# Efficient Parametric/Semiparametric Inference in Exchangeable/Partially Exchangeable Models With Applications

Hanxiang Peng
Joint work with X. Dang, S. Keeton, and Y. Wang

Department of Math Sciences
IUPUI

January 2, 2010

# Outline

# Modelling Correlated Data

• *Correlated data* embraces a multitude of data structures, such as multivariate observations, clustered data, repeated measurements, longitudinal data, and spatially correlated data.
• National Toxicology Program Data. The developmental toxicity studies are conducted at the Research Triangle Institute. These studies investigate the effects in mice of five chemicals: EG (ethylene glycol), TGDE (Triethylene Glycol Dimethyl Ether), DYME(diethylene glycol dimethyl ether), DEHP(di(2-ethylhexyl)phthalate) and THEO (theophylline).

• EG(Ethylene glycol, also called 1,2-ethanediol, HOCH2CH2OH): A high-volume industrial chemical with many applications including: antifreeze in cooling and heating systems, component of brake fluids, ingredient of inks, electrolytic condensers, plasticizers, synthetic fibers and waxes, solvent in the paint, softening agent for cellophane, stabilizer for soybean foam, etc.
• EG represents little hazard to human health in normal industrial handling. However, accidental or intentional ingestion of antifreeze products, of which approximately 95% is EG, is toxic and may result in death.
• EG Study in Mice and Rats: timed-pregnant CD-1 mice were dosed by gavage with EG in distilled water.

# DEHP, DYME, THEO

• DEHP(Di(2-ethylhexyl)phthalate, C24H38O4): used in vacuum pumps, extensively as plasticizers, desirable flexibility and clarity.
• DYME (Diethylene Glycol Dimethyl Ether, CH3O(CH2)2O(CH2)2OCH3J): component of industrial solvents, widely used in the manufacture of protective coatings such as lacquers, metal coatings, baking enamels, etc.
• THEO (Theophylline, C7H8N4O2) used in the treatment of respiratory diseases, as an anti-asthmatic, diuretic, etc. Theophylline has been shown to cross the human placenta and is secreted in breast milk. Therefore, there has been an increased interest in the teratogenetic potential of theophylline in rodents.

# Belgian Health Interview Survey

• The HIS was conducted to evaluate the usefulness of a periodic health-related survey, with the idea to collect information on the subjective health of the Belgian population, as well as on important predictor variables.
• The focus is on: (1) identification of health problems, (2) description of the health status and health needs of the population, (3) estimation of prevalence and distribution of health indicators, (4) analysis of social (in)equality in health and access to the health services and (5) study of health consumption and its determinants.
• The target population is defined as all people residing in Belgium at a particular point in time. The National Register is being used as the sampling frame (the homeless is excluded).

- The sampling of respondents took place in the following steps: (1) stratification by region and province, (2) selection of the municipalities within each stratum, (3) selection of a cluster of households within each municipality and (4) selection of respondents within a household.
- As a result, data are clustered at the levels of municipality and household.

# POPS Data

The Project On Preterm and Small-for-gestational age infants (POPS) collected information on 1338 infants born in the Netherlands in 1983 and having gestational age less than 32 weeks and/or birthweight less than 1500g (Verloove et al. 1988). The outcome of interest here concerns the situation after two years. The binary variable is 1 if an infant has died within two years after birth or survived with a major handicap, and 0 otherwise. Some of the recorded observations are from twins or triplets. So, one might have to account for the association between siblings of the same twin (or triplet, . .. ). Another interesting aspect is that there are observations on both cluster and individual level. For example, for a twin, the mothers age and the gestational age is the same for both siblings, while birthweight is subject specific.

# The Wisconsin Diabetes Study

In this data set there are records from 720 younger onset diabetic persons. Both eyes of each person are examined for the presence of macular edema. In the study there were 29 individuals where macular edema is present at only one eye, for 17 of them it is observed at both eyes, and for the remaining 674 persons, it was completely absent. One is interested in the probability of macular edema as a function of the patients systolic blood pressure, hereby taking the clustered nature of the data into account, as indeed the response values of both eyes are likely to be correlated.

# The E2 data (Dang, Keeton, and Peng, *Stat. Med.*, '09)

The E2 data (Brooks *et al.* 1997) records fetal control mortality in mouse litters. There are 211 litters in total with litter sizes varying from small (as to 3) to large (as to 19), having the mean litter size 12.9 and the standard deviation 2.68. The proportion of dead fetuses is 0.110. Among the 211 litters, there are 135 litters which have at least one fetal fetus.

# Bladder Cancer Study (Peng, Garner and Dang (2009), Quintana and Müller (2004))

- Measurements on bladder cancer

$$B = \begin{cases} 1, & \text{cancer} \\ 0, & \text{normal} \end{cases} \qquad X = \begin{cases} 1, & \text{treatment} \\ 0, & \text{control} \end{cases}$$

- All patients had bladder tumors when they entered the trial and the tumors were removed transurethrally.
- Reexamined every 3 months, any tumors recurrences were removed.
- Maximum sequence length: 12
- Sample size: 73

# Burn Wounds (Peng, Rayner, Wang and Tan, JSPI, '09))

The clinical dataset of retrospective study consists of 153 patients (age from 2 months to 82 years) with etiology of fire/flame, scald and contact burns, who were hospitalized from year 1985 to 2000. These patients were treated by an enzymatic debriding agent, in partial deep dermal or full thickness burn wounds. Each patient had 1-16 different burn wound locations on his/her body such as head, neck, left/right hand, left/right leg, etc. There were 19 pre-defined burn wound locations, such as head, left/right hand, left/right leg, etc., in the study protocol and case report form. There were a total of 393 burn wound locations among the 153 patients with burn wound locations up to 12. The mean and variance of the number of wound locations are 2.57 and 3.76 respectively.

# Disadvantages of Some Commonly Used Methods

- The statistical inference based on the conditional analysis does not use any information about the distribution of the latent effects since it is lost in the conditioning.
- The EM algorithm method by Mislevy (1985) is conceptually simple, but the computations may be formidable because each E-step in the computation may require a numerical integration.
- The empirical Bayes approach by Stiratelli, Laird, and Ware (1984) may also be computationally difficult.
- The quasi-likelihood and GEE approach (e.g. Liang, Qaqish and Zeger, 1992) only uses the first two moments, while higher order of correlation is approximated by a "working matrix".

# Proposed Exch. Models Indexed by Nonpara. Measure

- Suppose that $\mathbf{e}_1, \mathbf{e}_2, ...$ is an infinite sequence of *exchangeable* random vectors:

$$\mathbb{P}(\mathbf{e}_{i_1} = \mathbf{b}_1, ..., \mathbf{e}_{i_m} = \mathbf{b}_m) = \mathbb{P}(\mathbf{e}_{\pi_1} = \mathbf{b}_1, ..., \mathbf{e}_{\pi_m} = \mathbf{b}_m),$$

  where $\pi_1, ..., \pi_m$ is an arbitrary permutation of any finite subset $i_1, ..., i_m$ of the sequence of positive integers.
- By Hewitt and Savage (1955),

$$f(\mathbf{z}|Q) = \mathbb{P}(\mathbf{e}_1 = \mathbf{b}_1, ..., \mathbf{e}_m = \mathbf{b}_m) = \int_\Lambda g(\boldsymbol{\lambda}; \mathbf{z})\, dQ(\boldsymbol{\lambda}),\ \mathbf{z} \in \mathbb{Z} \tag{1}$$

  for some unique probability measure $Q$ on $\Lambda$.
- Alternatively (1) can also be described as a *mixture of distributions*, *random-effects model*, and *Bayesian model*.

# Discretized Exch. Models Indexed by Euclidean Parameters

- Expand $g(\boldsymbol{\lambda}; \mathbf{z})$ in a Maclaurin series in $\boldsymbol{\lambda}$,

$$g(\boldsymbol{\lambda}; \mathbf{z}) = \sum_{\mathbf{k}(\mathbf{z}) \in \mathcal{K}} a_{\mathbf{k}(\mathbf{z})}(\mathbf{z}) \boldsymbol{\lambda}^{\mathbf{k}(\mathbf{z})}, \quad \mathbf{z} \in \mathbb{Z}, \tag{2}$$

  where $\mathbf{k}(\mathbf{z}) \in \mathcal{K}$ is a vector of integer-valued components, and $a_{\mathbf{k}(\mathbf{z})}(\mathbf{z})$ is a scalar function of $\mathbf{z}$ independent of $\boldsymbol{\lambda}$.
- Substituting (2) in (1),

$$f(\mathbf{z}|\mathbf{p}) = \sum_{\mathbf{k}(\mathbf{z}) \in \mathcal{K}} a_{\mathbf{k}(\mathbf{z})}(\mathbf{z}) p_{\mathbf{k}(\mathbf{z})}, \quad \mathbf{z} \in \mathbb{Z}, \tag{3}$$

  where $\mathbf{p} = \{p_{\mathbf{k}} : \mathbf{k} \in \mathcal{K}\}$ are the (Euclidean) parameters with $p_{\mathbf{k}} = \int_{\Lambda} \boldsymbol{\lambda}^{\mathbf{k}} \, dQ(\boldsymbol{\lambda})$ the $\mathbf{k}$-th moment of $Q$, and $\mathcal{K}$ is a common index set.

# Exch. Binomial (Dang, Keeton and Peng, *Stat. Med.*, '09)

- Let $Y = B_1 + ... + B_m$. Then (1) holds with $g(\lambda; \mathbf{z}) = \lambda^y (1-\lambda)^{m-y}$ and (3) takes the form:

$$f(\mathbf{z}|\boldsymbol{\lambda}_m) = f(y, m|\boldsymbol{\lambda}_m) = \sum_{k=0}^{m-y} (-1)^k \binom{m-y}{k} \lambda_{y+k}, \qquad (4)$$

  where $\boldsymbol{\lambda} = \{\lambda_k : k \in \mathcal{K}\}$. Clearly $\mathbb{P}(Y = y) = \binom{m}{y} f(\mathbf{z}|\boldsymbol{\lambda}_m)$.
- The binomial is a special case: $\lambda_k = \lambda_1^k$.
- $Y$ is a mixture of binomials, $\mathbb{P}(Y = y) = \binom{m}{y} \int_0^1 \lambda^y (1-\lambda)^{m-y} \, dQ(\lambda)$.

$\boldsymbol{\lambda}$ is *completely monotone (CM)*:

$$(-1)^l \Delta^l \lambda_k \geq 0, \quad k = 0, 1, ..., m, \ l = 0, 1, ..., m - k, \qquad (5)$$

where $\Delta$ is the forward difference operator.

# Exch. Neg. Binom.(Peng, Rayner, Wang, Tan, *JSPI*'09)

- Let $Y$ be the number of trials to get the first $r$ successes. Then $g(\lambda; \mathbf{z}) = \lambda^r(1-\lambda)^{y-r}$ and (3) takes the form:

$$f(\mathbf{z}|\boldsymbol{\lambda}_{(r)}) = \sum_{k=0}^{y-r}(-1)^k \binom{y-r}{k} \lambda_{r+k}, \quad y = r, r+1, ..., \quad (6)$$

  where $\boldsymbol{\lambda}_{(r)} = \{\lambda_k : k = r, r+1, ...\}$ are the parameters. Clearly $\mathbb{P}(Y = y) = \binom{y-1}{r-1} f(\mathbf{z}|\lambda_{(r)})$.
- The negative binomial is a special case: $\lambda_k = \lambda_1^k$.
- The distribution of $Y$ is a mixture of negative binomials, $\mathbb{P}(Y = y) = \binom{y-1}{r-1} \int_0^1 \lambda^r(1-\lambda)^{y-r}\, dQ(\lambda)$.
- A nonparametric mixture of negative binomials is equivarant to a "parametric distribution" with infinitely many parameters $\lambda_r, \lambda_{r+1}, ...$ that are completely monotone.

# Exch. Trinomial (Ding, '07)

- Let $Y_1 = B_1 + ... + B_m$ and $Y_2 = C_1 + ... + C_m$, where $\{\mathbf{e}_i = (B_i, C_i) : i = 1, 2, ...\}$ are exchangeable random vectors of $\{0, 1\}$-valued components. Then $g(\boldsymbol{\lambda}; \mathbf{z}) = \lambda_1^{y_1} \lambda_2^{y_2} (1 - \lambda_1 - \lambda_2)^{m - y_1 - y_2}$ and (3) takes the form:

$$f(\mathbf{z}|\boldsymbol{\pi}_m) = \sum_{k=0}^{m-y} (-1)^k \binom{m-y}{k} \sum_{j=0}^{k} \binom{k}{j} \pi_{y_1+j,\, y_2+k-j}, \quad (7)$$

  where $y = y_1 + y_2 \leq m$, $\boldsymbol{\pi}_m = \{\pi_{j,k} : j, k = 0, 1, ..., m\}$ with $\pi_{0,\,0} = 1$ are the parameters. Clearly $\mathbb{P}(Y_1 = y_1, Y_2 = y_2) = \binom{m}{y_1, y_2} f(\mathbf{z}|\boldsymbol{\pi}_m)$,

- The trinomial is a special case: $\lambda_{j,k} = \lambda_{1,0}^j \lambda_{0,1}^k$.

- $(Y_1, Y_2)$ has a mixture of trinomials: $\mathbb{P}(Y_1 = y_1, Y_2 = y_2) = \binom{m}{y_1, y_2} \iint_{\lambda_1+\lambda_2\leq 1} \lambda_1^{y_1}\lambda_2^{y_2}(1-\lambda_1-\lambda_2)^{m-y_1-y_2}\, dQ(y_1, y_2)$.
- $\boldsymbol{\pi}_m = \{\pi_{i,j} : i, j = 0, 1, ..., m, i + j \leq m\}$ $(\pi_{0,0} = 1)$ is *simplex completely monotone (SCM)* for $i, j = 0, 1, ..., m, i + j \leq m$,

$$(-1)^r \bigtriangledown^r \pi_{i,j} \geq 0, \quad r = 0, 1, 2, ..., r \leq m - i - j. \tag{8}$$

where $\bigtriangledown$ is the two-dimensional forward difference operator:

$$\bigtriangledown \pi_{i,j} = \pi_{i+1,j} + \pi_{i,j+1} - \pi_{i,j}. \tag{9}$$

$\bigtriangledown^2 = \bigtriangledown(\bigtriangledown)$, and $\bigtriangledown^0$ the identity operator. Gupta (1999) calls such a SCM sequence *a partial G-multisequence of order m and degree* 2.

# Partially Exch. Binom. (Peng, Garner and Dang, '09)

- Diaconis and Freedman (1980) generalized *exchangeability* to *partial exchangeability* and extended de Finetti representation theorem.
- Suppose $B_0, B_1, B_2, ...$ are partially exchangeable and $\{0, 1\}$ valued. Then a representation analog of (1) holds with $g(\lambda_1, \lambda_2; \mathbf{z}) = \lambda_1^{t_{00}} \lambda_2^{t_{11}} (1-\lambda_1)^{t_{01}} (1-\lambda_1)^{t_{10}}$, where $t_{i,j}$ is the transition from $i$ to $j$ for $i, j = 0, 1$ of the sequence.
- Expanding the analog of representation (1):

$$f(\mathbf{z}|\boldsymbol{\pi}_m) = \sum_{j=0}^{t_{01}} \sum_{k=0}^{t_{10}} (-1)^{j+k} \binom{t_{01}}{j} \binom{t_{10}}{k} \pi_{t_{00}+j, t_{11}+k}, \qquad (10)$$

where $\boldsymbol{\pi}_m = \{\pi_{j,k} : j, k = 0, 1, ..., m\}$ with $\pi_{0,0} = 1$.

# Partially Exch. Models (Peng, Dang and Wang, '09))

- $f(\mathbf{z}|\boldsymbol{\pi}_m)$ is a mixture of Markov chains:
  $f(\mathbf{z}|\boldsymbol{\pi}_m) = \int_0^1\int_0^1 \lambda_1^{t_{00}}\lambda_2^{t_{11}}(1-\lambda_1)^{t_{01}}(1-\lambda_2)^{t_{10}}\, dQ(\lambda_1,\lambda_2)$.
- $\{\pi_{i,j} : i,j = 0,1,...,m\}$ are *rectangular completely monotone (RCM)*:

  $$(-1)^{r_1+r_2}\Delta_1^{r_1}\Delta_2^{r_2}a_{i,j} \geq 0, i,j = 0,1,...,r_1,r_2 = 0,...,m, r_1+r_2 \leq m, \tag{11}$$

  where $\Delta_1, \Delta_2$ are the (univariate) marginal forward difference operators, e.g., $\Delta_1 a_{i,j} = a_{i+1,j} - a_{i,j}$ etc.
- The binomial and Exchangeable binomial are special cases.
- The exchangeable trinomial and partially exchangeable binomial can be extended to multivariate cases.

# Parametric/Semiparametric Regression

Let $\mathbf{x}$ be a covariate vector such as the indicator of treatment/control groups, or death/alive. Let $W$ be a continuous covariate variable such as weight. In many situations, $p_{\mathbf{k}}$ can be interpreted as the probability of "$\mathbf{k}$ consecutive successes". We express the probability as $h_{\mathbf{k}}(\eta;\nu)$:

$$p_{\mathbf{k}} = h_{\mathbf{k}}(\eta;\boldsymbol{\nu}), \quad \mathbf{k} \in \mathcal{K}, \tag{12}$$

where $\eta$ is the systematic part either parametric $\eta = \boldsymbol{\beta}^{\top}\mathbf{x} + \gamma W$ or semiparametric $\eta = \boldsymbol{\beta}^{\top}\mathbf{x} + \rho(W)$ with regression parameter $\boldsymbol{\beta}$ and nonparametric smooth function $\rho$, and $\{h_{\mathbf{k}}(\eta;\nu) : \mathbf{k} \in \mathcal{K}\}$ are the link functions (the inverse link functions as used in generalized linear models) with $\boldsymbol{\nu} \in \mathbb{V}$ a nuisance parameter vector.

- Our approach generalizes the methodology in generalized linear models: *one* link function in GLM's is generalized to *several* link functions. The binomial, the negative binomial and the multinomial models in GLM's are generalized to the exchangeable binomial, the exchangeable negative binomial and the exchangeable multinomial.
- $\{h_{\mathbf{k}}(\eta;\nu) : \mathbf{k} \in \mathcal{K}\}$ must be CM so as to have probability density:

$$f(\mathbf{z}|\eta,\boldsymbol{\nu}) = \sum_{\mathbf{k}(\mathbf{z})\in\mathcal{K}} a_{\mathbf{k}}(\mathbf{z}) h_{\mathbf{k}(\mathbf{z})}(\eta;\boldsymbol{\nu}), \quad \mathbf{z} \in \mathbb{Z}, \tag{13}$$

**Theorem**

*Suppose the links* $\{h_{\mathbf{k}}(\theta) : \mathbf{k} \in \mathcal{K}\}$ $(h_{\mathbf{0}}(\theta) = 1)$ *are CM for each* $\theta$. *Then (13) is a probability density.*

# Univariate Completely Monotone Links.

Dang, Keeton and Peng ('09) gave a rich class of UCM links, including the incomplete Beta-, Gamma-, Normal-, and Poisson-binomial link; shown the class is closed under convex linear combinations, products, and composites; provided a forward model selection procedure about performing statistical inference on correlated binary data and overdispersed data.

Table: **Completely Monotonic Links.** $\theta = \theta$ **or** $(\theta_1, \theta_2)$**.**

| Name | Link $(j = 0, 1, ...)$ | Domain of $\theta$ |
|---|---|---|
| Ind-Bin | $\theta^j$ | $(0, 1)$ |
| MM-Bin | $\theta/(\theta + j)$ | $(0, \infty)$ |
| Beta-Bin | $B(\theta_1 + j, \theta_2)/B(\theta_1, \theta_2)$ | $(0, \infty)^2$ |
| Gamma-Bin | $(1 + \theta_2 j)^{-\theta_1}$ | $(0, \infty)^2$ |
| Poisson-Bin | $\exp(\theta(e^{-j} - 1))$ | $(0, \infty)$ |
| Normal-Bin | $2\exp((\theta j)^2/2)(1 - \Phi(\theta j))$ | $(0, \infty)$ |

Table: **Incomplete CM Links. $\theta = (\theta_1, \theta_2)$, or $(\theta_1, \theta_2, \theta_3)$, etc.**

| Name | Link ($j = 0, 1, ...$) | Domain of $\theta$ |
|---|---|---|
| Inc-A-Bin | $\frac{1}{j}\frac{\theta_2^j - \theta_1^j}{\ln\theta_2 - \ln\theta_1}$ | $[0,1]^2$ |
| Inc-MM-Bin | $\frac{\theta_3}{\theta_3+j}\frac{\theta_2^{j+\theta_3} - \theta_1^{j+\theta_3}}{\theta_2^{\theta_3} - \theta_1^{\theta_3}}$ | $(0,\infty)^3$ |
| Inc-Beta-Bin | $\frac{B(\theta_3+j;\theta_4)}{B(\theta_3;\theta_4)}\frac{F_{\text{beta}}(\theta_2;\theta_3+j,\theta_4) - F_{\text{beta}}(\theta_1;\theta_3+j,\theta_4)}{F_{\text{beta}}(\theta_2;\theta_3,\theta_4) - F_{\text{beta}}(\theta_1;\theta_3,\theta_4)}$ | $[0,1]^2 \times (0,\infty)^2$ |
| Inc-Gamma-Bin | $\frac{1}{(1+\theta_4 j)^{\theta_3}}\frac{\Gamma((\theta_4 j+1)\theta_2;\theta_3) - \Gamma((\theta_4 j+1)\theta_1;\theta_3)}{(\Gamma(\theta_2;\theta_3) - \Gamma(\theta_1;\theta_3))}$ | $(0,\infty)^4$ |
| Inc-Normal-Bin | $\exp(j^2\theta_3^2/2)\frac{\Phi(\theta_2/\theta_3+\theta_3 j) - \Phi(\theta_1/\theta_3+\theta_3 j)}{\Phi(\theta_2/\theta_3) - \Phi(\theta_1/\theta_3)}$ | $(0,\infty)^3$ |

# Multivariate Completely Monotone Links.

- RCM Links. Peng, Garner and Dang (2009) have presented four constructive methods to obtain RCM links, and shown that these links as a family are closed under convex linear combinations, products, and composites. A particular convenient construction of RCM links is to get RCM links from the rich family of UCM links. They have also provided a forward model selection method.
- SCM links.

**Theorem**

*Let $\varphi(s,t)$ be RCM. Then $h(s,t) = \gamma_1^s \ \gamma_2^t \ \varphi(s,t)$ is SCM where $0 \leq \ \gamma_1, \ \gamma_2$ and $\gamma_1 + \gamma_2 \leq 1$.*

Proof. See Ding (2007).

# Efficient Estimating Procedure.

Suppose we are available i.i.d. observations $\mathbf{Z}_j = (\mathbf{X}_j, W_j)$, $j = 1, ..., n$ of $(\mathbf{X}, W)$. Denote the individual log-likelihood function as a function of $\mathbf{Z}_i$ and $\eta, \boldsymbol{\nu}$ by

$$\ell_i(\eta|\boldsymbol{\nu}) = \log f(\mathbf{Z}_i|\eta;\boldsymbol{\nu}).$$

Write $\dot{\ell}_i, \ddot{\ell}_i, \dddot{\ell}_i$ the first, second, third derivative of $\ell_i$ with respect to $\eta$, so in particular $\dot{\ell}_i(\eta|\boldsymbol{\nu}) = (\dot{f}/f)(\mathbf{Z}_i|\eta;\boldsymbol{\nu})$ and $\ddot{\ell}_i(\eta|\boldsymbol{\nu}) = (\ddot{f}/f - (\dot{f}/f)^2)(\mathbf{Z}_i|\eta;\boldsymbol{\nu})$, where $\dot{f}(\mathbf{z}|\eta;\boldsymbol{\nu}) = (\partial f/\partial\eta)(\mathbf{z}|\eta;\boldsymbol{\nu})$ and $\ddot{f}(\mathbf{z}|\eta;\boldsymbol{\nu}) = (\partial^2 f/\partial\eta\partial)(\mathbf{z}|\eta;\boldsymbol{\nu})$, with

$$\frac{\partial^l f}{\partial\eta^l}(\mathbf{z}|\eta;\boldsymbol{\nu}) = \sum_{\mathbf{k}(\mathbf{z})\in\mathcal{K}} a_{\mathbf{k}(\mathbf{z})}(\mathbf{z})\frac{\partial^l h_{\mathbf{k}(\mathbf{z})}}{\partial\eta^l}(\eta;\boldsymbol{\nu}), \quad l = 0, 1, 2, 3.$$

# Estimating Nonparametric Part $\rho(w)$

Fix $w$ and $(\boldsymbol{\beta}, \boldsymbol{\nu})$, we estimate $\rho(w)$ by the local likelihood estimator $r = \hat{\rho}_n(w|\boldsymbol{\beta}, \boldsymbol{\nu})$ which solves the equation

$$\Phi_n(r|w, \boldsymbol{\beta}, \boldsymbol{\nu}) \equiv (nb)^{-1} \sum_{i=1}^{n} K_b(w - W_i)\, \dot{\ell}_i(\boldsymbol{\beta}^\top \mathbf{x}_i + r|\boldsymbol{\nu}) = 0, \quad (14)$$

where $K_b(\cdot) = K(\cdot/b)$ with $K$ a usual kernel and $b > 0$ a bandwidth.

# Estimating Regression Parameter $\beta$

We now assume that the nuisance parameter $\boldsymbol{\nu} = \boldsymbol{\nu}_0$ is known (otherwise take it to be the maximum profile likelihood estimate). Denote $\ell_i(\eta) = \ell_i(\eta|\boldsymbol{\nu}_0)$, $\hat{\rho}_n(W_i|\boldsymbol{\beta}) = \hat{\rho}_n(W_i|\boldsymbol{\beta}, \boldsymbol{\nu}_0)$, $\eta_{n,i}(\boldsymbol{\beta}) = \mathbf{X}_i^\top \boldsymbol{\beta} + \hat{\rho}_n(W_i|\boldsymbol{\beta})$, etc. The regression parameter $\boldsymbol{\beta}$ can by estimated by the maximum profile likelihood estimator $\hat{\beta}_n$ which solves the score equation

$$S_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \dot{\ell}_i(\eta_{n,i}(\boldsymbol{\beta}))\left(\mathbf{X}_i + \dot{\hat{\rho}}_n(W_i|\boldsymbol{\beta})\right) = 0,$$

where $\dot{\hat{\rho}}_n(\cdot|\boldsymbol{\beta}) = (\partial\hat{\rho}_n/\partial\boldsymbol{\beta})(\cdot|\boldsymbol{\beta})$ is a vector.

# Asymptotic Normality

Under certain regularity conditions (see Severini and Wong (1991) or Peng (2008)), the maximum profile likelihood estimator $\hat{\beta}_n$ has an asymptotic multinormal distribution with mean vector zero and convariance matrix $\Sigma$, i.e.,

$$\sqrt{n}(\hat{\beta}_n - \boldsymbol{\beta}_0) \Longrightarrow \mathcal{N}(0, \Sigma), \tag{15}$$

where the inverse $\Sigma^{-1}$ of $\Sigma = \Sigma(\beta_0, \rho_0)$ is given by

$$\Sigma^{-1} = -\mathbb{E}\left(\ddot{\ell}(\eta_0; \mathbf{Z})\left(\mathbf{x} + \dot{\rho}(W|\boldsymbol{\beta}_0)\right)^{\otimes 2} + \dot{\ell}(\eta_0; \mathbf{Z})\ddot{\rho}(W|\boldsymbol{\beta}_0)\right),$$

where $\eta_0 = \boldsymbol{\beta}_0^\top \mathbf{x} + \rho(W|\boldsymbol{\beta}_0)$. Estimate the above by

$$\hat{\Sigma}^{-1} = -\frac{1}{n}\sum_{j=1}^{n}\left\{\ddot{\ell}_j(\hat{\boldsymbol{\eta}}_j)\left(\mathbf{x}_j + \dot{\hat{\rho}}_n(W_j|\hat{\beta}_n)\right)^{\otimes 2} + \dot{\ell}_j(\hat{\boldsymbol{\eta}}_j)\ddot{\hat{\rho}}_n(W_j|\hat{\beta}_n)\right\},$$

where $\hat{\boldsymbol{\eta}}_j = \hat{\beta}_n^\top \mathbf{x}_j + \hat{\rho}_n(W_j|\hat{\beta}_n)$.

# Parametric Regression

**Parametric Regression.** The parametric regression $\eta = \mathbf{X}^\top \boldsymbol{\beta}$ can be considered as a special case of the semiparametric regressoin, $\rho(w) \equiv 0$. Because of the importance of parametric regression, we give here the asymptotic asymptotic normality of the maximum likelihood estimator $\hat{\beta}_n$ as follows:

$$\sqrt{n}(\hat{\beta}_n - \boldsymbol{\beta}_0) \Longrightarrow \mathcal{N}(0, \Sigma_0), \tag{16}$$

where the inverse $\Sigma_0^{-1}$ of $\Sigma_0(\boldsymbol{\beta}_0) = \Sigma(\beta_0, 0)$ is given by

$$\Sigma_0^{-1} = -\mathbb{E}\left(\ddot{\ell}(\boldsymbol{\beta}_0^\top \mathbf{x}; \mathbf{Z})\mathbf{x}^{\otimes 2}\right). \tag{17}$$

A plug-in estimator of $\Sigma_0$ can be obtained analogously.

**Algorithm for Semiparametric Regression.**

- Updating step for $\beta$:

$$\beta^{\text{new}} = (\widetilde{\mathcal{X}}^\top \mathcal{W} \widetilde{\mathcal{X}})^{-1} \widetilde{\mathcal{X}}^\top \mathcal{W} \widetilde{\mathbf{Z}},$$

where $\widetilde{\mathcal{X}} = (\mathcal{I} - \mathcal{S}_2)\mathcal{X}$, $\widetilde{\mathbf{Z}} = \widetilde{\mathcal{X}}\beta - \mathcal{W}^{-1}\mathbf{v}$ is an adjusted dependent variable with $\mathbf{v}_i = \dot{\ell}_i(\mathbf{x}_i^\top \beta + r_i)$ as the $i$-th component of vector $\mathbf{v}$, and $\mathcal{W} = \mathcal{W}_1 + \mathcal{W}_2$ with $\mathcal{W}_1 = \text{diag}(\ddot{\ell}_i(\mathbf{x}_i^\top \beta + r_i) : i = 1, ..., n)$ and $\mathcal{W}_2 = \text{diag}(\mathbf{v}^\top \mathcal{S}_3) - \mathcal{S}_3{}^\top \text{diag}(\mathbf{v})\mathcal{S}_2$.

- Updating step for $r_j$:

$$r_j^{\text{new}} = r_j - \frac{\sum_{i=1}^n \dot{\ell}_i(\mathbf{x}_i^\top \beta + r_j) K_b(W_i - W_j)}{\sum_{i=1}^n \ddot{\ell}_i(\mathbf{x}_i^\top \beta + r_j) K_b(W_i - W_j)}$$

- From the alogorithm it is evident that the parametric part of the model is updated by a parametric method with a nonparametrically modified design matrix $\widetilde{\mathcal{X}}$.

# Algorithm for Parametric Regression.

- Updating step for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}^{\text{new}} = (\mathcal{X}^\top \mathcal{W} \mathcal{X})^{-1} \mathcal{X}^\top \mathcal{W} \mathbf{Z},$$

where $\mathbf{Z} = \mathcal{X}\boldsymbol{\beta} - \mathcal{W}^{-1}\mathbf{v}$ is an adjusted dependent variable with $\mathbf{v}_i = \dot{\ell}_i(\boldsymbol{\beta}^\top \mathbf{X}_i)$ as the $i$-th component of $\mathbf{v}$, and $\mathcal{W} = \text{diag}(\ddot{\ell}_i(\boldsymbol{\beta}^\top \mathbf{X}_i) : i = 1, \ldots n)$ is a diagonal matrix.

# Model Assessment

• One way of assessing the adequacy of a model is to compare it with a saturated model — a more general model with the maximum number of parameters that can be estimated.

• The saturated model is the discretized form (3) with parameter vector $\mathbf{p} = \{p_{\mathbf{k}} : \mathbf{k} \in \mathcal{K}\}$.

• Estimate $\mathbf{p}$ by the MLE $\hat{\mathbf{p}} = \{\hat{p}_{\mathbf{k}} : \mathbf{k} \in \mathcal{K}\}$, so $\hat{\mathbf{p}}$ maximizes $L_n(\mathbf{p}) = \sum_{i=1}^n \log f(\mathbf{Z}_i|\mathbf{p}) = \sum_{i=1}^n \log \sum_{\mathbf{k}(\mathbf{Z}_i)\in\mathcal{K}} a_{\mathbf{k}(\mathbf{Z}_i)}(\mathbf{Z}_i) p_{\mathbf{k}(\mathbf{Z}_i)}$.

Differentiating w.r.t. $p_{\mathbf{j}}$ and setting the derivative to zero, $\hat{\mathbf{p}}$ solves

$$\frac{\partial L_n(\mathbf{p})}{\partial p_{\mathbf{j}}} = \sum_{i=1}^n \frac{a_{\mathbf{j}}(\mathbf{Z}_i)\mathbf{1}[\mathbf{k}(\mathbf{Z}_i) = \mathbf{j}]}{\sum_{\mathbf{k}(\mathbf{Z}_i)\in\mathcal{K}} a_{\mathbf{k}(\mathbf{Z}_i)}(\mathbf{Z}_i) p_{\mathbf{k}(\mathbf{Z}_i)}} = 0, \quad \mathbf{j} \in \mathcal{K},$$

subject to the constraints that $\mathbf{p}$ is completely monotone.

# The Deviance function

- The deviance is

$$D = 2(L_n(\hat{\mathbf{p}}) - L_n(\hat{\eta}, \hat{\boldsymbol{\nu}})),$$

where $L_n(\hat{\eta}, \hat{\boldsymbol{\nu}})$ denotes the log- likelihood of the model of interest given by

$$L_n(\hat{\eta}, \hat{\boldsymbol{\nu}}) = \sum_{i=1}^{n} \log f(\mathbf{Z}_i | \hat{\eta}_i, \hat{\boldsymbol{\nu}}) = \sum_{i=1}^{n} \log \sum_{\mathbf{k}(\mathbf{Z}_i) \in \mathcal{K}} a_{\mathbf{k}(\mathbf{Z}_i)}(\mathbf{Z}_i) h_{\mathbf{k}(\mathbf{Z}_i)}(\hat{\eta}_i, \hat{\boldsymbol{\nu}}).$$

- Under certain regularity assumptions, $D$ has an approximate non-central $\chi^2$. For parametric model, df $= \dim(\boldsymbol{\lambda}_K) - \dim(\boldsymbol{\beta}, \nu)$. For semiparametri model, df can be computed following the heuristic idea of Hastie and Tibshirani (1990, p. 155-158).

# The Deviance of the Exchangeable Binomial

- Data often consist of clusters of unequal sizes such as in developmental toxicity studies. In this case, complications arise, for example, it is difficult to specify a common model. Viewing cluster sizes as random, Bowman and George (1995) obtained the MLE of the parameters $\lambda_i$'s in the saturated EB model.
- Let $K$ be the max possible cluster size. Suppose data consist of randomly selected clusters with $m_i$ clusters of size $i, i = 1, ..., K$. Let $A_{r,m}$ be the number of clusters in data with cluster size $m$ and cluster sum $r$. Then the MLE $\hat{\lambda}_r$ of $\lambda_r$ is

$$\hat{\lambda}_r = \frac{1}{M} \sum_{m=r}^{K} \sum_{j=0}^{m-r} \frac{\binom{m-r}{j}}{\binom{m}{j}} A_{m-j,m}, \; r = 1, ..., K, \qquad (18)$$

where $M = m_1 + ... + m_K$ is the total number of clusters.

# Parametric Regression

In this case, $\eta = \mathbf{x}^\top \boldsymbol{\beta} + W\gamma$, the deviance for the **EB** model is

$$D(\tilde{\boldsymbol{\eta}}_n) = 2(L_n(\hat{\boldsymbol{\lambda}}_K) - L_n(\tilde{\boldsymbol{\eta}}_n|\mathbf{h})), \quad \text{where} \tag{19}$$

$$L_n(\hat{\boldsymbol{\lambda}}_K) = \sum_{i=1}^{n} \log \left( \sum_{k=0}^{m_i - Y_i} (-1)^k \binom{m_i - Y_i}{k} \hat{\lambda}_{Y_i+k} \right), \tag{20}$$

is the log- likelihood of the saturated model, and

$$L_n(\tilde{\boldsymbol{\eta}}_n|\mathbf{h}) = \sum_{i=1}^{n} \log \left( \sum_{k=0}^{m_i - Y_i} (-1)^k \binom{m_i - Y_i}{k} h_{Y_i+k}(\tilde{\eta}_i) \right), \tag{21}$$

is the log- likelihood of the parametric model of interest under link $\mathbf{h}$, where $\tilde{\boldsymbol{\eta}}_n = (\tilde{\eta}_1, ..., \tilde{\eta}_n)$ with $\tilde{\eta}_i = X_i^\top \tilde{\boldsymbol{\beta}}_n + W_i \tilde{\gamma}_n$. Under the usual regularity assumptions, $D$ has an approximate non-central $\chi^2$ distribution with degrees of freedom $\dim(\boldsymbol{\lambda}_K) - \dim(\boldsymbol{\beta})$ and non-centrality $v = 2(L_n(\boldsymbol{\lambda}_K) - L_n(\boldsymbol{\beta}|\mathbf{h}))$.

# Parametric Regression

Consider a null hypothesis versus an alternative,

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 = (\beta_1, ..., \beta_q)^\top \quad \text{vs} \quad H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1 = (\beta_1, ..., \beta_p)^\top.$$

where $q < p < n$. One can test $H_0$ versus $H_1$ using the difference of the deviance statistics

$$\begin{aligned} R &= 2[L_n(\hat{\boldsymbol{\lambda}}_K) - L_n(\hat{\boldsymbol{\beta}}_0|\mathbf{h})] - 2[L_n(\hat{\boldsymbol{\lambda}}_K) - L_n(\hat{\boldsymbol{\beta}}_1|\mathbf{h})] \\ &= 2[L_n(\hat{\boldsymbol{\beta}}_1|\mathbf{h}) - L_n(\hat{\boldsymbol{\beta}}_0|\mathbf{h})]. \end{aligned}$$

Under certain regularity conditions, $R$ has the $\chi^2$ distribution with $p - q$ degrees of freedom under the null hypothesis.

# Semiparametric Regression

• In this case, $\eta = \mathbf{x}^\top \boldsymbol{\beta} + \rho(W)$, and we want to compare the semiparametric model with the parametric model. The deviance is given by $D(\hat{\boldsymbol{\eta}}_n)$, an analog of (19) with $\hat{\eta}_i = X_i^\top \boldsymbol{\beta}_n + \hat{\rho}_n(W_i)$ and $L_n(\hat{\boldsymbol{\eta}}_n|\mathbf{h})$ replacing the log- likelihood in (21).

• The reference distribution of $D$ is still taken to be $\chi^2$ but with an approximate df calculated by $\dim(\boldsymbol{\lambda}_K) - \mathrm{df}^{\mathrm{err}}(\hat{\boldsymbol{\eta}}_n)$, where $\mathrm{df}^{\mathrm{err}}(\hat{\boldsymbol{\eta}}_n)$ is computed by,

$$\mathrm{df}^{\mathrm{err}}(\hat{\boldsymbol{\eta}}_n) = n - \mathbf{tr}(\mathcal{R}), \tag{22}$$

following Hastie and Tibshirani (1990).

• Denote the adjusted dependent variable $\mathbf{z} = \hat{\boldsymbol{\eta}}_n - A\hat{\mathbf{u}}_n$, where $A$ is the estimated information matrix and $\hat{\mathbf{u}}_n = \partial L_n / \partial \hat{\boldsymbol{\eta}}_n$. If at convergence of the iterative estimation $\hat{\boldsymbol{\eta}}_n = \mathcal{R}\mathbf{z}$ with a linear operator $\mathcal{R}$, then

$$D(\hat{\boldsymbol{\eta}}_n) \approx (\mathbf{z} - \hat{\boldsymbol{\eta}}_n)^\top \widetilde{\mathcal{W}} (\mathbf{z} - \hat{\boldsymbol{\eta}}_n)^\top$$

which has approximately

$$\mathrm{df}^{\mathrm{err}}(\hat{\boldsymbol{\eta}}_n) = \mathbb{E}(D(\hat{\boldsymbol{\eta}}_n)) = n - \mathbf{tr}(2\mathcal{R} - \mathcal{R}^\top \widetilde{\mathcal{W}} \mathcal{R} \widetilde{\mathcal{W}}^{-1}). \qquad (23)$$

In practice, the computation of the trace can be rather difficult. We use the simpler approximation (22), which were correct if $\mathcal{R}$ were a projection operator and $\widetilde{W}$ were the identity matrix.

- Müller (2000) gave the following a workable approximation

$$\mathcal{R} = \widetilde{\mathcal{X}}(\widetilde{\mathcal{X}}^\top \mathcal{W} \widetilde{\mathcal{X}})^{-1} \widetilde{\mathcal{X}}^\top \mathcal{W}(\mathcal{I} - \mathcal{S}_2) + \mathcal{S}_2,$$

where $\widetilde{\mathcal{X}}$, $\mathcal{S}_2$ and $\mathcal{W}$ are previously given.

- Now, for the comparison of the semiparametric $\hat{\boldsymbol{\eta}}_n$ and the parametric $\tilde{\boldsymbol{\eta}}_n$, the test statistic can be expressed by

$$R = D(\tilde{\boldsymbol{\eta}}_n) - D(\hat{\boldsymbol{\eta}}_n)$$

and should follow approximately the $\chi^2$ distribution with $\mathrm{df}^{\mathrm{err}}(\hat{\boldsymbol{\eta}}_n) - \mathrm{df}^{\mathrm{err}}(\tilde{\boldsymbol{\eta}}_n)$ degrees of freedom.

# Application I: Toxicological Data

Evaluations of reproductive and developmental toxicity studies are important components of the safety evaluation of environmental agents that could possibly harm developing fetuses or have adverse effects on reproductive health. These studies are generally conducted in laboratory animals. In developmental toxicity studies, for example, pregnant rats or mice are randomly assigned to expose to varying dose levels of toxin. These animals are sacrificed before their term and contents of the uterus were examined. The primary toxicity outcomes of interest are deaths/resorptions (early death), malformations (binary variable), and fetal body weights. These endpoints represent different degrees of responses to a toxic agent and occur in a dose-related manner (Kimmel and Gaylor, 1988). The goal of such studies is to assess the association between exposure level to the toxic substance and the incidence of developmental problem.

In this application, the data used comes from a study which was carried out at the National Center for Toxicological Research, U.S. Food and Drug Administration, of the herbicide 2,4,5-trichlorophenoxyacetic acid (2,4,5-T). One outbred (CD-1) and four inbred (C57BL/6, C3H/He, BALB/C, and A/JAX) strains of mice were tested with six or seven dose levels. Each female was mated and exposed to one of seven dose groups daily. The exposure levels were 0, 15, 20, 25, 30, 45, and 60 mg/kg of the herbicide. We consider only one strain of the mice A/JAX. Table 3 summarizes the A/JAX data.

Table: Summary Statistics of the A/JAX Data. nMalf=number of malformations, nImp=number of implants, pMalf=proportion of malformations, nLitt=number of litters, avgSiz=average litter size, sdSiz=standard deviation of litter size, avgWt=average weight.

| Dose$_{(mg/kg)}$ | nMalf | nImp | pMalf | nLitt | avgSiz | sdSiz | avgWt |
|---|---|---|---|---|---|---|---|
| 0 | 72 | 728 | 0.099 | 89 | 8.180 | 1.831 | 58.43 |
| 15 | 126 | 658 | 0.191 | 86 | 7.651 | 2.096 | 56.57 |
| 20 | 79 | 431 | 0.183 | 56 | 7.696 | 1.726 | 55.26 |
| 25 | 85 | 293 | 0.290 | 40 | 7.325 | 1.927 | 55.95 |
| 30 | 321 | 553 | 0.580 | 76 | 7.276 | 1.801 | 50.45 |
| 45 | 118 | 157 | 0.752 | 33 | 4.758 | 2.586 | 47.34 |
| 60 | 19 | 19 | 1.000 | 9 | 2.111 | 1.965 | 39.30 |

In Table 3, it is clear that as the dose level increases, the proportion of malformations increases, the mean fetal weight and the litter size decrease. The decreasing tendency of weight can also be observed from Fig. 1 which gives the curves of the kernel estimates of the conditional density of weight given the dose level. Here the bandwidth selection in Fig. 1 implemented the methods of Sheather and Jones (1991) using the pilot estimation of derivatives.

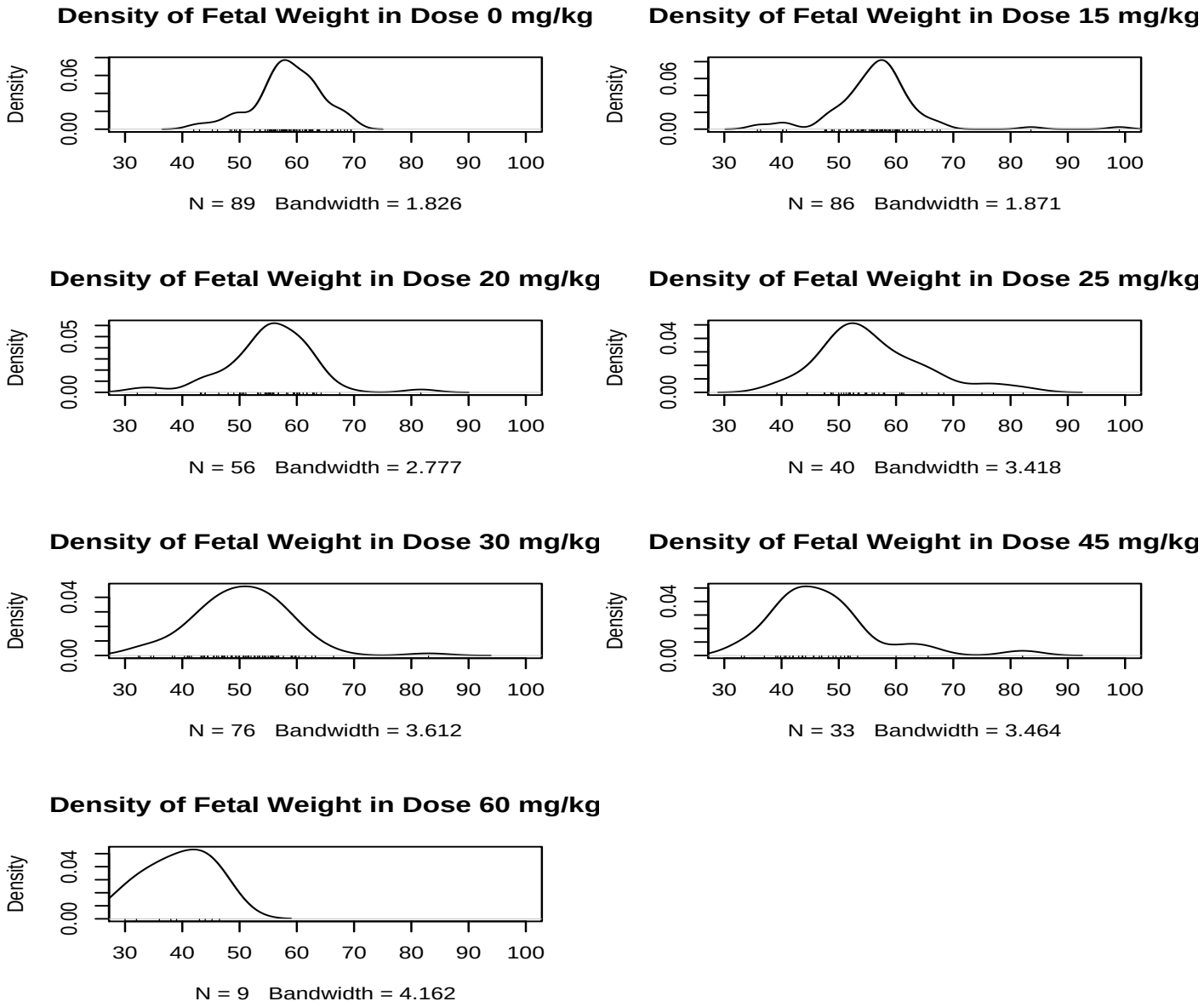

Density of Fetal Weight in Dose 0 mg/kg
Density
0.00
0.06
30 40 50 60 70 80 90 100
N = 89 Bandwidth = 1.826
Density of Fetal Weight in Dose 15 mg/kg
Density
0.00
0.06
30 40 50 60 70 80 90 100
N = 86 Bandwidth = 1.871
Density of Fetal Weight in Dose 20 mg/kg
Density
0.00
0.05
30 40 50 60 70 80 90 100
N = 56 Bandwidth = 2.777
Density of Fetal Weight in Dose 25 mg/kg
Density
0.00
0.04
30 40 50 60 70 80 90 100
N = 40 Bandwidth = 3.418
Density of Fetal Weight in Dose 30 mg/kg
Density
0.00
0.04
30 40 50 60 70 80 90 100
N = 76 Bandwidth = 3.612
Density of Fetal Weight in Dose 45 mg/kg
Density
0.00
0.04
30 40 50 60 70 80 90 100
N = 33 Bandwidth = 3.464
Density of Fetal Weight in Dose 60 mg/kg
Density
0.00
0.04
30 40 50 60 70 80 90 100
N = 9 Bandwidth = 4.162

Figure:

• Statistically, responses from a continuous variable (e.g. fetal weight) is more sensitive to treatment than responses from discrete variables such as presence or absence of a particular malformation.
• We include the factor of fetal weight as a covariate in the model and express it in the systematic part through a smooth nonparametric function of fetal weight, while the relationship with the effect of dose level is modeled as additive. With these considerations in mind, we express the probability of "consecutive sucesses" as *dose effect* + *smoth function of fetal weight* through completely monotone links:

$$\text{prob of } \text{“}k \text{ consecutive successes”} = \text{dose effect} + \rho(\text{fetal wight})$$

This link takes the following simple form,

$$h_j(\eta) = e^{\eta}/(e^{\eta} + j), \quad j = 0, 1, ...$$

Here we express the systematic part $\eta$ in the following three forms,

$$\eta = \alpha + \beta D, \tag{24}$$

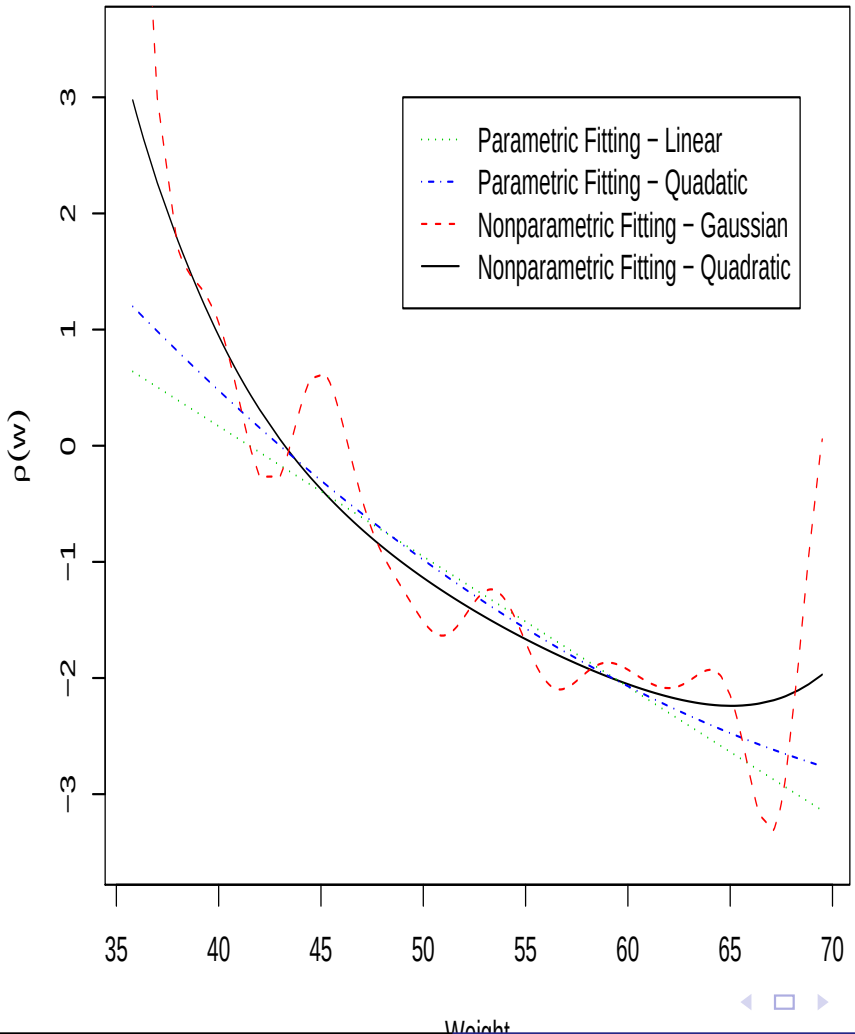
$$\eta = \alpha + \beta D + \gamma W, \tag{25}$$

$$\eta = \beta D + \rho(W), \tag{26}$$

where $D$ represents dose level and $W$ fetal weight, $\alpha$, $\beta$ and $\gamma$ are regression parameters, and $\rho$ is a unknown nonparametric smooth function.

Table: The Observed and Expected Number of Responses, Log-likelihood and Parameter Estimates under Different Models

| Dose(mg/kg) | Observed | Model (24) | Model (25) | Model (26) |
|---|---|---|---|---|
| 0 | 72 | 99 | 106 | 100 |
| 15 | 126 | 182 | 185 | 176 |
| 20 | 79 | 146 | 148 | 143 |
| 25 | 85 | 120 | 116 | 111 |
| 30 | 321 | 266 | 302 | 316 |
| 45 | 118 | 109 | 107 | 115 |
| 60 | 19 | 16 | 19 | 17 |
| Log-likelihood | | -667.09 | -609.66 | -610.34 |
| $\hat{\beta}$ | | 5.931 (??) | 4.242 (??) | 4.174 (.602) |


Parametric Fitting – Linear
Parametric Fitting – Quadatic
Nonparametric Fitting – Gaussian
Nonparametric Fitting – Quadratic
ρ(w)
3
2
1
0
−1
−2
−3
35
40
45
50
55
60
65
70
Weight

# Application II: Analyzing Exchangeable Binary Data (X. Dang, S.L. Keeton and H. Peng, *Stat. Med.*, 2009)

Given litter size $M$, we have observation $Y$ from parsimonious $\mathbf{EB}(\theta, M; \mathbf{h})$, associated with a covariate $X$. Typically $X$ is a vector of such as dose, weight, etc. Let $(Y_i, M_i, X_i), i = 1, ..., n$ be observations of $(Y, M, X)$. Write $\boldsymbol{\theta}$ as $(\theta, \vartheta)$, where $\theta \in \mathbb{R}$ is a parameter of interest, while $\vartheta$ is treated as a nuisance parameter. We model

$$\lambda_j = h_j(\tilde{\boldsymbol{\beta}}^\top X; \vartheta), \quad j = 0, 1, 2, ..., M.$$

# Model Selection

**Model Selection.**

1. Start from simple one-parameter links such as MM, MM-log, one-parameter Gamma, one-parameter Gamma-log-bin, Poisson-log, positive stable link, etc. By AIC or BIC, the best two or three models are selected.

# Model Selection

**Model Selection.**

1. Start from simple one-parameter links such as MM, MM-log, one-parameter Gamma, one-parameter Gamma-log-bin, Poisson-log, positive stable link, etc. By AIC or BIC, the best two or three models are selected.
2. Next, consider large two-parameter models which should include the previously selected models as sub-models. At this step, may look at models resulted from linear combination, power composite, and nonnegative polynomial composite of previous ones. Test the significance between the large and reduced model by the asymptotic likelihood ratio test (LRT).

# Model Selection

**Model Selection.**

1. Start from simple one-parameter links such as MM, MM-log, one-parameter Gamma, one-parameter Gamma-log-bin, Poisson-log, positive stable link, etc. By AIC or BIC, the best two or three models are selected.
2. Next, consider large two-parameter models which should include the previously selected models as sub-models. At this step, may look at models resulted from linear combination, power composite, and nonnegative polynomial composite of previous ones. Test the significance between the large and reduced model by the asymptotic likelihood ratio test (LRT).
3. R codes of link functions are available at `http://www.olemiss.edu/~xdang`.

Table: **Fitting The E2 Data**. Observed: $(p, q) = (0.110, 0.640)$.

| Models (npr) | $\hat{p}(s.d.)$ | $\hat{\phi}(s.d.)$ | $\hat{q}$ | $-2\log L$ | AIC | BIC |
|---|---|---|---|---|---|---|
| Bin (1) | 0.113(.006) | 0.000 | | 765.6 | 767.6 | 767.9 |
| Correlated Bin (2) | 0.131(.010) | 0.073(.012) | | 720.5 | 724.5 | 731.2 |
| Beta-Bin (2) | 0.112(.009) | 0.101(.017) | 0.612 | 689.8 | 693.8 | 700.5 |
| Two Bin (3) | 0.111 | 0.114 | | 682.4 | 688.4 | 698.5 |
| Three Bin (5) | 0.121 | 0.101 | | 679.7 | 689.7 | 706.5 |
| Beta-Bin with Bin (4) | 0.135 | 0.189 | | 680.2 | 688.2 | 701.6 |
| Kuk's Q-power (2) | 0.119 | 0.209 | 0.648 | 687.1 | 691.1 | 697.8 |
| Kuk's P-power (2) | 0.109 | 0.080 | 0.595 | 698.8 | 702.8 | 709.5 |
| Gamma-Bin with $\theta_2 = 1$(1) | 0.118(.0015) | 0.191(.0009) | 0.543 | 697.8 | 699.8 | 700.1 |
| Gamma-Bin (2) | 0.110(.0087) | 0.093(.0181) | 0.619 | 679.9 | 683.9 | **684.5** |
| Inc. Gamma-Bin (4) | 0.109(.0118) | 0.101(.0301) | **0.633** | **675.2** | **683.2** | 696.6 |
| Piecewise-Flogit (1) | 0.111(.0154) | 0.112(.0362) | 0.601 | 680.8 | **682.8** | **683.1** |
| Piecewise-Flogit Power (2) | 0.111(.0175) | 0.112(.0411) | 0.601 | 680.8 | 684.8 | 691.5 |
| Inc. Beta-Bin (4) | 0.110(.0055) | 0.102(.0384) | 0.632 | 676.4 | 684.4 | 697.8 |
| Inc. A-Bin (2) | 0.115(.0085) | 0.096(.0172) | 0.624 | 681.6 | 685.6 | 692.3 |

The estimated response probability $\hat{p}$, intra-litter correction $\hat{\phi}$, probability $\hat{q}$ of the affected litters, along with negative twice log-likelihood, AIC and BIC. The upper, middle, and lower table are from Brooks *et* al. (1997) and Brooks (2001), Kuk (2005), and the proposed framework, respectively. The standard deviations are included when they are available. Highlighted are the optimal models.

# Application III: a full likelihood procedure of exch neg. bin. for analyzing overdispersed count data (H. Peng, G.J. Rayner, X. Wang, F.Tan, JSPI, '09)

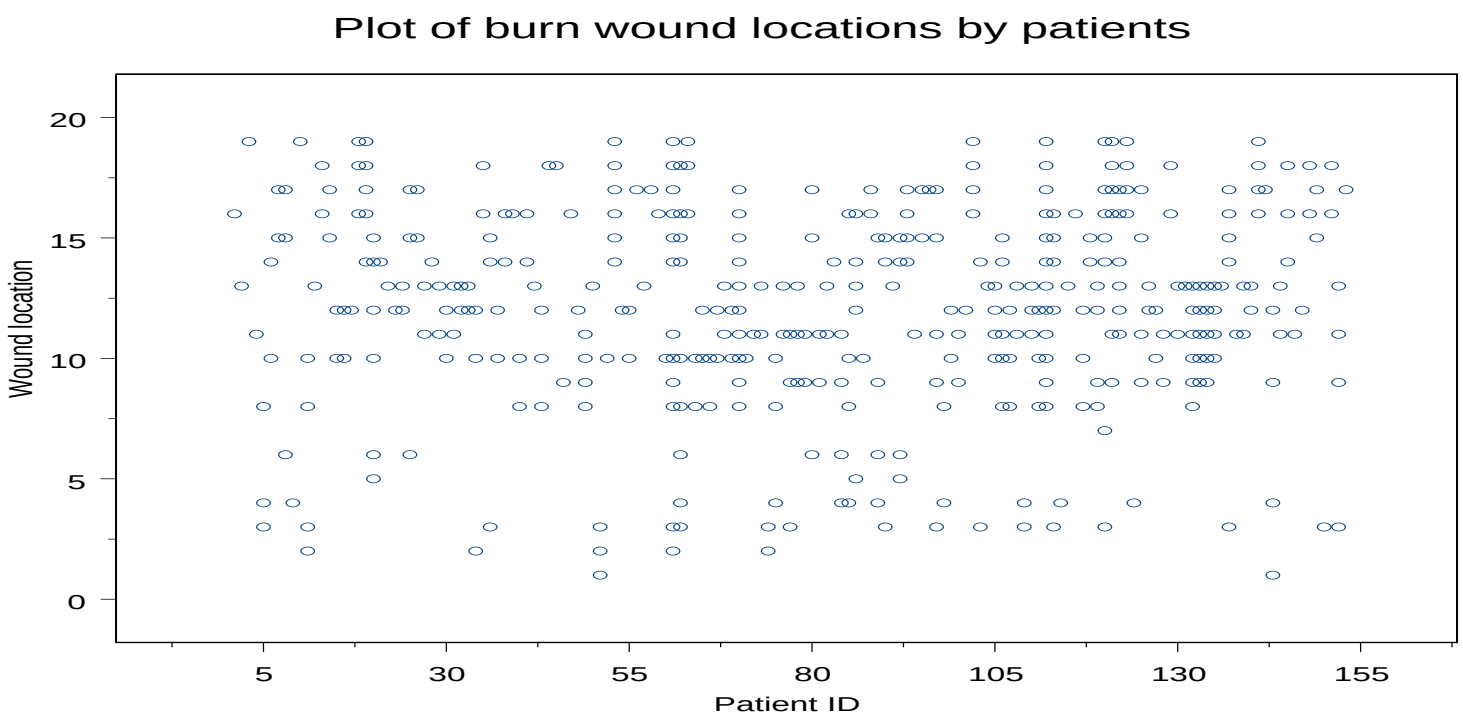



Figure:

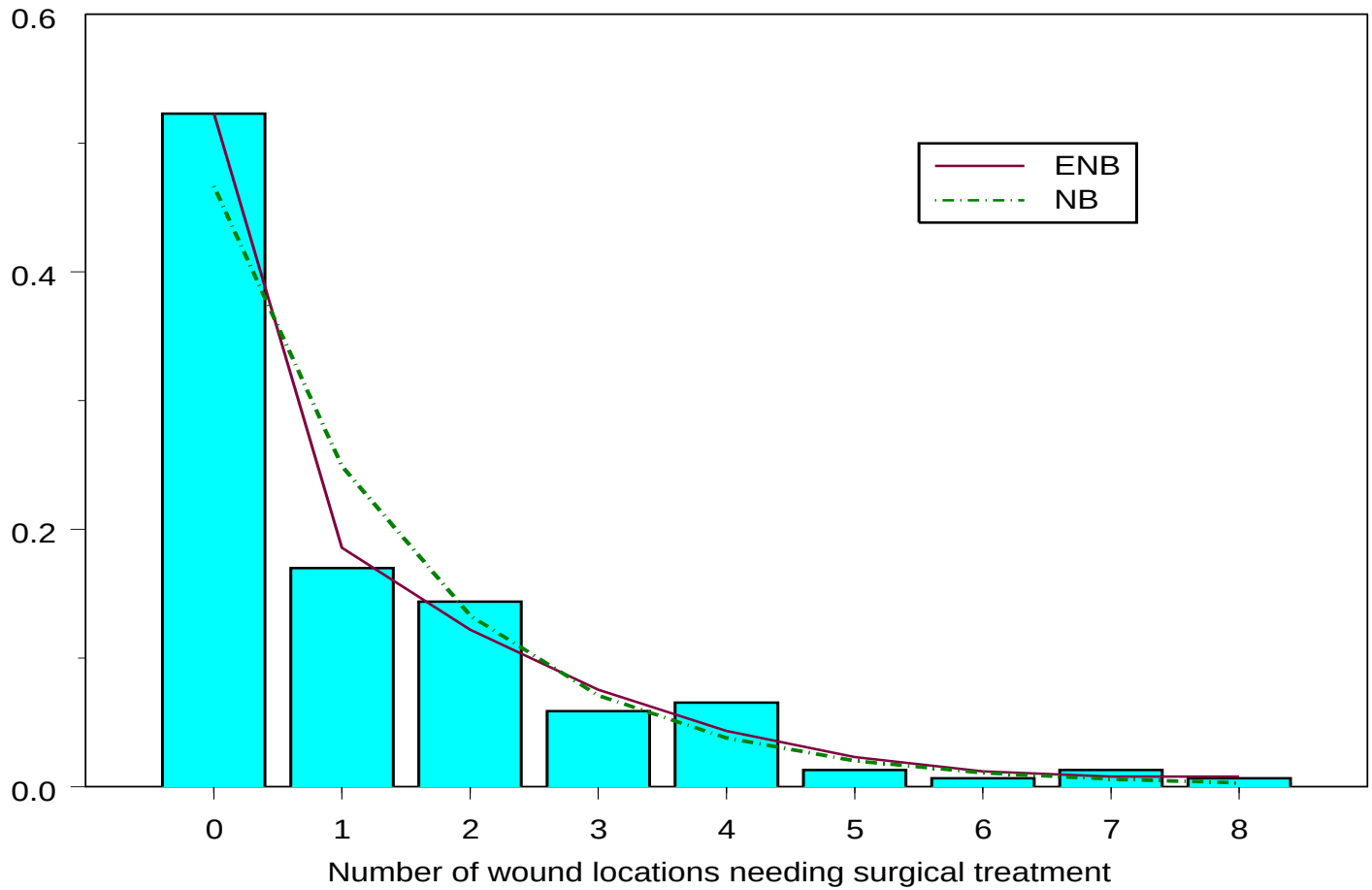



Figure: The histogram of the frequencies of wound locations in need of

# Application IV: Analyzing Partial Exchangeable Data (Peng, Garner and Dang ('09))

Bladder Cancer Study

- Measurements on bladder cancer

$$B = \begin{cases} 1, & \text{cancer} \\ 0, & \text{normal} \end{cases} \qquad X = \begin{cases} 1, & \text{treatment} \\ 0, & \text{control} \end{cases}$$

- All patients had bladder tumors when they entered the trial and the tumors were removed transurethrally.
- Reexamined every 3 months, any tumors recurrences were removed.
- Maximum sequence length: 12
- Sample size: 73

# The Results of Bladder Cancer Data

Table: **Estimated & Observed Transition Probabilities for the Bladder Cancer Data.**

| | Control | | Treatment | |
|---|---|---|---|---|
| | $P_{00}$(s.d.) | $P_{11}$(s.d.) | $P_{00}$(s.d.) | $P_{11}$(s.d.) |
| Observed | 0.829 | 0.374 | 0.942 | 0.333 |
| $\text{Ga}_s(\alpha_1, \beta_1) * \text{Ga}_t(\alpha_2, \beta_2)$ | 0.788(.036) | 0.311(.057) | 0.903(.027) | 0.264(.087) |
| $\text{Ga}_s(\alpha_1, \beta_1) * \text{MM}_t(\alpha_2, \beta_2)$ | 0.788(.036) | 0.362(.064) | 0.903(.027) | 0.307(.098) |
| $\textbf{Bin}_s(\alpha_1, \beta_1) * \textbf{MM}_t(\alpha_2, \beta_2)$ | **0.829(.005)** | **0.361(.062)** | **0.943(.004)** | **0.320(.098)** |
| $\text{MM}_s(\alpha_1, \beta_1) * \text{MM}_t(\alpha_2, \beta_2)$ | 0.748(.036) | 0.362(.064) | 0.930(.016) | 0.308(.094) |
| $\text{Biga}_{s,t}(\alpha_1, \beta_1, \alpha_2, \beta_2, \theta_3, \rho)$ | 0.787(.036) | 0.312(.058) | 0.903(.027) | 0.264(.087) |
| $\text{Ga}_s(\alpha_1, \beta_1, \theta) * \text{MM}_t(\alpha_2, \beta_2)$ | 0.785(.041) | 0.362(.064) | 0.901(.030) | 0.307(.098) |
| $\text{Kbiga}_{s,t}(\alpha_1, \beta_1, \alpha_2, \beta_2, \rho, v = 1)$ | 0.749(.036) | 0.365(.064) | 0.930(.016) | 0.307(.094) |
| $\text{Kbiga}_{s,t}(\alpha_1, \beta_1, \alpha_2, \beta_2, v, \rho = 0)$ | 0.753(.195) | 0.412(.145) | 0.926(.109) | 0.351(.162) |

**THANKS**

THANKS THANKS

*THANKS THANKS THANKS*

*THANKS THANKS THANKS THANKS*

thanks thanks thanks thanks thanks thanksTHANKS