(\hat{\theta})^{-1/2} \mathbf{1}_{p+q+3} \right],
\end{aligned}$$

where $\mathbf{J}(\hat{\theta})$ is defined by substituting $\hat{\theta}$ into $\mathbf{J}(\theta)$ and $\mathbf{1}_{p+q+3}$ is a $p+q+3$ dimensional vector all of whose elements are 1. Needless to say, in case (ii) we omit \mathbf{J}_{a4} ($1 \leq a \leq 4$) and \mathbf{J}_{45} from $\mathbf{J}(\theta)$ and similarly in case (iii) we omit \mathbf{J}_{a5} ($1 \leq a \leq 5$). Then $\mathbf{1}_{p+q+3}$ is rewritten as $\mathbf{1}_{p+q+2}$ in each case.

4. Illustrative Real Data Analyses

In this section, we apply HLRM to several real data sets using "HLREG" which we developed. The HLREG is a free software. Everybody can download it from the web site, <http://apollo.rbm.hiroshima-u.ac.jp/cdrom/index.htm#house>. This software can be used for fitting HLRM to regression data and diagnosing the goodness of fit. Q-Q plot of total log-normal residuals may be useful for detecting lack of fit of HLRM.

Table 1. Notation for data on price of detached house.

Notation		
<i>Price</i>	: in units of 10 thousand yen	The price of the house
<i>Area</i>	: m^2	The lot size
<i>Size</i>	: m^2	Size of building
<i>Age</i>	: <i>years</i>	Age of building
<i>JR</i>	: 1 if train is available conveniently, 0 else	Degree of commuting convenience
<i>Distance</i>	: <i>km</i>	Distance from city center of Hiroshima
<i>Period</i>	: 0 if May 1998, 1 if June 1999	The period of investigation

Table 2. Estimates of variance parameters and *AIC* with the optimal, fully additive, fully multiplicative and intuitive models by variance structure for the house data.

Mean Structure	Variance Structure	Estimates		
		$\sqrt{\hat{\xi}^2}$	$\sqrt{\hat{\sigma}^2}$	<i>AIC</i>
Intuitive	Dual	.08099	.17170	2195.8
	Additive	.19590	.00000	2201.7
	Multiplicative	.00000	.19226	2194.7
Fully additive	Dual	.09841	.15763	2194.4
	Additive	.19455	.00000	2203.5
	Multiplicative	.00000	.19299	2195.9
Fully multiplicative	Dual	.14237	.14438	2212.7
	Additive	.20434	.00000	2210.7
	Multiplicative	.00000	.20226	2210.7

4.1. Price of Detached House

Takahashi et al. (2000) studied the relationship between price of detached houses located in the suburbs of Hiroshima and their environmental condition. One hundred and fifty eight sets of house data were collected from advertising in newspapers on May 1998 and June 1999. The notation for this house data is explained in Table 1. In this subsection, we ascertain an our question that the lot size or a size of building contribute additively on the price, on the other hand, a distance from city center or an age of building contribute multiplicatively on the price. The price of the house, *Price*, was used as the response variable and the others were used as explanatory variables. Further, the identity function was assumed for the link function *g* since a price of house is always positive. All possible combinations of sets of additive and multiplicative explanatory variables and errors were examined. Our intuitive model is expressed as

Table 3. Estimates of regression coefficients with several models for the house data.

[Intuitive model]						
Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		3186.1	(3087.2	3288.0)	501.734	.000
β_1	<i>Area</i>	.14716E-02	(.63432E-03	.23089E-02)	3.445	.001
β_2	<i>Size</i>	.72264E-02	(.51813E-02	.92715E-02)	6.926	.000
γ_1	<i>Period</i>	-.86740E-01	(-.15154	-.21938E-01)	-2.624	.009
γ_2	<i>JR</i>	.17355	(.87697E-01	.25941)	3.962	.000
γ_3	<i>Age</i>	-.17737E-01	(-.21916E-01	-.13557E-01)	-8.318	.000
γ_4	<i>Distance</i>	-.23214E-01	(-.31024E-01	-.15405E-01)	-5.826	.000
$R^2 = .6953, \sqrt{\hat{\zeta}^2} = .00000, \sqrt{\hat{\sigma}^2} = .19226, AIC = 2194.7$						
[Fully additive model]						
Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		3248.8	(3122.2	3380.5)	398.742	.000
β_1	<i>Period</i>	-.73447E-01	(-.15201	.51152E-02)	-1.832	.067
β_2	<i>Area</i>	.90459E-03	(-.30952E-04	.18401E-02)	1.895	.058
β_3	<i>Size</i>	.64104E-02	(.40837E-02	.87370E-02)	5.400	.000
β_4	<i>JR</i>	.14611	(.40261E-01	.25197)	2.705	.007
β_5	<i>Age</i>	-.15535E-01	(-.20345E-01	-.10725E-01)	-6.330	.000
β_6	<i>Distance</i>	-.17176E-01	(-.26689E-01	-.76627E-02)	-3.539	.000
$R^2 = .6898, \sqrt{\hat{\zeta}^2} = .09841, \sqrt{\hat{\sigma}^2} = .15763, AIC = 2194.4$						
[Fully multiplicative model]						
Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		3113.2	(3016.6	3213.0)	499.861	.000
γ_1	<i>Period</i>	-.74887E-01	(-.14286	-.69105E-02)	-2.159	.031
γ_2	<i>Area</i>	.10182E-02	(.21421E-03	.18222E-02)	2.482	.013
γ_3	<i>Size</i>	.67919E-02	(.48044E-02	.87795E-02)	6.698	.000
γ_4	<i>JR</i>	.16894	(.78393E-01	.25949)	3.657	.000
γ_5	<i>Age</i>	-.18952E-01	(-.23090E-01	-.14814E-01)	-8.976	.000
γ_6	<i>Distance</i>	-.20854E-01	(-.29070E-01	-.12637E-01)	-4.975	.000
$R^2 = .6953, \sqrt{\hat{\zeta}^2} = .00000, \sqrt{\hat{\sigma}^2} = .20226, AIC = 2210.7$						

[Intuitive model:]

$$\begin{aligned}
 \text{Price} = & \mu(1 + \beta_1 \text{Area} + \beta_2 \text{Size} + \tau) \exp(\gamma_1 \text{Period} + \gamma_2 \text{JR} \\
 & + \gamma_3 \text{Age} + \gamma_4 \text{Distance} + \varepsilon).
 \end{aligned} \tag{6}$$

The following two models were compared with the optimal model:

[Fully additive model:]

$$Price = \mu(1 + \beta_1 Period + \beta_2 Area + \beta_3 Size + \beta_4 JR + \beta_5 Age + \beta_6 Distance + \tau) \exp(\varepsilon). \quad (7)$$

[Fully multiplicative model:]

$$Price = \mu(1 + \tau) \exp(\gamma_1 Period + \gamma_2 Area + \gamma_3 Size + \gamma_4 JR + \gamma_5 Age + \gamma_6 Distance + \varepsilon). \quad (8)$$

The values of *AIC* for these models are given in Table 2. Our intuitive model is a better one than fully additive and multiplicative ones. So, it can be said that our view is right. Therefore, HLRM can be give a better result than ordinary models whose covariates have either an additive or multiplicative effect on response. The estimates of regression coefficients under these models are listed in Table 3.

4.2. Cost of Nuclear Power Plants

Next sub-section, we search the optimal model by using HLRM. The example comes from 32 light water reactor (LWR) power plants constructed in the US, on which Mooz (1978) reported. Further, Cox and Snell (1981) analyzed these data by applying a multivariate linear model.

The notation for this power plants data is explained in Table 4. The cost of construction, *C*, was used as the response variable and the others were used as explanatory variables. Further, the identity function was assumed for the link function *g*.

Cox and Snell (1981) analyzed these data using the following model:

[Cox and Snell's model:]

$$\log C = \beta_0 + \beta_1 D + \beta_2 \log S + \beta_3 NE + \beta_4 CT + \beta_5 \log N + \beta_6 PT + \varepsilon. \quad (9)$$

Table 4. Notation for data on cost of nuclear power plants.

Notation		
<i>C</i>	: dollars × 10 ⁻⁶	Cost of construction
<i>D</i>	:	Date the construction permit was issued
<i>S</i>	: /MWe	Net capacity
<i>NE</i>	: 0 if the plant was constructed in the north-east US, 0 otherwise	Region
<i>CT</i>	: 1 if used, 0 otherwise	Use of cooling tower
<i>N</i>	: /number	Cumulative number of power plants constructed by each architect-engineer
<i>PT</i>	: 1 if so, 0 if not	Partial turnkey plant

Table 5. Estimates of regression coefficients with Cox and Snell's model for the power plant data.

Parameter	Variable	Coefficient	Standard error
β_0		-13.26	3.140
β_1	D	0.2124	0.0433
β_2	$\log S$	0.7234	0.1188
β_3	NE	0.2490	0.0741
β_4	CT	0.1404	0.0604
β_5	$\log N$	-0.0876	0.0415
β_6	PT	-0.2261	0.1135

$R^2 = .8569, \quad \sqrt{\hat{\sigma}^2} = 0.15922,$

Table 6. Estimates of variance parameters and AIC with the optimal, fully additive and fully multiplicative by variance structure for the power plant data.

Mean Structure	Variance Structure	Estimates		
		$\sqrt{\hat{\zeta}^2}$	$\sqrt{\hat{\sigma}^2}$	AIC
Optimal	Dual	.00000	.13835	311.71
	Additive	.15310	.00000	317.00
	Multiplicative	.00000	.13835	309.71
Fully additive	Dual	.00000	.14589	315.10
	Additive	.16355	.00000	322.30
	Multiplicative	.00000	.14590	313.10
Fully multiplicative	Dual	.10363	.10433	315.52
	Additive	.14763	.00000	313.52
	Multiplicative	.00000	.14684	313.52

The estimates of regression coefficients under this model are listed Table 5. These estimates were obtained by using the ordinary method. This model means that all the variables have a multiplicative contribution on C .

Next, we analyzed these data by applying the HLRM. All possible combinations of sets of additive and multiplicative explanatory variables and errors were examined. The following model was obtained as the optimal one, with which AIC attains the minimum value:

[Optimal model:]

$$C = \mu(1 + \beta_1 D + \beta_2 NE + \beta_3 CT + \tau) \exp(\gamma_1 S + \gamma_2 N + \gamma_3 PT + \varepsilon). \quad (10)$$

The following two models were compared with the optimal model:

[Fully additive model:]

$$C = \mu(1 + \beta_1 D + \beta_2 S + \beta_3 NE + \beta_4 CT + \beta_5 N + \beta_6 PT + \tau) \exp(\varepsilon). \quad (11)$$

Table 7. Estimates of regression coefficients with several models for the power plant data.

[Optimal model]

Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		450.73	(425.62	477.33)	208.881	.000
β_1	<i>D</i>	.25496	(.86223E-01	.42369)	2.962	.003
β_2	<i>NE</i>	.24825	(.77538E-01	.41896)	2.850	.004
β_3	<i>CT</i>	.16474	(.23769E-01	.30572)	2.290	.022
γ_1	<i>S</i>	.94498E-03	(.66778E-03	.12222E-02)	6.682	.000
γ_2	<i>N</i>	-.11682E-01	(-.28681E-01	.53165E-02)	-1.347	.178
γ_3	<i>PT</i>	-.11079	(-.54256	.32099)	-.503	.615

$R^2 = .8617, \sqrt{\hat{\zeta}^2} = .00000, \sqrt{\hat{\sigma}^2} = .13835, AIC = 309.71$

[Fully additive model]

Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		455.03	(388.56	532.88)	75.961	.000
β_1	<i>D</i>	.24256	(-.22919E-01	.50804)	1.791	.073
β_2	<i>S</i>	.88618E-03	(.43743E-03	.13349E-02)	3.871	.000
β_3	<i>NE</i>	.24268	(.60546E-02	.47931)	2.010	.044
β_4	<i>CT</i>	.13511	(-.15914E-01	.28613)	1.753	.080
β_5	<i>N</i>	-.11528E-01	(-.51296E-01	.28240E-01)	-.568	.570
β_6	<i>PT</i>	-.78395E-01	(-.56402	.40723)	-.316	.752

$R^2 = .8464, \sqrt{\hat{\zeta}^2} = .00000, \sqrt{\hat{\sigma}^2} = .14590, AIC = 313.10$

[Fully multiplicative model]

Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		431.44	(410.04	453.96)	233.725	.000
γ_1	<i>D</i>	.21292	(.12322	.30262)	4.652	.000
γ_2	<i>S</i>	.92893E-03	(.64626E-03	.12116E-02)	6.441	.000
γ_3	<i>NE</i>	.23993	(.10756	.37229)	3.553	.000
γ_4	<i>CT</i>	.14111	(.28180E-01	.25405)	2.449	.014
γ_5	<i>N</i>	-.10880E-01	(-.23658E-01	.18979E-02)	-1.669	.095
γ_6	<i>PT</i>	-.24320	(-.44767	-.38733E-01)	-2.331	.020

$R^2 = .8442, \sqrt{\hat{\zeta}^2} = .00000, \sqrt{\hat{\sigma}^2} = .14684, AIC = 313.52$

[Fully multiplicative model:]

$$C = \mu(1 + \tau)\exp(\gamma_1 D + \gamma_2 S + \gamma_3 NE + \gamma_4 CT + \gamma_5 N + \gamma_6 PT + \varepsilon). \quad (12)$$

The values of *AIC* for these models are given in Table 6. It is shown that additive contribution is expected for *D*, *NE* and *CT*. Further, the additive error term can be negligible. The estimates of regression coefficients under these models are listed in Table

Table 8. Notation for data on crime rate.

Notation		
R	Crime rate	: the number of offenses reported to the police per 1,000,000 population
Age	Age distribution	: the number of males aged 14-24 per 1,000 of total state population
Ed	Educational level	: the mean number of years of schooling in the population, 25 years old and over
E_{X_0}	Police expenditure	: the per capita expenditure on police protection by state and local government in 1960
U_2	Unemployment rate	: unemployment rate of urban males per 1,000 in the age-group 35-39
W	Wealth	: wealth as measured by the median value of transferable goods and assets or family income (unit 10 dollars)
X	Income inequality	: the number of families per 1,000 earning below one-half of the median income

7. From the coefficients of determination, it seems that the optimal model is better than Cox and Snell's model.

4.3. Crime Rate

Finally, we deal with crime rate data for 47 states of the US collected in 1960 reported by Vandaele (1978).

The notation for this crime rate is explained in Table 8. The crime rate, R , was used as the response variable and the others were used as explanatory variables. Further, the identity function was assumed for the link function g .

Table 9. Estimates of variance parameters and AIC with the optimal, fully additive and fully multiplicative by variance structure for the crime rate data.

Mean Structure	Variance Structure	Estimates		
		$\sqrt{\hat{\zeta}^2}$	$\sqrt{\hat{\sigma}^2}$	AIC
Optimal	Dual	.00000	.19477	543.36
	Additive	.20368	.00000	546.93
	Multiplicative	.00000	.19476	541.36
Fully additive	Dual	.00835	.21044	551.58
	Additive	.21738	.00000	554.53
	Multiplicative	.00000	.21044	549.58
Fully multiplicative	Dual	.14806	.15033	550.64
	Additive	.21280	.00000	548.64
	Multiplicative	.00000	.21044	548.64

Table 10. Estimates of regression coefficients with several models for the crime rate data.

[Optimal model]

Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		874.63	(729.95	1048.0)	73.422	.000
β_1	Age	.15704E-01	(.75512E-02	.23857E-01)	3.775	.000
β_2	U_2	.92319E-02	(.49948E-03	.17964E-01)	2.072	.038
β_3	W	.17335E-02	(-.39891E-02	.74562E-02)	.594	.553
β_4	X	.75196E-02	(.35377E-03	.14685E-01)	2.057	.040
γ_1	Ed	.16212E-01	(.79214E-02	.24502E-01)	3.833	.000
γ_3	E_{X_0}	.10760E-01	(.27718E-02	.18749E-01)	2.640	.008

$R^2 = .7708, \sqrt{\hat{\zeta}^2} = .00000, \sqrt{\hat{\sigma}^2} = .19476, AIC = 541.36$

[Fully additive model]

Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		876.06	(808.39	949.40)	165.188	.000
β_1	Age	.11796E-01	(.47225E-02	.18869E-01)	3.269	.001
β_2	Ed	.17410E-01	(.56668E-02	.29154E-01)	2.906	.004
β_3	E_{X_0}	.10095E-01	(.63606E-02	.13829E-01)	5.299	.000
β_4	U_2	.77559E-02	(-.14766E-02	.16988E-01)	1.647	.100
β_5	W	.14808E-02	(-.42240E-03	.33841E-02)	1.525	.127
β_6	X	.75196E-02	(.43253E-02	.10714E-01)	4.614	.000

$R^2 = .7304, \sqrt{\hat{\zeta}^2} = .00000, \sqrt{\hat{\sigma}^2} = .21044, AIC = 549.58$

[Fully multiplicative model]

Parameter	Variable	Coefficient	95% Confidence bounds		T	P
			Lower	Upper		
μ		832.89	(784.26	884.54)	219.079	.000
γ_1	Age	.13130E-01	(.61813E-02	.20078E-01)	3.704	.000
γ_2	Ed	.18607E-01	(.90955E-02	.28118E-01)	3.834	.000
γ_3	E_{X_0}	.10146E-01	(.66755E-02	.13616E-01)	5.730	.000
γ_4	U_2	.84963E-02	(.31994E-04	.16961E-01)	1.967	.049
γ_5	W	.23188E-02	(.45917E-03	.41784E-02)	2.444	.015
γ_6	X	.94982E-02	(.59039E-02	.13092E-01)	5.179	.000

$R^2 = .7322, \sqrt{\hat{\zeta}^2} = .00000, \sqrt{\hat{\sigma}^2} = .21044, AIC = 548.64$

All possible combinations of sets of additive and multiplicative explanatory variables and errors were examined. The following model was obtained as the optimal one, with which *AIC* attains the minimum value:

[Optimal model:]

$$R = \mu(1 + \beta_1 \text{Age} + \beta_2 U_2 + \beta_3 W + \beta_4 X + \tau) \exp(\gamma_1 \text{Ed} + \gamma_2 E_{X_0} + \varepsilon). \quad (13)$$

The following two models were compared with the optimal model.

[Fully additive model:]

$$R = \mu(1 + \beta_1 \text{Age} + \beta_2 \text{Ed} + \beta_3 E_{X_0} + \beta_4 U_2 + \beta_5 W + \beta_6 X + \tau) \exp(\varepsilon). \quad (14)$$

[Fully multiplicative model:]

$$R = \mu(1 + \tau) \exp(\gamma_1 \text{Age} + \gamma_2 \text{Ed} + \gamma_3 X_{E_0} + \gamma_4 U_2 + \gamma_5 W + \gamma_6 X + \varepsilon). \quad (15)$$

The values of AIC for these models and the parameter estimates are given in Tables 9 and 10, respectively. It shown that an additive contribution is expected for Age , U_2 , W and X . Further, the additive error term can be negligible.

In these analyses of real data sets, we see that models having variables with both additive and multiplicative contributions at the same time can be better than ones having variables with only additive or multiplicative contributions.

Appendix

From (4) we have the following partial derivatives up to second order,

$$\begin{aligned} \frac{\partial \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \alpha} &= \sum_{i=1}^n \frac{\log y_i - \eta_i}{\psi_i^2}, \\ \frac{\partial \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i=1}^n \left\{ \frac{\zeta^2 x_{ij}}{\psi_i^2 \rho_i (\rho_i^2 + \zeta^2)} + \frac{x_{ij} (\log y_i - \eta_i)}{\psi_i^2 \rho_i} \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right) - \frac{\zeta^2 x_{ij} (\log y_i - \eta_i)^2}{\psi_i^4 \rho_i (\rho_i^2 + \zeta^2)} \right\}, \\ \frac{\partial \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \gamma_j} &= \sum_{i=1}^n \frac{z_{ij} (\log y_i - \eta_i)}{\psi_i^2}, \\ \frac{\partial \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \zeta^2} &= -\frac{1}{2} \sum_{i=1}^n \left\{ \frac{\log y_i - \eta_i + 1}{\psi_i^2 (\rho_i^2 + \zeta^2)} - \frac{(\log y_i - \eta_i)^2}{\psi_i^4 (\rho_i^2 + \zeta^2)} \right\}, \\ \frac{\partial \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \sigma^2} &= -\frac{1}{2} \sum_{i=1}^n \left\{ \frac{1}{\psi_i^2} - \frac{(\log y_i - \eta_i)^2}{\psi_i^4} \right\}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \alpha^2} &= -\sum_{i=1}^n \frac{1}{\psi_i^2}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \alpha \partial \beta_j} &= -\sum_{i=1}^n \left\{ \frac{x_{ij}}{\psi_i^2 \rho_i} \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right) - \frac{2\zeta^2 x_{ij} (\log y_i - \eta_i)}{\psi_i^4 \rho_i (\rho_i^2 + \zeta^2)} \right\}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \alpha \partial \gamma_j} &= -\sum_{i=1}^n \frac{z_{ij}}{\psi_i^2}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \alpha \partial \zeta^2} &= \frac{1}{2} \sum_{i=1}^n \left\{ \frac{1}{\psi_i^2 (\rho_i^2 + \zeta^2)} - \frac{2(\log y_i - \eta_i)}{\psi_i^4 (\rho_i^2 + \zeta^2)} \right\}, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \alpha \partial \sigma^2} &= - \sum_{i=1}^n \frac{\log y_i - \eta_i}{\psi_i^4}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \beta_j \partial \beta_k} &= \sum_{i=1}^n \left[\frac{\zeta^2 x_{ij} x_{ik}}{\psi_i^2 \rho_i^2 (\rho_i^2 + \zeta^2)} \left\{ \frac{2\zeta^2}{\psi_i^2 (\rho_i^2 + \zeta^2)} + \frac{2\zeta^2}{\rho_i^2 + \zeta^2} - 3 \right\} \right. \\ &\quad - \frac{x_{ij} x_{ik}}{\psi_i^2 \rho_i^2} \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right)^2 + \frac{4\zeta^2 x_{ij} x_{ik} (\log y_i - \eta_i)}{\psi_i^4 \rho_i^2 (\rho_i^2 + \zeta^2)} \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right) \\ &\quad + \frac{x_{ij} x_{ik} (\log y_i - \eta_i)}{\psi_i^2 \rho_i^2} \left\{ \frac{2\zeta^4}{(\rho_i^2 + \zeta^2)^2} - \frac{3\zeta^2}{\rho_i^2 + \zeta^2} - 1 \right\} \\ &\quad \left. - \frac{\zeta^2 x_{ij} x_{ik} (\log y_i - \eta_i)^2}{\psi_i^4 \rho_i^2 (\rho_i^2 + \zeta^2)} \left\{ \frac{4\zeta^2}{\psi_i^2 (\rho_i^2 + \zeta^2)} + \frac{2\zeta^2}{\rho_i^2 + \zeta^2} - 3 \right\} \right], \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \beta_j \partial \gamma_k} &= - \sum_{i=1}^n \left\{ \frac{x_{ij} z_{ik}}{\psi_i^2 \rho_i} \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right) - \frac{2\zeta^2 x_{ij} z_{ik} (\log y_i - \eta_i)}{\psi_i^4 \rho_i (\rho_i^2 + \zeta^2)} \right\}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \beta_j \partial \zeta^2} &= - \frac{1}{2} \sum_{i=1}^n \left[\frac{x_{ij}}{\psi_i^2 \rho_i (\rho_i^2 + \zeta^2)} \left\{ \frac{2\zeta^2}{\psi_i^2 (\rho_i^2 + \zeta^2)} + \frac{\zeta^2}{\rho_i^2 + \zeta^2} - 3 \right\} \right. \\ &\quad + \frac{2x_{ij} (\log y_i - \eta_i)}{\psi_i^2 \rho_i (\rho_i^2 + \zeta^2)} \left\{ \frac{\zeta^2}{\rho_i^2 + \zeta^2} - \frac{1}{\psi_i^2} - 1 \right\} \\ &\quad \left. - \frac{2x_{ij} (\log y_i - \eta_i)^2}{\psi_i^4 \rho_i (\rho_i^2 + \zeta^2)} \left\{ \frac{2\zeta^2}{\psi_i^2 (\rho_i^2 + \zeta^2)} + \frac{\zeta^2}{\rho_i^2 + \zeta^2} - 1 \right\} \right], \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \beta_j \partial \sigma^2} &= - \sum_{i=1}^n \left\{ \frac{\zeta^2 x_{ij}}{\psi_i^4 \rho_i (\rho_i^2 + \zeta^2)} + \frac{x_{ij} (\log y_i - \eta_i)}{\psi_i^4 \rho_i (\rho_i^2 + \zeta^2)} \left(\frac{\zeta^2}{\rho_i^2 + \zeta^2} + 1 \right) \right. \\ &\quad \left. - \frac{2\zeta^2 x_{ij} (\log y_i - \eta_i)^2}{\psi_i^6 \rho_i (\rho_i^2 + \zeta^2)} \right\}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_k} &= - \sum_{i=1}^n \frac{z_{ij} z_{ik}}{\psi_i^2}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \gamma_j \partial \zeta^2} &= \frac{1}{2} \sum_{i=1}^n \left\{ \frac{z_{ij}}{\psi_i^2 (\rho_i^2 + \zeta^2)} - \frac{2z_{ij} (\log y_i - \eta_i)}{\psi_i^4 (\rho_i^2 + \zeta^2)} \right\}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \gamma_j \partial \sigma^2} &= - \sum_{i=1}^n \frac{z_{ij} (\log y_i - \eta_i)}{\psi_i^4}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial (\zeta^2)^2} &= \frac{1}{4} \sum_{i=1}^n \left\{ \frac{1 + 2(\log y_i - \eta_i)}{\psi_i^2 (\rho_i^2 + \zeta^2)^2} - \frac{2(\log y_i - \eta_i)^2}{\psi_i^4 (\rho_i^2 + \zeta^2)^2} \right\} \left(\frac{2}{\psi_i^2} + 1 \right), \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial \zeta^2 \partial \sigma^2} &= \frac{1}{2} \sum_{i=1}^n \left\{ \frac{\log y_i - \eta_i + 1}{\psi_i^4 (\rho_i^2 + \zeta^2)} - \frac{2(\log y_i - \eta_i)^2}{\psi_i^6 (\rho_i^2 + \zeta^2)} \right\}, \\ \frac{\partial^2 \ell(\mathbf{y}; \boldsymbol{\theta})}{\partial (\sigma^2)^2} &= \frac{1}{2} \sum_{i=1}^n \left\{ \frac{1}{\psi_i^4} - \frac{2(\log y_i - \eta_i)^2}{\psi_i^6} \right\}. \end{aligned}$$

Acknowledgement

The authors wish to thank Dr. John Cologne, Radiation Effects Research Foundation, Prof. Yasunori Fujikoshi, Chuo University, Prof. Takashi Nakamura, The Institute of Statistical Mathematics, Prof. Sadanori Konishi, the editor of Bulletin of Informatics and Cybernetics, and referees for their valuable comments.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory* (B. N. Petrov and F. Csáki, eds.). Akadémiai Kiadó, Budapest, 267–281.
- Barakat, R. (1976). Sums of independent lognormally distributed random variables. *Journal of the Optical Society of America*, **66**, 211–216.
- Cox, D. R. and Snell, E. J. (1981). *Applied Statistics, Principles and Examples*. Chapman & Hall/CRC, London.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distribution, Volume 1* (2nd ed.). John Wiley & Sons, New York.
- Lehmann, E. L. (1983). *Theory of Point Estimation*. John Wiley & Sons, New York.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). Chapman & Hall/CRC, London.
- Mooz, W. E. (1978). Cost analysis of light water reactor power plants. *Report R-2304-DOE*. Rand Corp., Santa Monica, California.
- Moran, P. A. P. (1971). Maximum-likelihood estimation in non-standard conditions. *Proceedings of the Cambridge Philosophical Society*, **70**, 441–450.
- Ohtaki, M. and Izumi, S. (1999). Globally convergent algorithm without derivatives for maximizing a multivariate function. *Proceedings of Symposium on "Exploratory Methods and Analysis for Nonlinear Structures of Data with Random Variation" in Hiroshima*.
- Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–610.
- Takahashi, A., Yanagihara, H., Ohtaki, M. and Munaka, M. (2000). Nonparametric regression analysis of the relationship between price of detached houses and their environmental conditions. *Proceedings of Journal of the Faculty for Human Development Hiroshima Jogakuin University*, **7**, 57–65 (in Japanese).
- Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Science*, **153**, 12–18 (in Japanese).
- Vandaele, W. (1978). Participation in illegitimate activities: Ehrlich revisited. *Deterrence and Incapacitation* (A. Blumstein, J. Cohen and D. Nagin, eds.). National Academy of Science, Washington, D. C.

Received November 7, 2003

Revised April 29, 2004



Epidural compression in neuroblastoma: Diagnostic and therapeutic aspects

Bruno De Bernardi^{a,*}, Walentyna Balwierz^b, Joanna Bejent^c, Susan L. Cohn^d,
Maria Luisa Garrè^a, Tomoko Iehara^e, Dominique Plantaz^f,
Thorsten Simon^g, Paola Angelini^a, Armando Cama^h, Wendy B. Londonⁱ,
Kim Kramer^j, Howard M. Katzenstein^k, Paolo Tortori-Donati^l,
Andrea Rossi^l, Giulio J. D'Angio^m, Audrey E. Evansⁿ

^aDepartment of Haematology–Oncology, Giannina Gaslini Children's Hospital,
Largo Gerolamo Gaslini, 5, 16148 Genova, Italy

^bPolish-American Institute of Paediatrics, Jagiellonian University Medical College, Krakow, Poland

^cDepartment of Paediatrics, Great Ormond Street Hospital, London, UK

^dDepartment of Pediatrics, Northwestern University and Children's Memorial Hospital, Chicago, IL, USA

^eDepartment of Paediatrics, Kyoto, Japan

^fDepartment of Pediatrics, CHU Grenoble, Grenoble, France

^gDepartment of Paediatric Oncology and Haematology, University of Cologne, Cologne, Germany

^hDivision of Neurosurgery, Giannina Gaslini Children's Hospital, Genova, Italy

ⁱDepartment of Statistics, University of Florida and Children's Oncology Group Statistical Office, Gainesville, FL, USA

^jDepartment of Paediatrics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

^kAFLAC Cancer Center and Blood Disorders Service, Children's Healthcare of Atlanta, Atlanta, GA, USA

^lService of Neuroradiology, Giannina Gaslini Children's Hospital, Genova, Italy

^mRadiation Oncology, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

ⁿDepartment of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

Received 11 January 2005; accepted 30 January 2005

Abstract

The involvement by tumour of intervertebral foramina and the consequent invasion of the spinal space, accompanied or not by neurological symptoms, represent a well-recognised pattern of presentation of neuroblastoma. The main peculiarity of this condition stands in the fact that, in case of its late detection or inadequate treatment, severe, permanent neurological compromise may ensue. Surprisingly enough, remarkable disagreements still exist regarding its optimal treatment and the related literature provide contrasting indications at this respect. The neurosurgical and the chemotherapeutic approaches have equally convinced supporters, while the use of radiation therapy is uncommon, possibly without good reasons. This mini-review intends to report

* Corresponding author. Address: Division of Oncology, Department of Paediatric Haematology–Oncology, Giannina Gaslini Children's Hospital, Largo Gerolamo Gaslini, 5, 16148 Genova, Italy. Tel.: +39 010 56 36 464; fax: +39 010 56 36 714.

E-mail address: brunodebernardi@ospedale-gaslini.ge.it (B.D. Bernardi).

the clinical experiences of the major Paediatric Oncology Groups with the aim to collect as many data as possible in the perspective of establishing common guidelines for proper diagnosing and treatment of this important complication.

© 2005 Elsevier Ireland Ltd. All rights reserved.

Keywords: Neuroblastoma; Epidural compression; Spinal cord compression; Late effects

1. Introduction (by Audrey E. Evans, Philadelphia)

Signs of cord compression from a paravertebral neuroblastoma make up one of the few emergencies that may occur in paediatric oncology [1–4]. The decision regarding appropriate treatment often depends upon the specialist who first sees the patient: the neurosurgeon or the oncologist. The former knows that decompression can relieve the problem promptly; and it is to hope, is equally aware of the truly major orthopaedic problems with normal growth and development that ensue. The chemotherapist's efforts are usually effective, and no important late complications can be expected from the modest doses of chemotherapy that are usually employed. Days may be needed to relieve the problem, however, is there time? Although not consulted initially, the radiation therapist also knows that irradiation is rapidly beneficial; and with the low doses used to day, has minimal long-term side effects. The choice among these three treatments that are equally effective revolves around the two major issues: time and late effects. The extraordinary resilience of the childhood spinal cord here enters the equation. The dicta concerning irreversible paralysis always supervening after more than a few hours of frank paresis from cord compression do not hold in paediatrics. Full recovery has been observed by me even after several days. Emergency, middle-of-the-night surgery is therefore seldom warranted in this setting and can be planned more deliberately. To be sure, laminectomy uniquely has the advantage of guaranteed immediate relief. It also provides tissue for accurate diagnosis and for various tests and research projects, although these last are a very secondary consideration in this matter. These advantages are offset by the very major scoliotic deformities that ensue. Insofar as radiation therapy is concerned, relatively early responses can be expected without serious adversities later on. A series of treatments over 5 days or so is necessary nonetheless. Moreover,

not all institutions have radiotherapy readily available, or the radiotherapy facility may be off-site making daily transport necessary. These several competing aspects were discussed and debated by experts in each of the disciplines involved. Their deliberations are detailed in the report that follows.

2. Diagnostic imaging (by Paolo Tortori-Donati and Andrea Rossi, Genova)

Neuroblastoma invading the intervertebral foramina and the spinal canal typically displays a dumbbell aspect [5]. The intraspinal component usually remains confined extradurally. X-rays may directly show calcified paraspinal masses. More often, indirect signs such as lytic or sclerotic changes in the adjacent bone are seen. Enlargement of one or more neural foramina and widened interpeduncular distances are all also frequently found. MRI adequately depicts the intraspinal extension of the mass. Often, the intraspinal component extends for several metamerics both cranially and caudally to the entrance foramen (Fig. 1A and B). Such development is optimally depicted on the coronal plane. The lesion is usually hypointense on T1-weighted images and iso- to hypointense on T2-weighted images due to high cellularity and nuclear-to-cytoplasmic ratio. However, necrotic-cystic changes, haemorrhages, and calcifications can result in heterogeneous signal behaviour. Contrast enhancement is usually marked. MRI also detects compression and dislocation of the spinal cord. Intramedullary hyperintense signal on T2-weighted images reflects cord oedema in the setting of compression-related myelopathy (Figs. 2 and 3).

Differential diagnosis includes a variety of tumours, such as lymphomas, skeletal and extraosseous Ewing's sarcomas, neuroblastoma and malignant nerve sheath tumours.

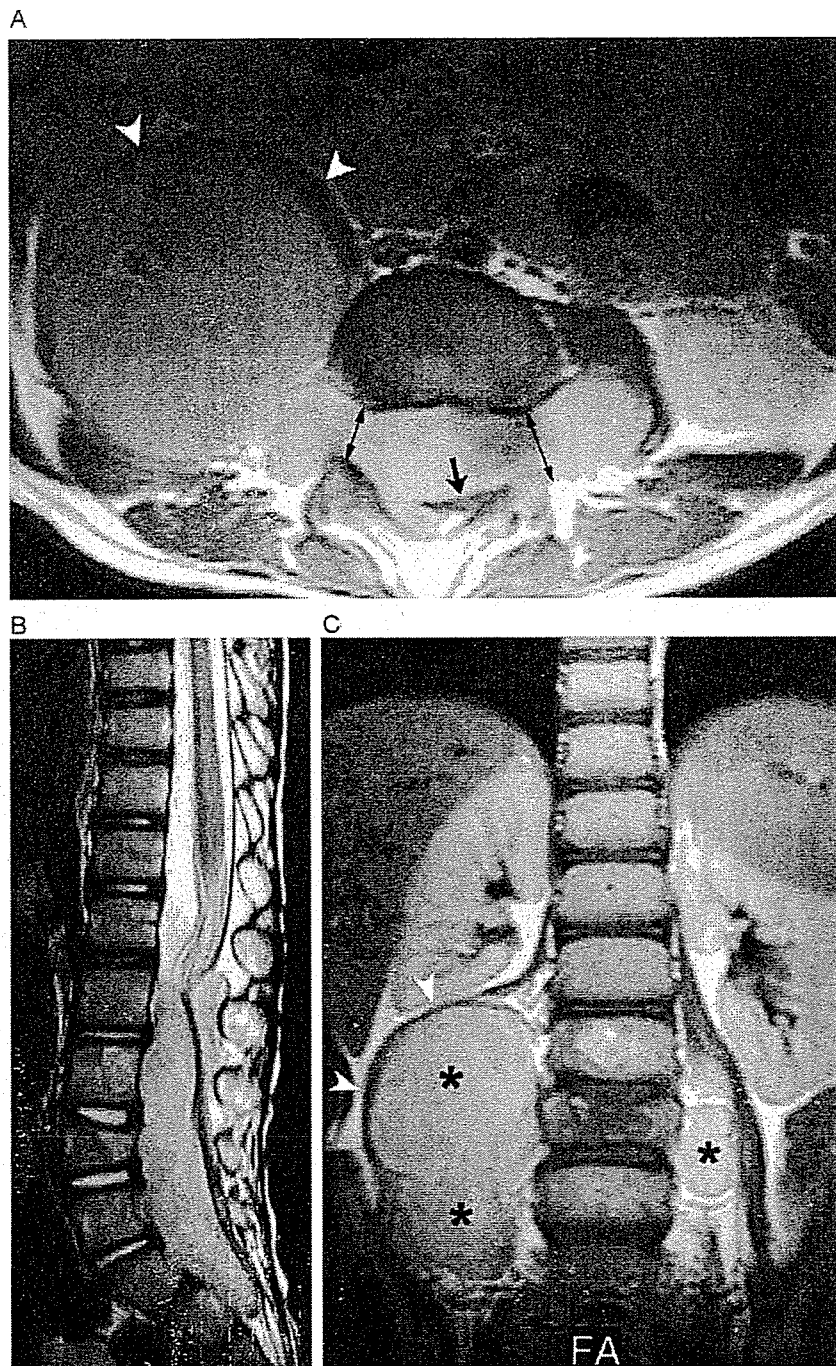


Fig. 1. Retroperitoneal neuroblastoma with bilateral intraspinal extension in a 3-year-old boy. (A) Gd-enhanced coronal T1-weighted image. (B) Sagittal T2-weighted image. (C) Gd-enhanced axial T1-weighted image. Coronal image shows large tumour masses to both sides of the lumbar spine (asterisks, A) and a pathologic L4 vertebral body. The size of the intraspinal extradural component is huge (B). Axial image shows bilateral dumbbell extension through widened neural foramina (double arrows, C) and marked compression of the thecal sac (single arrow, C). Notice marked elevation and displacement of the right psoas muscle (arrowheads, A,C).

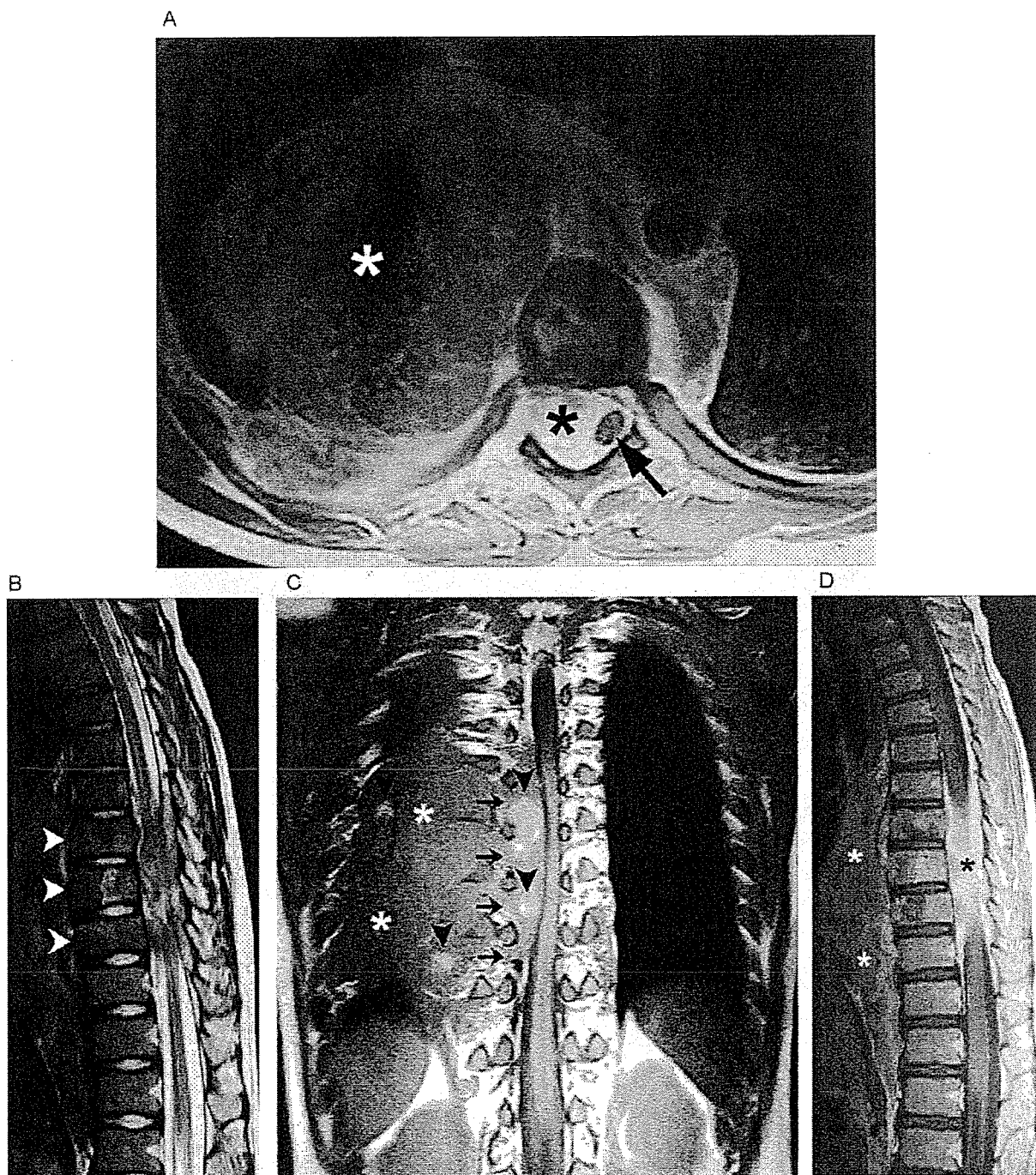


Fig. 2. Thoracic neuroblastoma with intraspinal invasion in a 6-year-old boy. (A) Coronal T1-weighted image. (B) Sagittal T2-weighted image. (C) Gd-enhanced sagittal T1-weighted image. (D) Axial Gd-enhanced axial T1-weighted image. Huge thoracic mass (black asterisks, A–D) extends an intraspinal component through four, markedly enlarged neural foramina (arrows, A). This component, hypointense on T2-weighted images (B) and enhancing with contrast material administration (black asterisk, C,D), compresses and engulfs the spinal cord (arrow, D). The mass has numerous hemorrhagic components (arrowheads, A). Three pathologic vertebrae are seen (arrowheads, B).

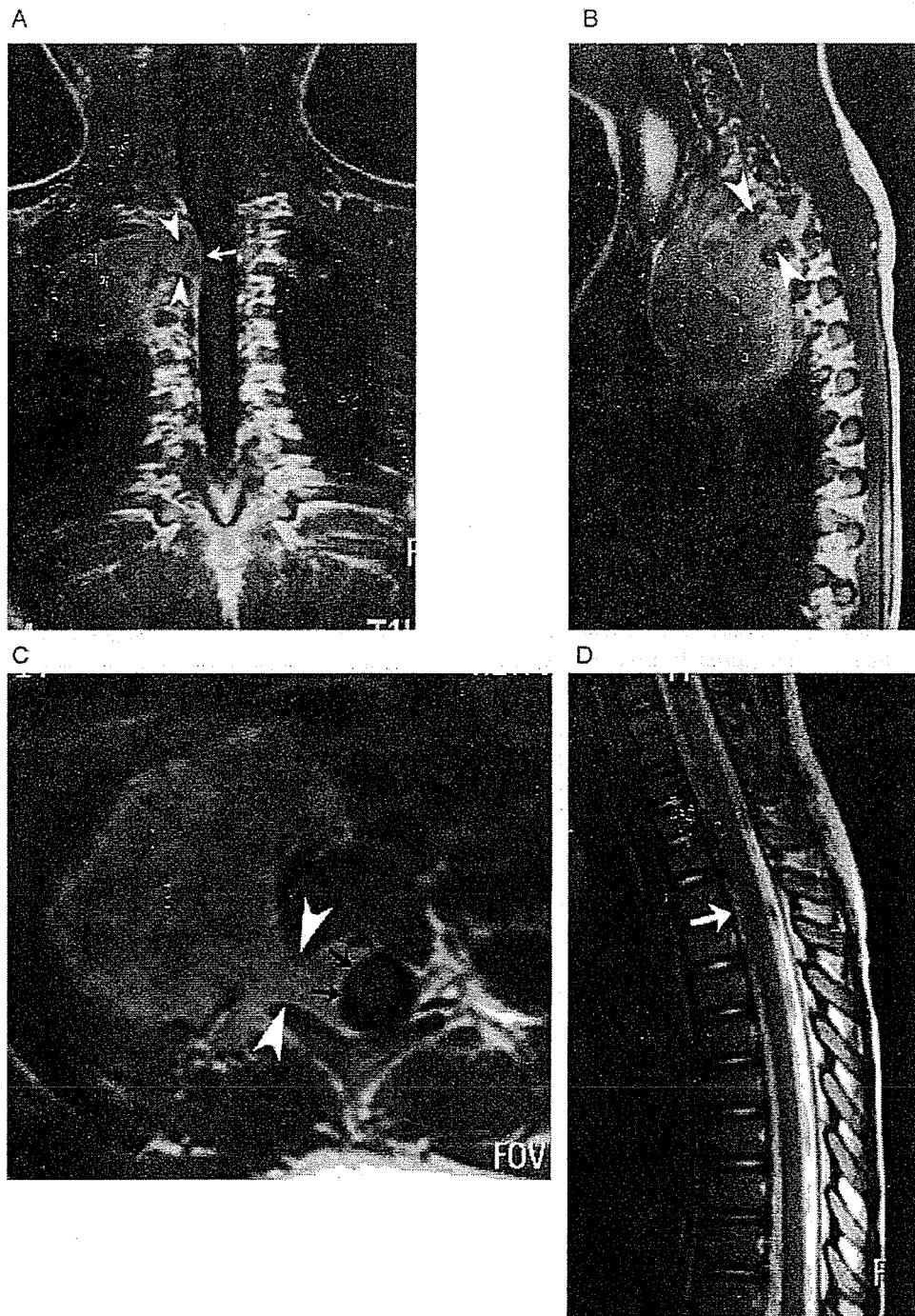


Fig. 3. Thoracic neuroblastoma with intraspinal extension in a 2-year-old girl. (A) Gd-enhanced coronal T1-weighted. (B) Gd-enhanced parasagittal T1-weighted. and (C) Gd-enhanced axial T1-weighted image show thoracic mass with a dumbbell extension through an enlarged neural foramen (arrowheads), invading the epidural space (arrows, A). Notice that, although the thecal sac containing the spinal cord is slightly displaced contralaterally, preservation of the subarachnoid spaces around the cord (arrows, C) indicates absence of spinal cord compression. (D) Midsagittal T2-weighted image shows mild amount of pathological tissue in the ventral epidural space (arrow) and a normal spinal cord.