Identifiability of Bivariate Mixtures: An Application to Infant Mortality Models

Eric Y. Frimpong[2], Timothy B. Gage[1,2], Howard Stratton[2].
[1]Department of Anthropology, University at Albany, SUNY.
[2]Department of Epidemiology and Biostatistics, University at Albany, SUNY.

Abstract

Identifiability of the parameters for a Mixture of Bivariate Densities (MBD) in the form

$$f(x,y;\phi,\theta,\pi) =$$
$$\pi f(y|x;\phi_1) f(x;\theta_1) + (1-\pi) f(y|x;\phi_2) f(x;\theta_2)$$

is considered. Particular attention is given to $\theta_1 \neq \theta_2$ (i.e. marginal of *x* is a nondegenerate mixture). Characterzations of identifiability that includes extensions of Hennig (2000), Hunter (2007) and Gage (2004) models are given. These identified models are applied to characterize latent subpopulations related to infant mortality and survival in several NCHS linked Birth/Death data sets.

1. Introduction

Mixture models are widely applied due to their ease of interpretation. Researchers often interpret the parameters of the model by viewing the fitted components as distinct clusters that form heterogeneous subpopulations within the population.
Indeed, populations are unlikely to be homogeneous due to unobservable genetic, environmental and/or social factors. Factors discussed in the literature are age, gender, species, geographical origin and cohort status. (McLachlan and Peel, 1996).

2. The MBD

Mixture of Bivariate Densities is given by the joint density:

$$f(x,y;\phi,\theta,\pi) = \pi f(x,y;\phi_1,\theta_1) + (1-\pi) f(x,y;\phi_2,\theta_2)$$

where the distribution parameters (θ_1,ϕ_1) and (θ_2,ϕ_2) are the associated parameters characterizing the two component densities with weights π and $1-\pi$ respectively.
Taking *x* as a covariate and *y* as the dependent variable, the MBD format used here allows the parameters to be separated as follows,

$$f(x,y;\phi,\theta,\pi) = \pi f(y|x;\phi_1) f(x;\theta_1) + (1-\pi) f(y|x;\phi_2) f(x;\theta_2) \qquad (1)$$

where $(x,y) \in \square$ are observables, $\theta = (\theta_1,\theta_2)$ denotes the parameters of the components associated with covariate *x* (the marginal of *x*). The parameters associated with the variable *y* are $\phi = (\phi_1 = (\beta_1,\alpha_1), \phi_2 = (\beta_2,\alpha_2))$ where $\beta = (\beta_1,\beta_2)$ denotes the regression coefficients for *y* in the respective components and $\alpha = (\alpha_1,\alpha_2)$ are other distributional parameters. Let $\omega = (\pi,\theta,\phi)^T \in \Omega$ denote the vector of all the unknown parameters in the MBD. The estimation and the consistency of the parameters of mixture models such as the one described above are known to be problematic. Estimation procedures may not be well-behaved if a model is not identifiable.

2.1 Identifiability

The problem of identifiability is basic to statistical methods and data analysis. In the past four decades various attempts have been made to build on the identifiability for mixtures which goes back to the papers by Teicher. (Teicher 1961, 1963).Various treatments of identifiability in the context of finite mixture models with specified distributions have been considered. (Yakowitz and Spragins 1968, Blischke 1962, Ord 1972, Robbins and Pitman, 1949, Kent, 1983, Gerard Gory, 1998, Holzman et.al, 2006, Jiang and Tanner, 1999).

2.1.1 Definition

Let $\{ f(x,y;\omega = (\pi,\theta,\phi), \omega \in \Omega) \}$ be a MBD parametric family, then it is said to be identifiable for $\omega \in \Omega$ if for all allowed *x* and *y*
$f(x,y;\omega) \equiv f(x,y;\tilde{\omega})$ implies up to labelling that
$\pi = \tilde{\pi}$, $\theta = \tilde{\theta}$ and $\phi = \tilde{\phi}$.
Estimation procedures for the MBD may have no unique parameters to estimate if the model is not identifiable. To achieve identifiability one may need to impose restrictions which may or may not have natural interpretations.

2.1.2 Natural restrictions

Before considering the Mixture of Bivariate Densities (MBD) in detail we make the following assumptions

i) $(\theta_1,\phi_1) \neq (\theta_2,\phi_2)$

ii) $(\frac{1}{2} < \pi < 1)$, π is the mixing proportion

The first assumption is to make sure that there are two unique components. The second assumption resolves the issue of non uniqueness of labelling.

As cautioned by McLachlan and Peel (1996), under these natural restrictions, the logistic regression mixture model may not be identifiable without additional "unrealistic" restrictions on the covariates. Conditions for which the parameters of the MBD with a logistic regression submodel identified are given in the main proposition.

The above MBD can be expressed in different forms. Though *x* and *y* are observable, the parameters and the cluster memberships are not.

2.2 Case 1: $\theta_1 = \theta_2$ (Homogeneous *x* marginal)

If we assume that $\theta_1 = \theta_2 = \tilde{\theta}$ in the above MBD model as expressed in equation (1), the distribution $f(x;\theta)$ in the MBD factors out and yields a Clusterwise Regressions model Hennig (2000):

$$f(x,y;\phi,\theta=(\tilde{\theta},\tilde{\theta}),\pi)$$
$$= \pi f(y|x;\phi_1) f(x;\tilde{\theta}) + (1-\pi) f(y|x;\phi_2)) f(x;\theta$$
$$= (\pi f(y|x;\phi_1) + (1-\pi) f(y|x;\phi_2)) f(x;\tilde{\theta}) \quad (2)$$

Thus the latent clusters are characterized by a mixture of *y* conditional densities with corresponding different mixture relations. For example, the mixture of conditional densities $f_y(y|x,\phi) = \pi f(y|x;\phi_1) + (1-\pi) f(y|x;\phi_2)$ leads to other mixtures of interest such as mixture of means (regressions) and possibly of variances. For each cluster it is often assumed that the means are linear regressions with distinct parameters

$\phi = (\phi_1 = (\beta_1,\alpha_1), \phi_2 = (\beta_2,\alpha_2))$ defining distinct clusters. More particularly:

***Theorem* 1**

If $\theta_1 = \theta_2$ in (1) then (i)

$f(x,y;\phi,\theta,\pi) = (\pi f(y|x;\phi_1) + (1-\pi) f(y|x;\phi_2)) f(x;\theta)$

(ii) $f(y|x,\phi) = \pi f(y|x;\phi_1) + (1-\pi) f(y|x;\phi_2)$ is a mixture of conditional densities (mixing regressions).

(iii) $E(y|x) = \pi\mu_1(x) + (1-\pi)\mu_2(x)$

(iv) $V(y|x) = \pi\sigma_1^2(x) + (1-\pi)\sigma_2^2(x)$ if and only if

$\mu_1(x) = \mu_2(x) \quad \forall x$.

2.3 Case 2: $\theta_1 \neq \theta_2$ (Nonhomogeneous *x* marginal)

If $\theta_1 \neq \theta_2$ in the above equations then $f(x;\theta_1) \neq f(x;\theta_2)$ and we have two distinct *x* marginal densities set apart by the parameters θ_1 and θ_2.

To better understand the nondegenerate mixture in this case we "sum out" the other variable *y*.

Thus $f(x;\theta) = \pi f(x;\theta_1) + (1-\pi) f(x;\theta_2)$

and the density $f(x;\theta)$ determines the latent clusters. Unlike Hennig's (2000) case the mixing proportion π is determined by the marginal of *x*. The parameters

$\theta_1 = (\mu_1,\sigma_1), \theta_2 = (\mu_2,\sigma_2)$

of the covariate *x* define distinct clusters. As in the case of Gage's model, the latent clusters are determined by the density/distribution of birth weight and/or gestational age.

3. Identifiability of the Models

The main objective is to explore the Identifiability of MBD as expressed in several special forms. The next step is to determine when an identifiable MBD is still identifiable for parameters associated with exogenous variables introduced as covariates for the MBD parameters. These exogenous variables may be discrete or continuous.

As pointed out earlier, estimation of parameters of models such as MBD is only meaningful if the parameters π, θ and ϕ are uniquely determined. There has been extensive work done on the identifiability of mixture models pertaining to latency. As pointed out by Hennig (1996, 2000) and Hunter (2007), there are still no clear or simple necessary and sufficient conditions for the identifiability of finite mixtures. For instance, Hennig investigated the identifiability of the parameters of models for data generated by different linear regression distributions with Gaussian errors. He concluded that such models are identifiable under additional restrictions.

Hunter et. al investigated the Identifiability of finite symmetric mixtures from a family of shifts. They noted that beyond certain numbers of component mixtures the identifiability of their models becomes complicated. We give a general proposition and provide proof (see appendix) that shows when the two general forms of a MBD are identified.

3.1 Definition

The distribution function *f* is said to be Two-mixing identified for $\pi, \tilde{\pi} > \frac{1}{2}$ iff

$$\pi f(y;\eta_1) + (1-\pi) f(y;\eta_2) = \tilde{\pi} f(y;\tilde{\eta}_1) + (1-\tilde{\pi}) f(y;\tilde{\eta}_2)$$
$$\Rightarrow \pi = \tilde{\pi}, \quad \eta_1 = \tilde{\eta}_1 \text{ and } \eta_2 = \tilde{\eta}_2$$

The identifiability of several families of functions has already been established. For instance, the mixtures of Gamma, Poisson and binomial by Jiang and Tanner

(1999). Ahmad (1998) also proved the identifiability of the Weibull, Lognormal, chi, Pareto, and power functions. With this we proceed with the general proposition for the bivariate mixtures expressed in equation (1).

3.2 General Statement

Theorem 2
Assuming:
(i) $f(x;\theta)$ determines θ
(ii) $f(y|x;\phi) = g(y;\eta = (h(x;\beta),\alpha))$ where $h(x;\beta)$ determines β over the support of $f(x;\theta)$ Then MBD
$f(x,y;\pi,\theta,\phi = (\beta,\alpha)) =$
$\pi\, f(y|x;\phi_1)\, f(x;\theta_1) + (1-\pi)\, f(y|x;\phi_2)\, f(x;\theta_2)$
is identified if:
(A) g is two-mixing or
(B) g ~ Bernoulli $(h(x;\beta))$ is not a constant on support of $f(x;\theta)$ and $f(x;\theta)$ is two mixing.

Example:
In the case where the y observations are Bernoulli (g ~ Bernoulli $(h(x;\beta))$), we have
$P(Y=1|x) = \pi P_1(x;\beta^{(1)}) + (1-\pi)P_2(x;\beta^{(2)})$.
For f two mixing and
$h(x;\beta) = P_i(x;\beta^{(i)}) = \dfrac{e^{\alpha_i+\beta_i x}}{1+e^{\alpha_i+\beta_i x}}$; $i = (1,2)$, the support of $f(x;\theta)$ has to contain at least four points for the model to be identified.

A special case of the above proposition is the symmetric mixtures considered by Hunter et.al (2007). Notice that for g, the distribution function
$f(y|x;a) = g(y-a)$, $a \in \mathbb{R}$ can be written as
$f(y|x;\beta) = g(y - h(x;\beta))$ where $h(x;\beta)$ is a function of x in place of a shift parameter a. Common forms for $h(x;\beta)$ are:

$$(i)\ a+bx \quad (ii)\ a+bx+cx^2 \quad (iii)\ e^{a+bx} \quad (iv)\ \frac{e^{a+bx}}{1+e^{a+bx}}$$

3.3 MBD with exogenous covariates

As questioned earlier, would the MBD still be identifiable if we introduce exogenous variables (z) in equation (1) as polynomial functions of the parameters of the models. For example would the Gage (2004) model with covariate x (birth weight) still be identifiable if discrete or continuous variables such as smoking, educational level and maternal age are covariates for the parameters π, ϕ, θ.

For example, if $\theta_1 \neq \theta_2$ and the assumptions of Theorem 2 holds for each z, then $\pi(z)$, $\phi(z)$, $\theta(z)$ are identified and usually what is left is to see if say;

$$i)\ \pi(z,\gamma) = \frac{\exp\left(\sum_{i=0}^{k}\gamma_i z^i\right)}{1+\exp\left(\sum_{i=0}^{k}\gamma_i z^i\right)} \text{ determines } \gamma$$

$$ii)\ \phi(z,\tau) = \sum_{i=0}^{k}\tau_i z^i \text{ determines } \tau$$

$$iii)\ \theta(z,\alpha) = \sum_{i=0}^{k}\alpha_i z^i \text{ determines } \alpha$$

In this case, since every polynomial of the nth degree with real coefficients has precisely n zeros (fundamental theorem of algebra) γ,τ and α are determined if z has at least n distinct values.

4. Application

4.1 Data and Method.

The dataset used for illustrations and analyses of the proposed models in this study are obtained from the national linked birth death files for the birth cohort born in 2001. It consists of all non-Hispanic African and European American singleton births. Cases with missing birth weight or gender were excluded

Let T denote the infant survival time (days) and x be their Birth weight .We are particularly interested in modeling and determining the association/ relationship between the survival time and the covariate x. This paper focuses on infant survival time within the latent subpopulations and the significant effects of birth weight on infant survival within a subpopulation. In addition we look at the mortality within each subpopulation over time.

4.1.1 MBD with survival probability (MBD-Survival).

For the joint model we use the form already discussed in case 2. The mixture sub model $f_i(x;\theta_i)$ is given by two Gaussian densities which are commonly associated with birth weight. The conditional survival submodel $f(y|x;\phi_i = (\beta_i,\alpha_i))$ is taken to be the Weibull distribution. The hazard function of the distribution is monotone increasing when the shape parameter $\alpha > 1$, decreasing when $\alpha < 1$ and constant when the $\alpha = 1$. The hazard function at time t for a given covariate x (birth weight) is modeled as $h(t|x) = h_o(t)\exp(\nu(x))$ i.e. as a function of x where $h_o(t)$ is called the baseline hazard function is assumed Weibull and $v(x)$ is the linear predictor relating to the covariate x. The predictor could be generalized to include higher order polynomials. The MBD-survival is fitted in a single joint likelihood by minimising the negative log likelihood using the ms() minimiser in the S-PLUS library. Details of the fitting procedure are outlined in Gage (2004).

To better understand the latent subpopulations, we adopt the usual convention of assigning the subpopulation with the larger proportion as subpopulation 1 (majority of births) (Gage 2004, McLachlan and Peel 2000). For simplicity, we will refer to this subpopulation as the primary subpopulation. The associated survival and mortality in this subpopulation are respectively referred to as primary survival and primary mortality. The subpopulation with the smaller proportion is called secondary subpopulation and the associated survival/mortality as secondary survival/mortality.

Table1:Descriptive statistics for the populations

Birth Cohort	Total Births	Infant Death Rate
N.H European Am. Female	1,015,923	3.42
N.H African Am. Female	251,684	7.12

N.H = Non-Hispanic
Am.= American

Bias corrected confidence intervals for the parameter estimates are estimated with bootstraps .The 95 percentile confidence limits obtained indicated that all the parameter estimates are statistically significant. These bootstrap results are also compared to classical confidence intervals computed from the Hessian.

4.2 Results

Descriptive statistics of total births and infant death rate are presented in table 1. The results of the fitted model for non-Hispanic European and African Females populations are presented in table 2. The upper section of table 2 gives the estimates of the x marginal. About 9 per cent of infants in the African American female population are identified as being in the secondary subpopulation whereas the 90 percent appear to be in the primary subpopulation. For the European American female about 6 percent are in the secondary and 94 percent appear in the primary subpopulation. Estimates show that the mean of the primary subpopulation is much larger than the mean of the secondary subpopulation in both populations. On the other hand the standard deviation of the secondary subpopulation is much larger than the primary subpopulation. A graphical comparison of specific birth weight total mortality through the year indicates that infant mortality is lower in the non-Hispanic European American within the normal range (3500g-4500g) compared to non-Hispanic African American female. However, it's higher at weights less or equal to 2500g (Figure1). The mortality curves over time within a population shows that primary mortality is generally higher in comparison to secondary mortality in almost all birth weights (Figure 2). Birth weight specific hazard for the total population is fairly constant in each subpopulation at each birth weight after about 50 days. (Figure 3 A, B, C). A cross section of the mortality curve for the first day (Birth–Day mortality), the 28th day (neonatal) and 364th day (infant mortality) also suggest a fairly reasonable fit of our model. (Figure 3: D-H)

4.3 Discussion

The MBD can be used in any field or area where bivariate mixtures are prominent. The choice of distribution for the conditional submodels is open to the researcher's interest. This article addresses the issues of identifiability of bivariate mixtures that occurs in practice and have established a general proposition with a comprehensive proof for the discussed cases. In this situation the parameters of all cases are uniquely determined and hence the parameter estimates are well behaved. It is also applicable in the special situations such as symmetric mixtures (Hunter (2007). The study demonstrated how these identified models can be applied to survival and mortality data.

The conditional submodel distributions used for the subpopulations in this demonstration were assumed to be Weibull. However other lifetime distributions are also applicable. The Gage model of covariate defined mixtures of logistic regression is seen to be a submodel of MBD when t = 365. The fits of the more general MBD are seen to be consistent with those obtained by Gage. It gives a resolution of the phenomenon that at t = 365, lower birth weight specific infant mortalities among African Americans are smaller compared to European Americans despite their larger infant mortality to all values of t.
The overall birth age specific hazard suggests that the instantaneous risk of death of an infant within the first year of birth is fairly stable/ after approximately 50 days (i.e. remains substantially constant). However this constant risk is still higher for an infant within the primary subpopulation compared to the secondary subpopulation (Figure 3 A, B, C).

Table 2: Parameter estimates for the MBD-Survival model (Weibull)

Weibull

	Non Hispanic African American Female		Non Hispanic European American Female	
	Secondary	Primary	Secondary	Primary
	X Marginal		*X* Marginal	
π	0.09		0.06	
μ_g	2035.56	3169.586	2620.89	3380.08
σ_g	1296.23	455.0998	1116.37	456.45
	Joint Model		Joint Model	
π	0.09		0.06	
μ_g	2031.02	3169.40	2617.56	3379.97
σ_g	1298.14	455.48	1117.73	456.70
β_{1g}	0.02	0.02	0.02	0.03
β_{2g}	3.15E-06	3.21E-06	2.88E-06	3.45E-06
β_{0g}	-2.05	-8.10	-1.63	-10.85
α_g	0.24	0.39	0.22	0.35
λ_g	111.86	1.26E+08	42.48953	7.15E+10

g = (1 = Primary, 2 = Secondary)

π Mixing Proportion

μ Mean birth weight

σ standard deviation of birth weight

β_0 Intercept

β_i Regression coefficients

Weibull Parameters

α Shape

λ Scale

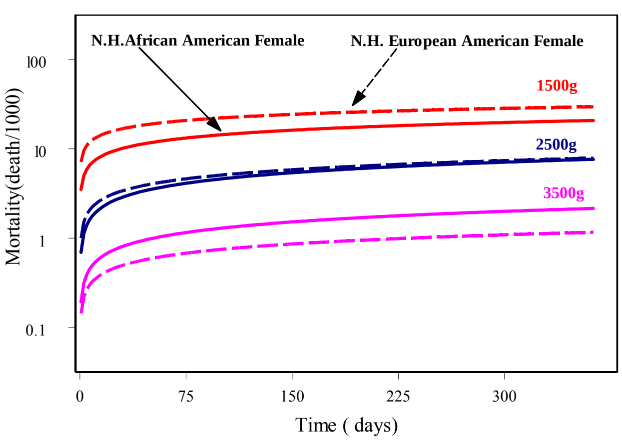


Figure 1 : Comparison of total Mortality within the Non-Hispanic African and European American Females

4.4 Conclusions

1. The study developed conditions for a bivariate mixture model that establish its identifiability.

2. Hazard/mortality in the Secondary subpopulation at each birth weight is lower compared to the Primary subpopulation

3. Birth weight specific hazard is fairly constant in each subpopulation at each birth weight after about 50 days.

4. Lower Birthweight-specific infant mortalities for African Americans are ***smaller*** than European American infants despite their ***larger*** total infant mortality.

.

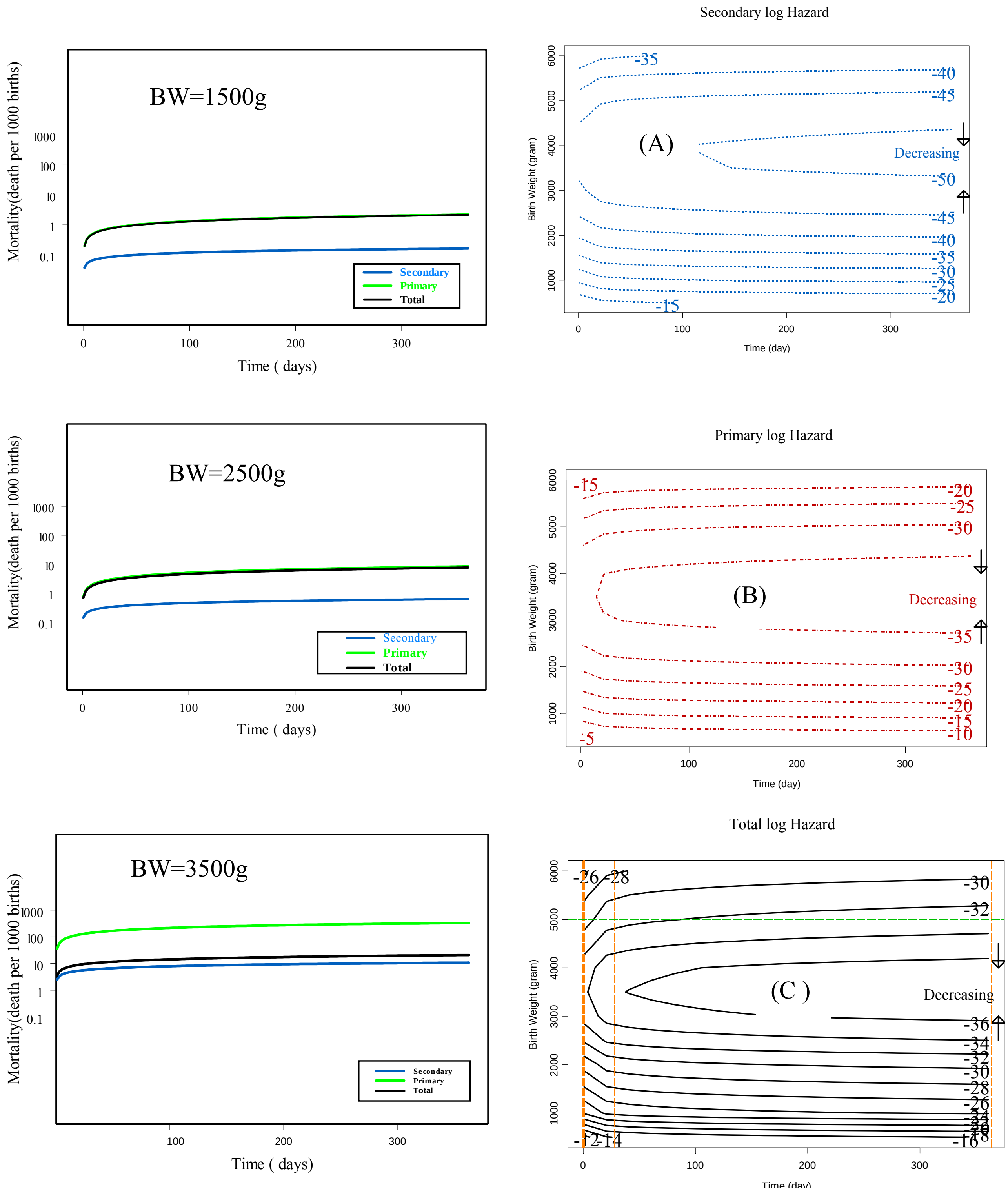
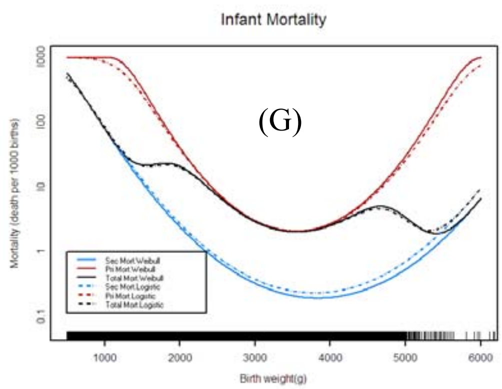


Figure 2 : Secondary Mortality(-) vs Primary Mortality(-) within Non-Hispanic African American Female population for Very Low (1500g), Low (2500g) and Normal (3500) Birth weights



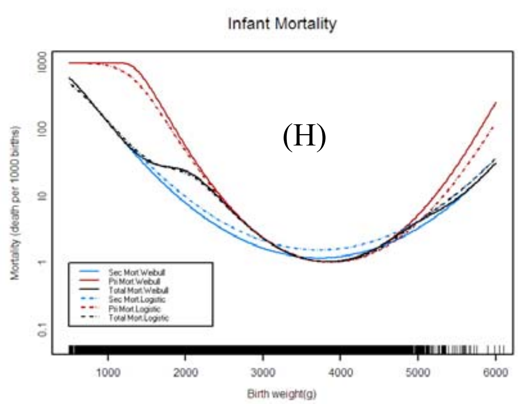
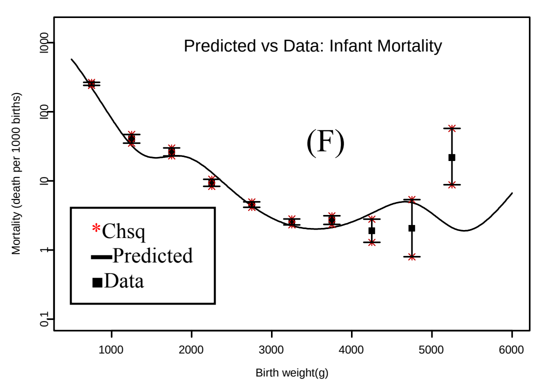
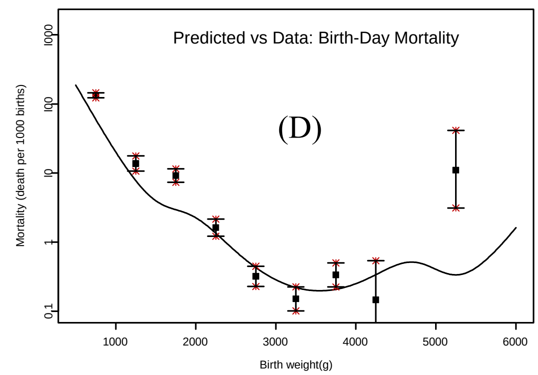
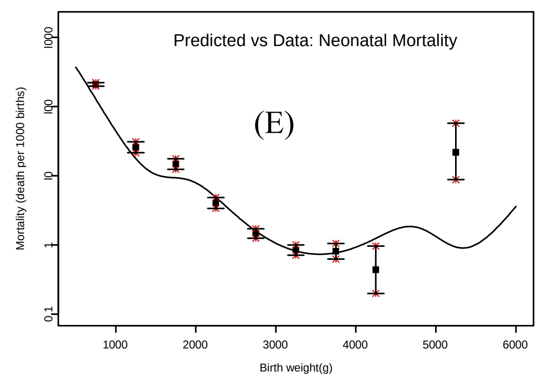


Figure 3 : log hazard contours for non-Hispanic African American Female population (A) Secondary, (B) Primary and (C) Total
Predicted vs Data: First day (D) Neonatal Mortality (E) Infant Mortality (F) Weibull vs Logistic: Infant mortality for the non-Hispanic African (G) &European(H) American Female







References

Ahmad K.E. Identifiability of finite mixtures using a new transform. *Ann. Inst. Statist. Math.* Vol. 40, No. 2, 261-265 (1998).

Allman E.S, Rhodes J.A, The identifiability of tree topology for phylogenetic models, including covarion and mixture models.*J Comput Biol*. 2006 Jun;13(5):1101-13.

Behboodian J. On the Distribution of a Symmetric Statistic From a Mixed Population. *Technometrics* Vol. 14, No. 4 November, 1972.

Everitt B.S. , Hand D.J. (1981). *Finite Mixture Distributions*. Chapman and Hall. New York.

Gage T., Modeling Birthweight and Gestational Age Distributions: Additive vs. Multiplicative Processes. *American Journal of Human Biology* 14: 728-734 (2002)

Gage T, Bauer MJ, Heffner N, and Stratton H. Pediatric Paradox: Heterogeneity in the Birth Cohort.

Human Biology, June 2004, v.76, no.3, pp.327-342.
Gage T, Therriault (1998). Variability of Birth-weight distribution by sex and ethnicity: Analysis using mixture models. *Hum. Biol*. 70:517-534.
Gory G. Parameter Identifiability and Model Selection in Capture-Recapture Models: A Numerical Approach. *Biometrical Journal* 40 (1998) 3, 313-325.
Hennig C. Identifiability of Models for Clusterwise Linear Regression. *Jounal of Classification* 17: 273-296 (2000)
Holzmann H, Munk A, Zucchini W. On identifiability in capture-recapture models. *Biometrics*, 2006 Sep;62(3):934-6; discussion 936-9
Hunter D.R., Shaoli W. and Hettmansperger T.P., Inference for Mixture of Symetric Distributions. *The annals of Statistics* 2007, Vol. 35, No. 1. 224-251
Jiang, W, Tanner M.A, On Identifiability of Mixtures –of-experts. *Neural Networks* 12 (1999) June;1253-1258
Kent J.T, Identifiability of Finite Mixtuers for Direction Data. *The Annals of Statistics,* Vol. 11, No. 3. (Sep., 1983), pp. 984-988.
Kiefer NM, Discrete Variation: Efficient Estimation of a Switching Regression Model. *Econometrica,* Vol. 46, No. 2 (Mar., 1978), pp. 427-434
Link W. A. (2004) Individual heterogeneity and identifiability in capture-recapture models. *Animal Biodiversity and Conservation* 27.1: 87-91
Link W. A. (2003) Nonidentifiability of population size from capture-recapture data with heterogenous detection probabilities. *Biometrics*, 59: 1125-1132.
Mallows C.L., Andrews D.F., Scale Mixtures of Normal Distributions. Journal of the Royal Statistical Society. Sereies B (Methodological), Vol. 36, No.1.(1974), pp. 99-102.
MacLachlan G. and Peel D. A. (2000). *Finite Mixture Models*. Wiley, New York. MR1789474.
Redner R, Note on the Consistency of the Maximum Likelihood Estimate for Nonidentifiable Distributions. *The Annals of Statistics*, Vol. 9, No. 1. (Jan., 1981), pp. 225-228.
Teicher H. (1961). Identifiability of mixtures. *Ann. Math. Statis*., 33,244-248
Teicher H. (1963). Identifiability of finite mixtures. *Ann. Math. Statis*., 34, 1265-1269
Titterington D.M., Smith A.F. and Makov U.E, (1985). *Statistical Analysis of Finite Mixture Distribution*. Wiley, Chichester. MR0838090
Wedel M, Desarbo W., Market Segment Derivation and Profiling Via a Finite Model Mixture Model Framework. *Marketing Letters* 13:1, 17-25, 2002
Yakowitz S.J, Spragins J.D. On the Identifiability of Finite Mixtures. *The Annals of Mathematical Statistics*, Vol. 39, No.1 (Feb., 1968), pp. 209-214.

Appendix

We prove the general theorem.

Proof:

(A) Suppose that

$$\pi\, f(y|x;\phi_1))\, f(x;\theta_1) + (1-\pi)\, f(y|x;\phi_2))\, f(x;\theta_2)$$
$$= \tilde{\pi}\, f(y|x;\tilde{\phi}_1))\, f(x;\tilde{\theta}_1) + (1-\tilde{\pi})\, f(y|x;\tilde{\phi}_2))\, f(x;\tilde{\theta}_2) \qquad (**)$$

Summing out y, gives the x-marginal:

$$f(x;\theta) = \pi\, f(x;\theta_1) + (1-\pi)\, f(x;\theta_2)$$
$$= \tilde{\pi}\, f(x;\tilde{\theta}_1) + (1-\tilde{\pi})\, f(x;\tilde{\theta}_2)$$

Dividing (**) by the x-marginal,

$$\frac{\pi f(x;\theta_1)}{\pi f(x;\theta_1)+(1-\pi) f(x;\theta_2)} \underbrace{f(y|x;\phi_1)}_{g(y;\eta_1(x)=(h(x;\beta_1),\alpha_1)} + \frac{(1-\pi) f(x;\theta_2)}{\pi f(x;\theta_1)+(1-\pi) f(x;\theta_2)} \underbrace{f(y|x;\phi_2)}_{g(y;\eta_2(x)=(h(x;\beta_2),\alpha_2)}$$

$$= \frac{\tilde{\pi} f(x;\tilde{\theta}_1)}{\tilde{\pi} f(x;\tilde{\theta}_1)+(1-\tilde{\pi}) f(x;\tilde{\theta}_2)} \underbrace{f(y|x;\tilde{\phi}_1)}_{g(y;\tilde{\eta}_1(x)=(h(x;\tilde{\beta}_1),\tilde{\alpha}_1)} + \frac{(1-\tilde{\pi}) f(x;\tilde{\theta}_2)}{\tilde{\pi} f(x;\tilde{\theta}_1)+(1-\tilde{\pi}) f(x;\tilde{\theta}_2)} \underbrace{f(y|x;\tilde{\phi}_2)}_{g(y;\tilde{\eta}_2(x)=(h(x;\tilde{\beta}_2),\tilde{\alpha}_2)}$$

For each fixed x, this is a mixture of the conditionals of g and so by g being 2-mixing identified:

$\eta_1(x) = (h(x;\beta_1),\alpha_1) = \tilde{\eta}_1(x) = (h(x;\tilde{\beta}_1),\tilde{\alpha}_1)$

$\Rightarrow (i)\ \alpha_1 = \tilde{\alpha}_1, \alpha_2 = \tilde{\alpha}_2$, and

$(ii)\ h(x;\beta_1) = h(x;\tilde{\beta}_1)$ But h determines β on the support of $f(x;\theta_1) \Rightarrow \beta_1 = \tilde{\beta}_1$ and $\beta_2 = \tilde{\beta}_2$

$$(2)\ \frac{\pi f(x;\theta_1)}{\pi f(x;\theta_1)+(1-\pi) f(x;\theta_2)} = \frac{\tilde{\pi} f(x;\tilde{\theta}_1)}{\tilde{\pi} f(x;\tilde{\theta}_1)+(1-\tilde{\pi}) f(x;\tilde{\theta}_2)} \quad \forall x$$

Recall equal denominators in (2)

$\Rightarrow \quad \pi f(x;\theta_1) = \tilde{\pi} f(x;\tilde{\theta}_1) \quad \forall x,$

Integrating over x:

$\Rightarrow \quad \pi = \tilde{\pi}$

$\Rightarrow \quad f(x;\theta_1) = f(x;\tilde{\theta}_1) \quad \forall x$

since $f(x;\theta)$ determines θ

$\Rightarrow \theta_1 = \tilde{\theta}_1$ Likewise for parameters indexed by 2.