A Bayesian change-point recursive model:
An application on litter size and number of stillborn piglets

N. Ibáñez - Escriche[1*], E. López de Maturana[2], J. L. Noguera[1], L. Varona[3]

[1]Genètica i Millora Animal, IRTA-Lleida, 25198 Lleida. [2]Dpto. de Mejora Genética Animal, INIA, 28040 Madrid. [3]Dpto. de Anatomía, Embriología y Genética Animal, Universidad de Zaragoza, 50013 Zaragoza.

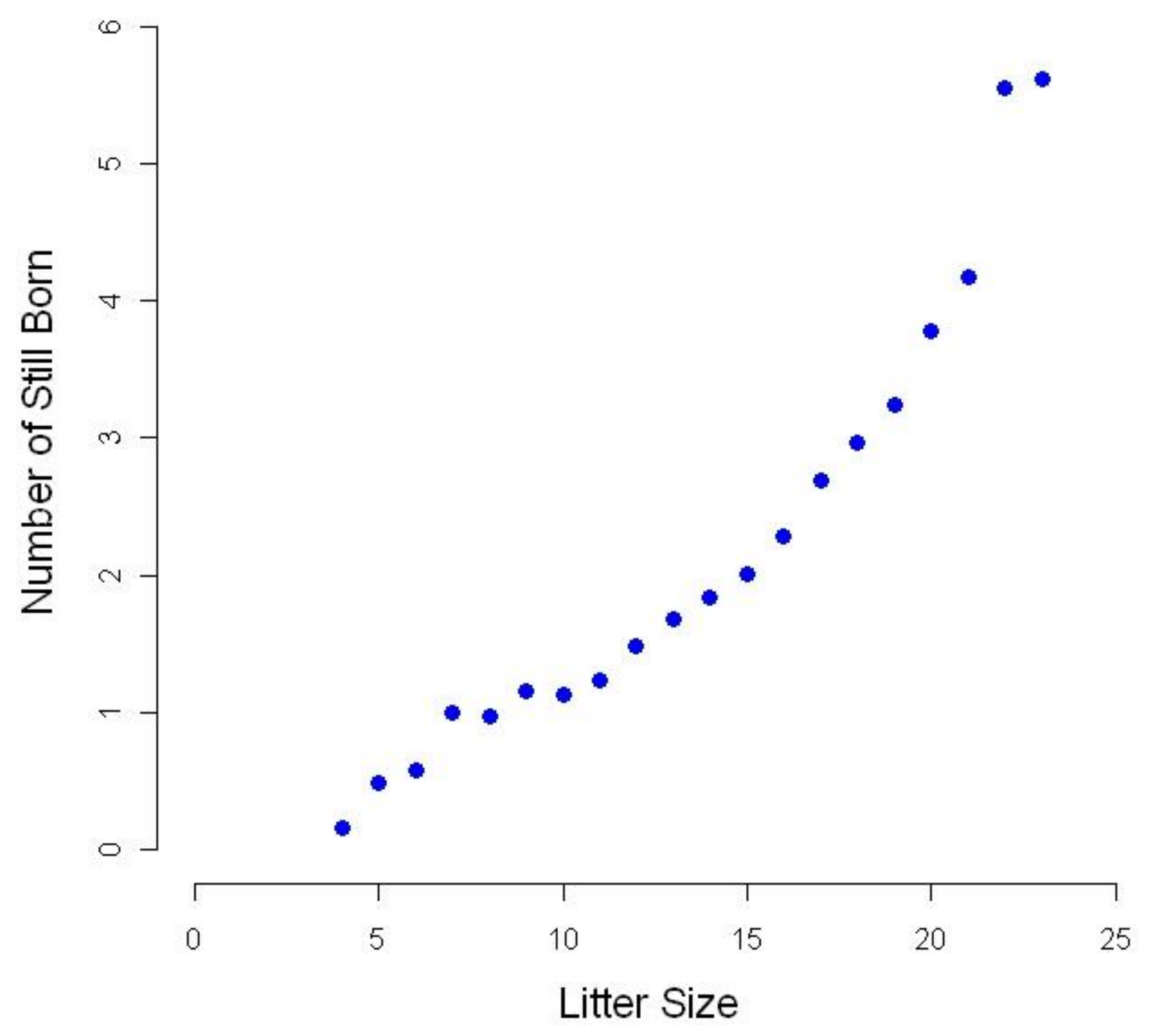
OBJECTIVE

The purpose of this study was to develop change-point recursive models for the investigation of the relationships between litter size (LS) and number of stillborn piglets (NSB)

MATERIAL AND METHODS

Data base

- Nucleus of Large White selected by LS
- Records from 1999 -2006
- LS and NSB of 4462 farrowings
- 1070 sows



Statistical Analysis

Four **change point recursive** models and a standard mixed model (SMM) were used.

The data for individual *i* belonging to subpopulation *k* and model *j (j=1 ,…, 4)* were modeled as

$$\boldsymbol{\Lambda}_{kj}\,\mathbf{y}_i = \mathbf{X}_i\mathbf{b}_j + \mathbf{Z}_i\mathbf{a}_j + \mathbf{W}_i\mathbf{p}_j + \mathbf{e}_j,$$

where

$$\mathbf{y}_i \mid \boldsymbol{\Lambda}_{kj}, \mathbf{t}_{kj}, \mathbf{b}_j, \mathbf{a}_j, \mathbf{p}_j, \mathbf{R}_0 \sim N\left(\boldsymbol{\Lambda}^{-1}_{kj}\left(\mathbf{X}_i\mathbf{b}_j + \mathbf{Z}_i\mathbf{a}_j + \mathbf{W}_i\mathbf{p}_j\right), \boldsymbol{\Lambda}'^{-1}_{kj}\ \mathbf{R}_0\boldsymbol{\Lambda}^{-1}_{kj}\right)$$

- $\mathbf{y}_i$**:** 2 x 1 data LS and NSB
- **b** : parity order (6) and year season (31)
- **a** : additive genetics(1530)
- **p**: permanents (1070)
- $\mathbf{X}_i$, $\mathbf{Z}_i$, $\mathbf{W}_i$**:** known incidence matrices

Where Λ_{kj} is the matrix of structural coefficients corresponding to subpopulation *k* of model *j* for LS, and it takes the form

$$\Lambda_{kj} = \begin{pmatrix} 1 & 0 \\ -\lambda_{NSB \leftarrow LS_{kj}} & 1 \end{pmatrix}$$

Priors:

- $\boldsymbol{\Lambda}_{kj}$**, b, a, p** → Multivariate
- $\mathbf{G}_0$**,** $\mathbf{P}_0$**,** $\mathbf{R}_0$ → Inverted Wishart
- $\mathbf{t}_{kj}$ → Uniform

$\mathbf{t}_{kj}$ is the vector of change points and the *k* subpopulation for model *j* is defined by the records of LS between the change points t_k and t_{k+1} for *k=1, …, j* , where t_1 and t_{m+1} are equal to the smallest and largest records of LS.

McMC: Gibbs and Metropolis Hasting (**t**) algorithms

Model comparison

Deviance Information Criterion

$$DIC = 2\overline{D} - D(\overline{\theta}_M)$$

K fold Cross Validation:

- Data partitioned into 5 subsets
- Pearson's correlation (PC)
- Mean squared error (MSE)

Leave-one-out Cross-Validation:

- Conditional predictive ordinates CPO_i
- Bayesian squared standardized residual D^2

RESULTS

Table 1. Monte Carlo estimates of posterior means and 95% highest posterior density intervals (between brackets) of change points (t) and structural coefficients (λ_j) of models 1, 2, 3 and 4.

Parameter	Model 1	Model 2	Model 3	Model 4
t_2	-	16 (16, 16)	16 (15, 16)	12 (5, 16)
t_3	-	-	20 (17, 23)	16 (15, 20)
t_4	-	-	-	20 (19, 23)
λ_1	0.15 (0.14, 0.16)	0.13 (0.12, 0.13)	0.13 (0.12, 0.13)	0.11 (0.03, 0.14)
λ_2	-	0.17 (0.16, 0.17)	0.16 (0.15, 0.17)	0.13 (0.12, 0.16)
λ_3	-	-	0.21 (0.18, 0.26)	0.16 (0.15, 0.20)
λ_4	-	-	-	0.22 (0.18, 0.27)

(Co) Variances:

Genetic additive correlations were low and close to zero.

Permanent and residual correlations between LS and NSB were positive across models and their HPD95% did not include the zero.

Although only residual correlation of models 2, 3 and 4 showed a clear increase with the subpopulations, posterior means of genetic and permanent correlations showed a similar trend.

Table 2. Monte Carlo estimates of D^2, the sum of Log(CPO), (DIC), the Deviance ($\overline{D}$), the effective number of parameters (*pD*) and the average of MSE and PC for the five subsets.

Model	D^2	Log(CPO)	DIC	$\overline{D}$	*pD*	MSE	PC
1	1.51	-15728	9773	8832	941	3.80	0.384
2	0.84	-15371	9791	8832	959	3.76	0.404
3	0.84	-15368	9803	8832	971	3.73	0.408
4	0.84	-15366	9805	8832	973	3.82	0.41
SMM	1.69	-15872	9880	8925	956	4.23	0.286

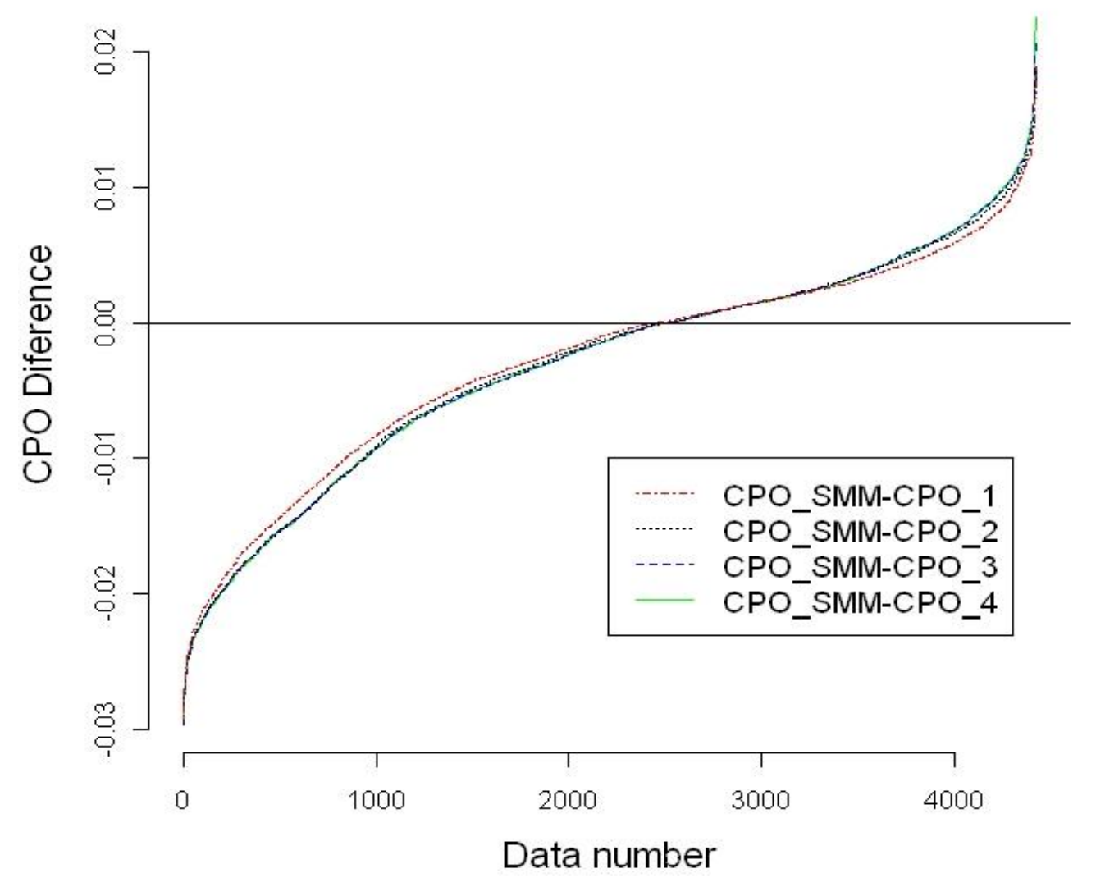


Figure 2. Difference in CPOs SMM and recursive models, sorted from smallest to largest difference.

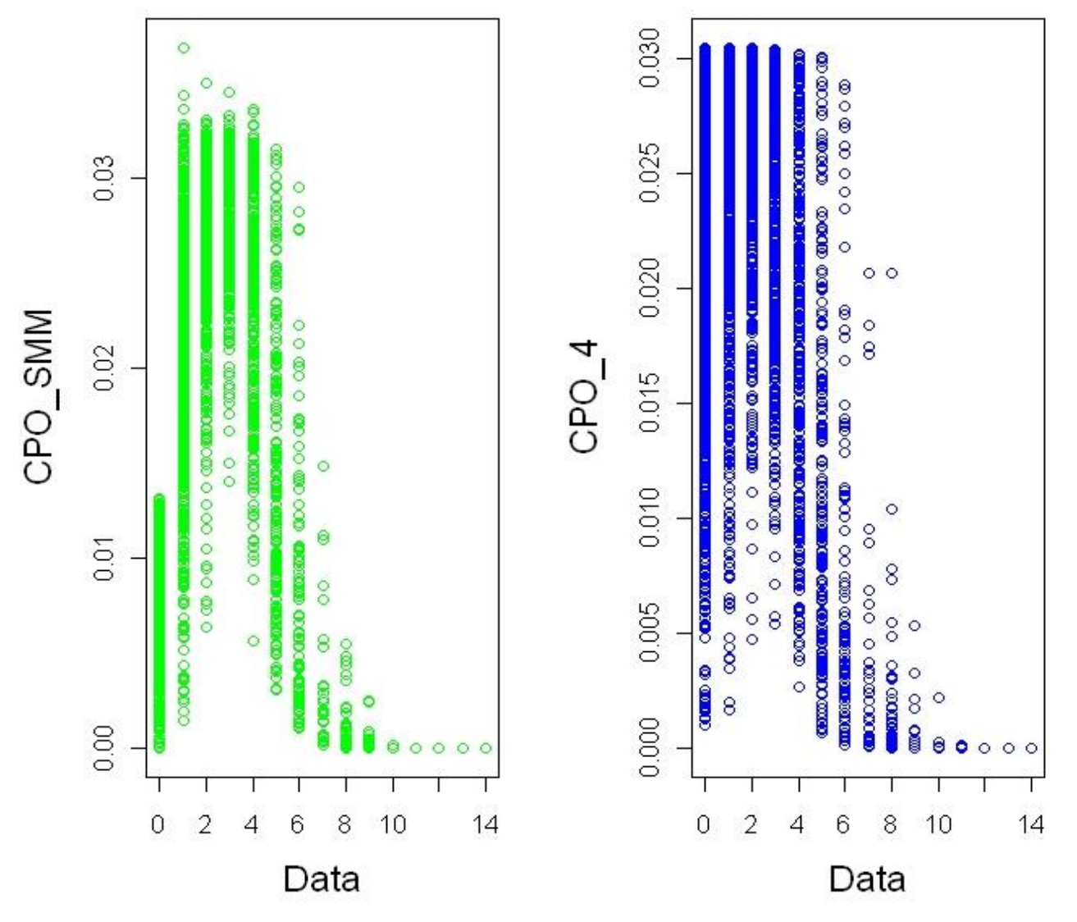


Figure 1. Right CPOs from SMM plotted against number of stillborn. Left. CPOs from model 4 plotted against number of stillborn.

CONCLUSIONS

Recursive change point models revealed a positive relationship between LS and NSB, that is increased when LS takes extreme values

In terms of model comparison, the model 1 was preferred on account of its smallest deviance value (degree of parameters), whereas model 2, 3 and 4 provided a better fit and predictive ability based on the cross validation approaches