A new method for fitting semi parametric variance regression

Kristy Robledo

Supervisor: Prof Ian Marschner
Macquarie University

23 November 2016

Variance heterogeneity

Usual regression model:

$$X_i \sim N\left(\sum \beta x, \sigma^2\right)$$

- random error in regression can sometimes depend on covariates
- variance is a function of x: $\sigma^2 = f(x)$
- leads to variance heterogeneity models

Variance heterogeneity

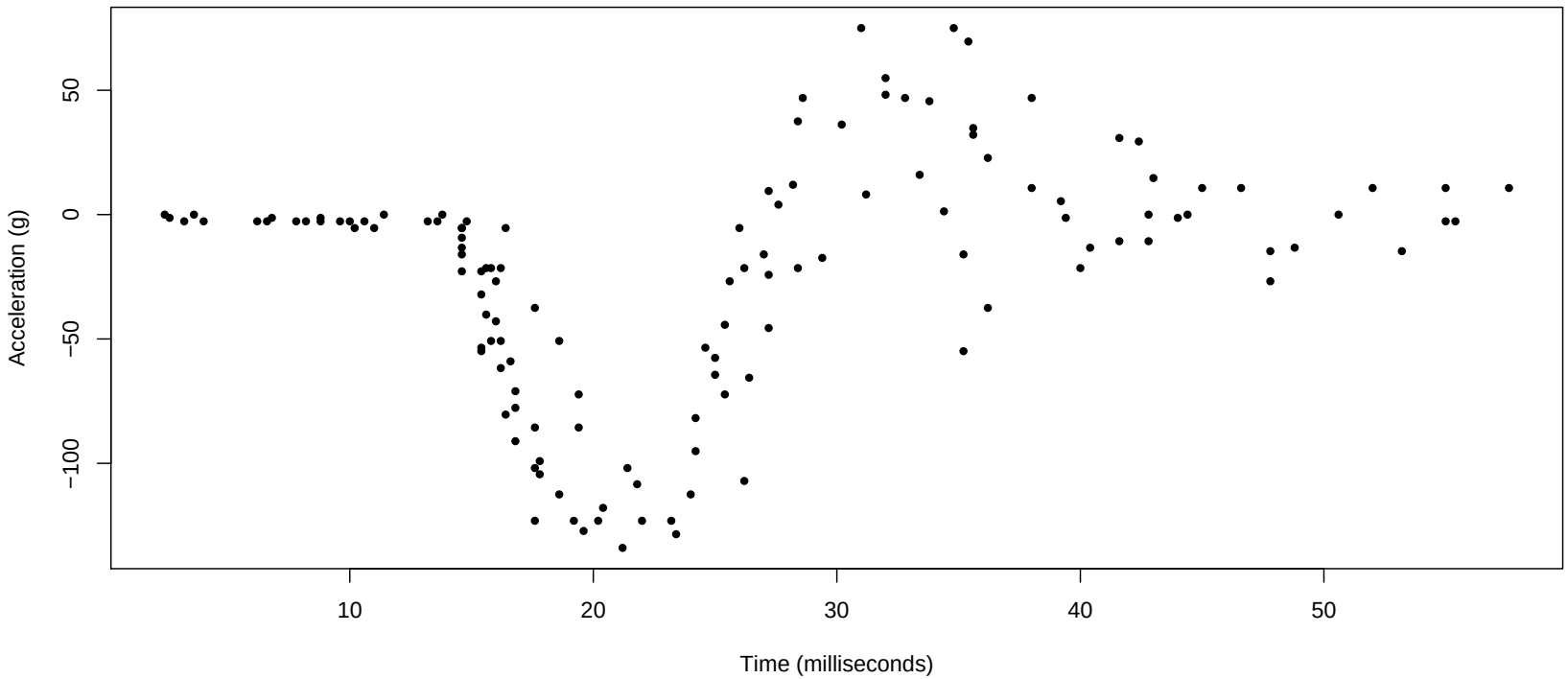


Figure: Motorcycle crash dataset

Semi parametric variance regression

$$X_i \sim N\left(f\left(x\right), g\left(x\right)\right)$$

- $f(x)$ flexible non-linear function in the mean
- $g(x)$ flexible non-linear function in the variance
- can be extended to non-normal models eg. shape/skew regression

B-spline model

$$X_i \sim N\left(f\left(x\right), g\left(x\right)\right)$$

$$f\left(x\right) = \beta_0 + \sum_{p=1}^{P} \beta_j B_p(x)$$

$$g\left(x\right) = \alpha_0 + \sum_{q=1}^{Q} \alpha_q B_q(x)$$

- Regression function of f and g is a linear combination of the B-spline basis functions
- The B-splines can also be easily restricted to be monotonic.

Example of B-splines

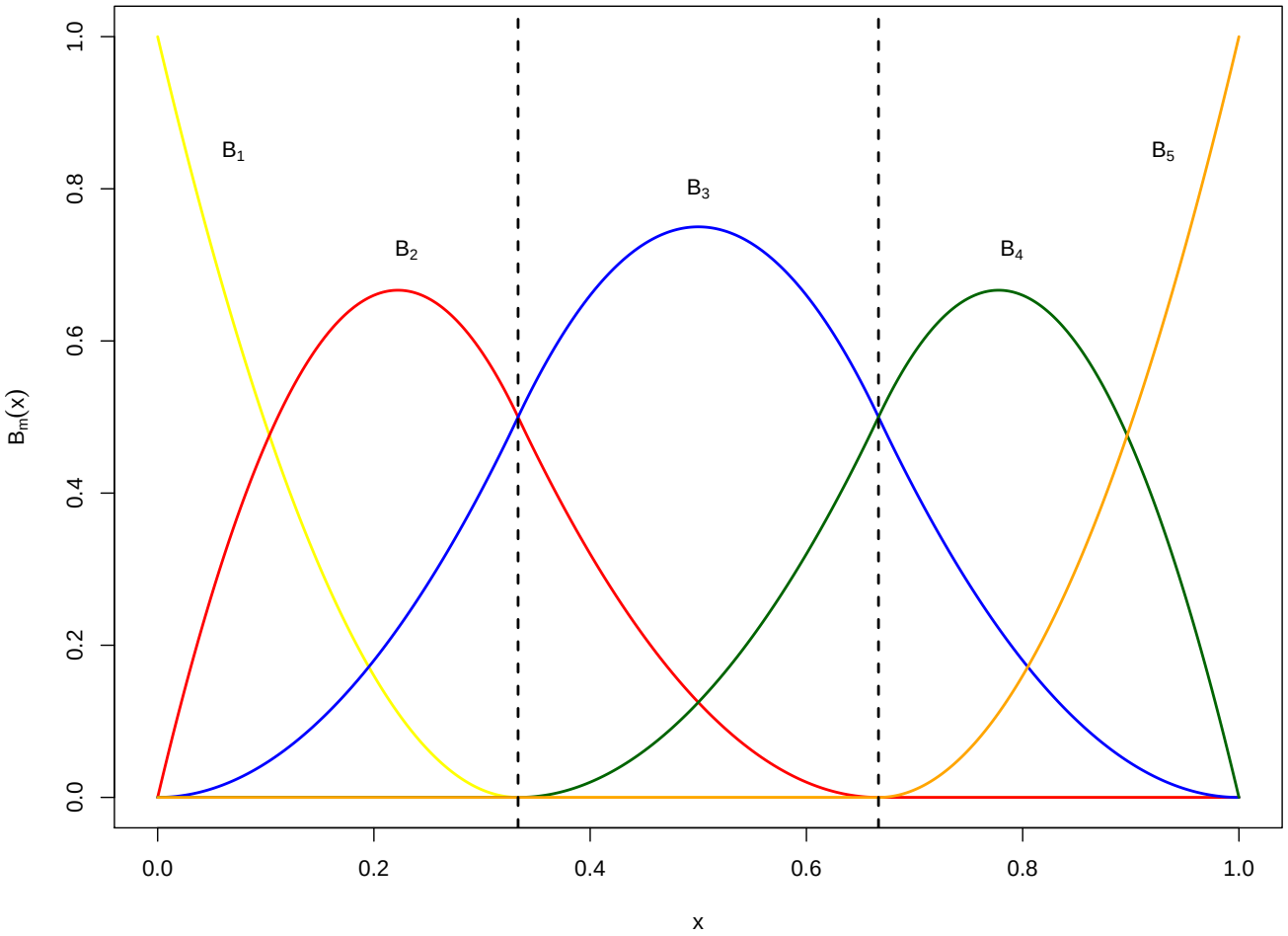


Figure: B-spline basis function for given x with two internal knots (dashed lines)

B-spline model

the B-spline model is simply a special case of

$$X_i \sim N\left(\sum_p \beta_p x_p, \sum_q \alpha_q x_q\right)$$

- existing methods can struggle with additive variance structure
- eg constrained non stationary MLE are difficult with Newton-type algorithms
- new computational and model fitting methodology are always of interest for complex models

Review of EM algorithm

EM is the Expectation - Maximization algorithm, which is an iterative method for finding the MLE when there is missing data or latent variables.

We illustrate it for a simple variance regression model:

$$X_i \sim N(0, \alpha_0 + \alpha_1 x_i)$$

Suppose that $X_i = Y_i + Z_i$ where Y_i and Z_i are latent, unobserved variables (missing).

$$Y_i \sim N(0, \alpha_0) \text{ and } Z_i \sim N(0, \alpha_1 x_i)$$

If Y_i and Z_i are observed, estimation of α_0 and α_1 would be easy.

EM algorithm

"E" step

Calculate the expectation of $\log \ell(\alpha_0, \alpha_1; Y, Z)$ given X and current estimates $\hat{\alpha_0}$ and $\hat{\alpha_1}$

"M" step

Compute parameter estimates to maximise this expected log likelihood. These estimates are used in the next "E" step, and so on (until convergence).

- more stable than existing alternatives

Extensions of EM algorithm

- multiple x_i in the variance model:

$$X_i \sim N\left(0, \sum_q \alpha_q x_q\right)$$

- Extended to mean model via ECME with weighted LS step:

$$X_i \sim N\left(\sum_p \beta_p x_p, \sum_q \alpha_q x_q\right)$$

- semi parametric x_i:

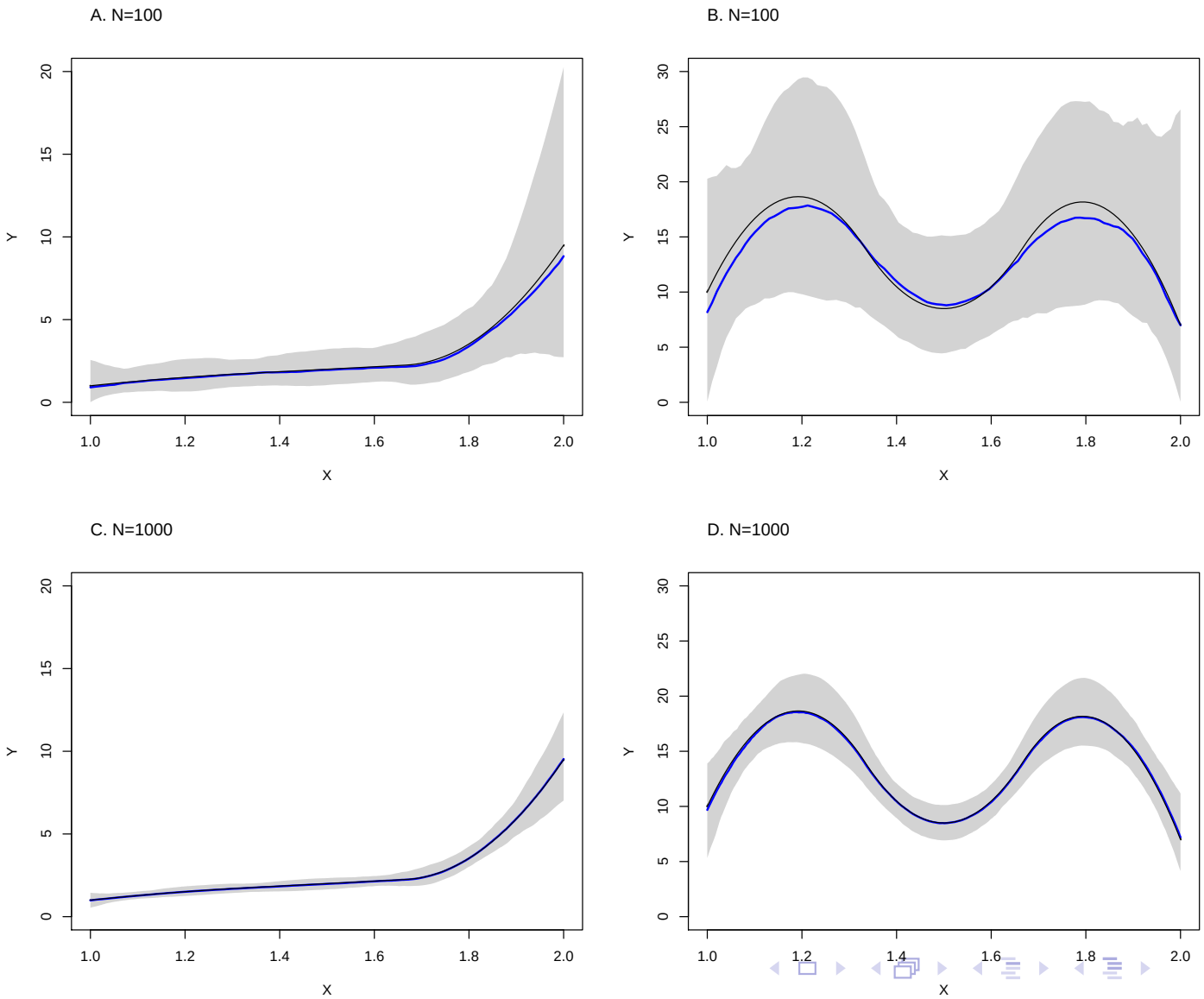
$$X_i \sim N\left(f\left(x\right), g\left(x\right)\right)$$

where $f(x)$ and $g(x)$ are specified using B-spline basis functions.

Simulation study

- one monotonically increasing variance function with zero mean
- one periodically increasing and decreasing variance function with zero mean
- both variance functions with two internal knots
- four sample sizes (100, 250, 500, 1000)
- 500 simulations performed for each

Simulation study: Part One
Reliably estimating a known function



Simulation study: Part Two
Automatic choice of model complexity

- both models are of known complexity (two internal knots = 5 parameters in variance)
- each simulation has 6 models performed; linear (2 parameters) to 4 knots (7 parameters)
- calculate AIC, AICc, HQC, BIC for all models
- choose an optimal model for each criteria
- rinse and repeat for the 500 simulations

Simulation study: Part Two
Automatic choice of model complexity

		Increasing variance				Periodic variance			
N	Statistic	AIC	AICc	HQC	BIC	AIC	AICc	HQC	BIC
100	Median	3.0	3.0	2.0	2.0	3.0	3.0	2.0	2.0
	Mean	3.5	3.3	2.9	2.5	3.8	3.5	2.8	2.2
	Mode	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
250	Median	4.0	4.0	4.0	3.0	5.0	5.0	5.0	2.0
	Mean	4.2	4.2	3.5	2.9	4.9	4.9	3.7	2.6
	Mode	4.0	4.0	4.0	2.0	5.0	5.0	5.0	2.0
500	Median	5.0	5.0	4.0	4.0	5.0	5.0	5.0	5.0
	Mean	4.6	4.6	4.1	3.5	5.2	5.2	4.6	3.6
	Mode	4.0	4.0	4.0	4.0	5.0	5.0	5.0	5.0
1000	Median	5.0	5.0	4.0	4.0	5.0	5.0	5.0	5.0
	Mean	5.0	5.0	4.5	4.1	5.4	5.4	5.1	4.9
	Mode	5.0	5.0	4.0	4.0	5.0	5.0	5.0	5.0

Application to Mcycle data

Table: Table of the AIC from the 100 different mean and variance models for the motorcycle crash data. The lowest AIC is in bold.

Variance	Mean									
	Linear	0 knots	1 knot	2 knots	3 knots	4 knots	5 knots	6 knots	7 knots	8 knots
Linear	1377	1367	1347	1291	1263	1224	1204	1188	1196	1200
0 knots	1332	1355	1326	1289	1233	1203	1190	1175	1183	1186
1 knot	1363	1320	1327	1273	1226	1185	1163	1144	1151	1155
2 knots	1346	1309	1299	1277	1226	1187	1165	1147	1148	1152
3 knots	1341	1252	1299	1266	1220	1184	1160	1144	1146	1150
4 knots	1279	1264	1266	1244	1231	1179	1152	1138	1141	1146
5 knots	1252	1239	1241	1231	1225	1174	1141	1118	1130	1136
6 knots	1241	1234	1235	1235	1230	1145	1107	**1075**	1101	1112
7 knots	1241	1231	1232	1217	1219	1147	1122	1103	1115	1117
8 knots	1240	1220	1223	1202	1193	1149	1107	1089	1100	1104

Application to Motorcycle data

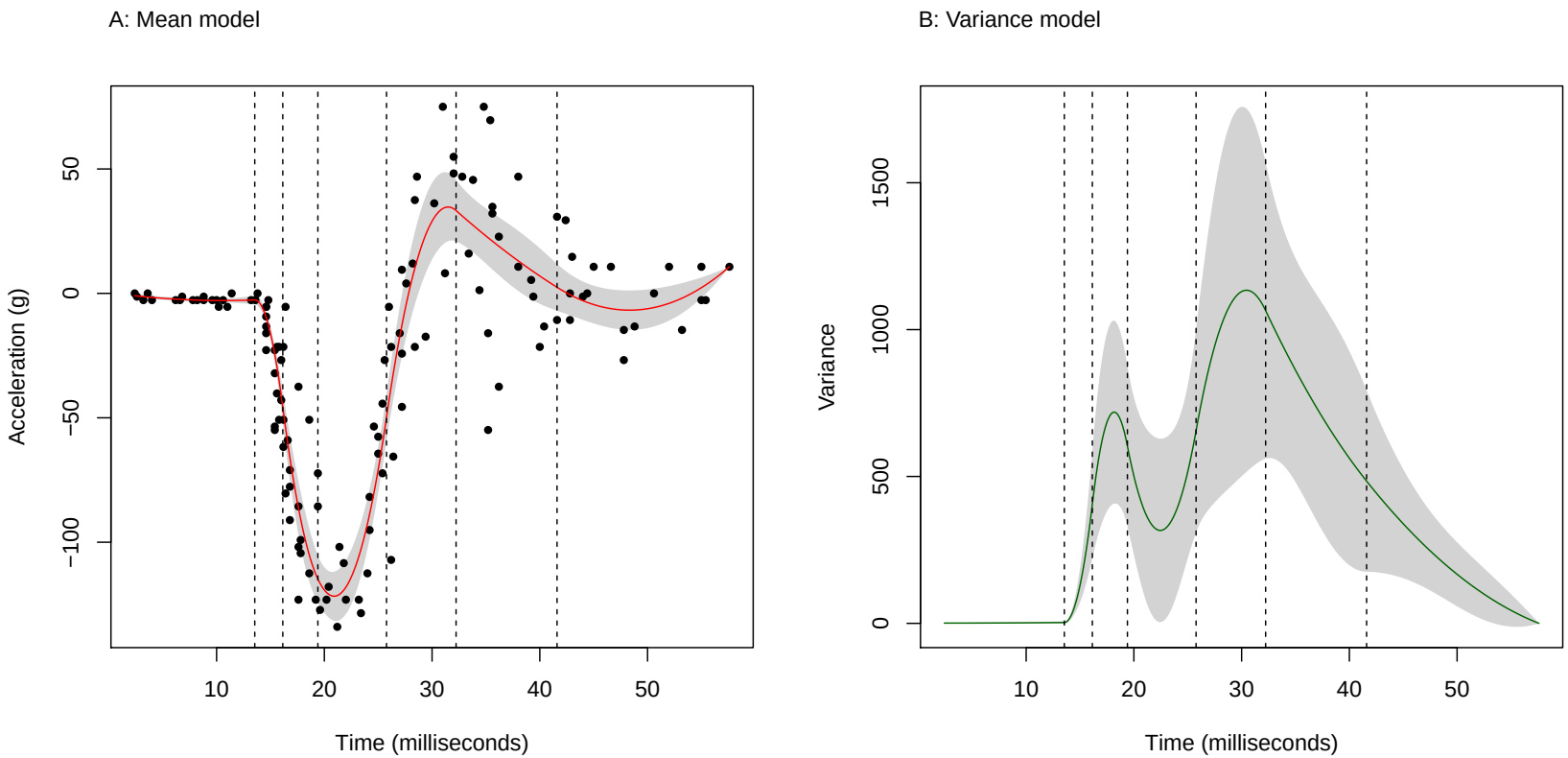


Figure: Optimal model for Motorcycle crash dataset

Application to Motorcycle data

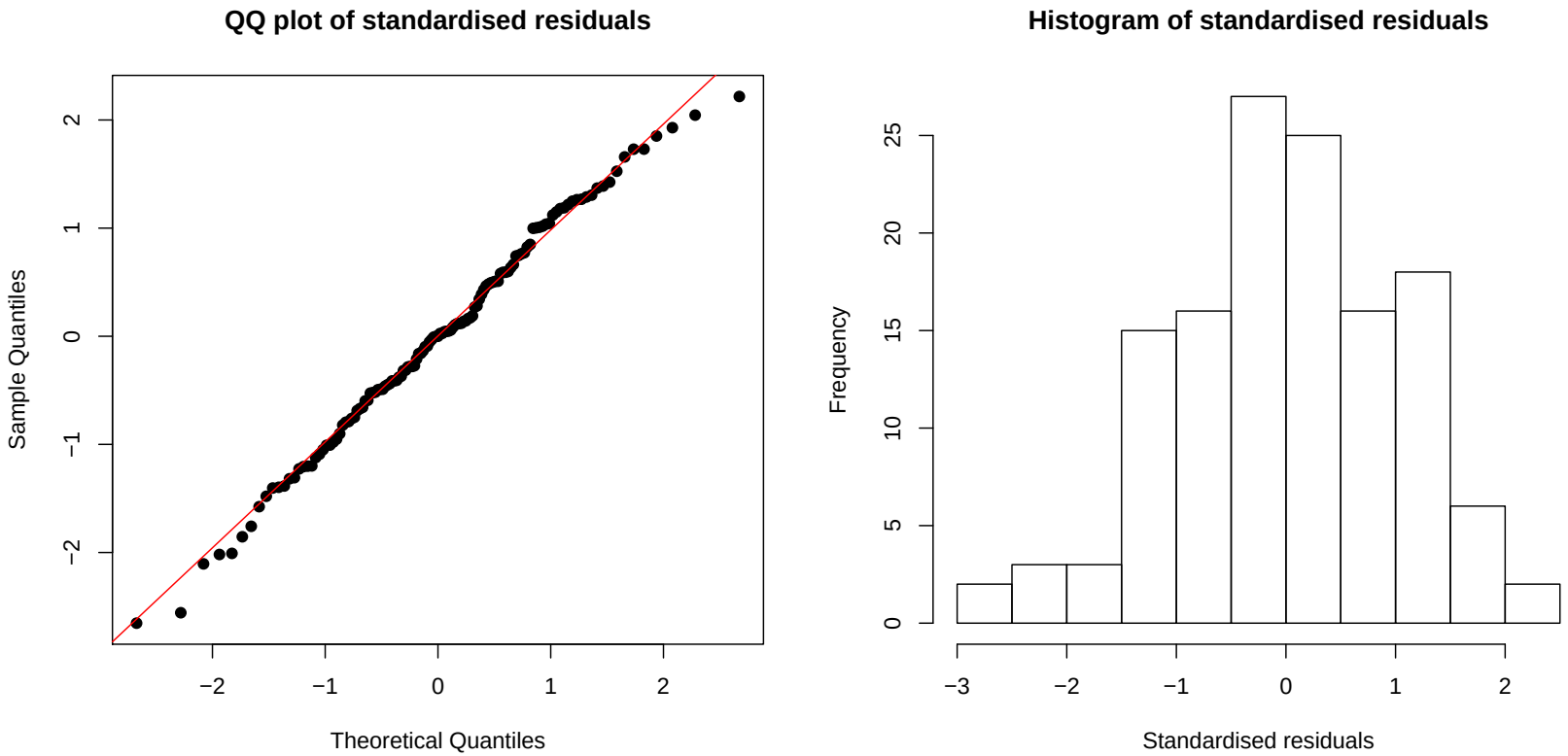
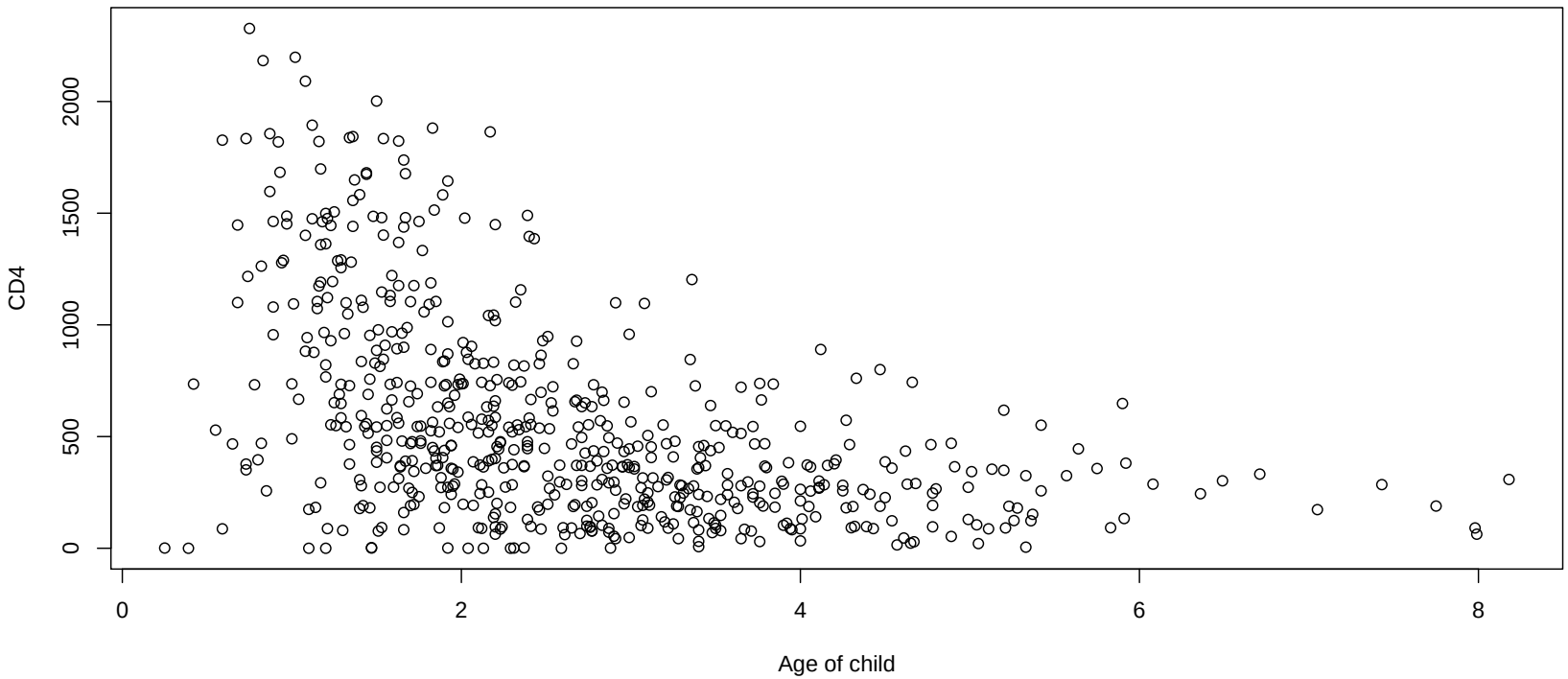


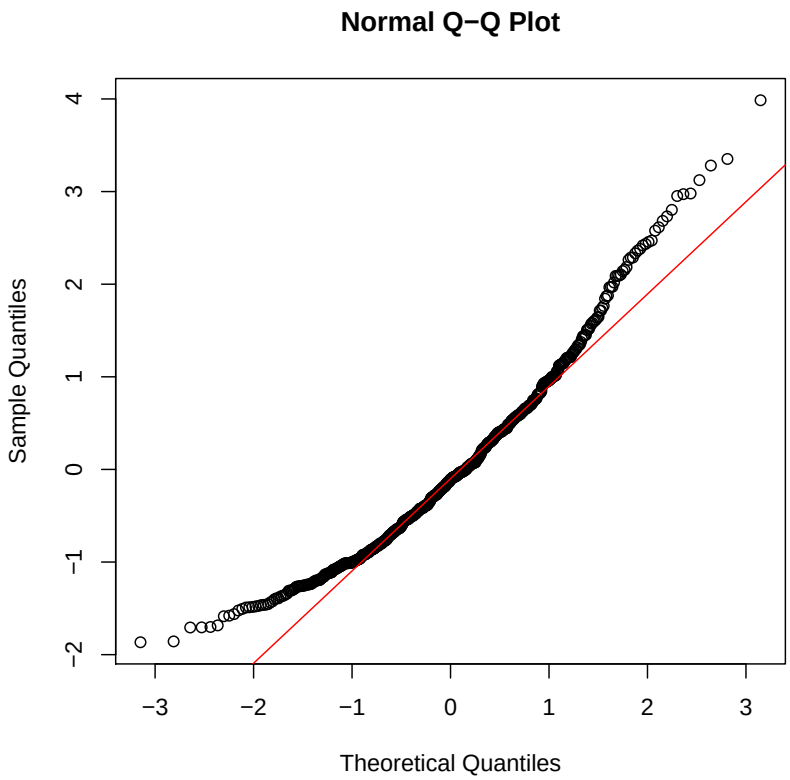
Figure: Optimal model residuals

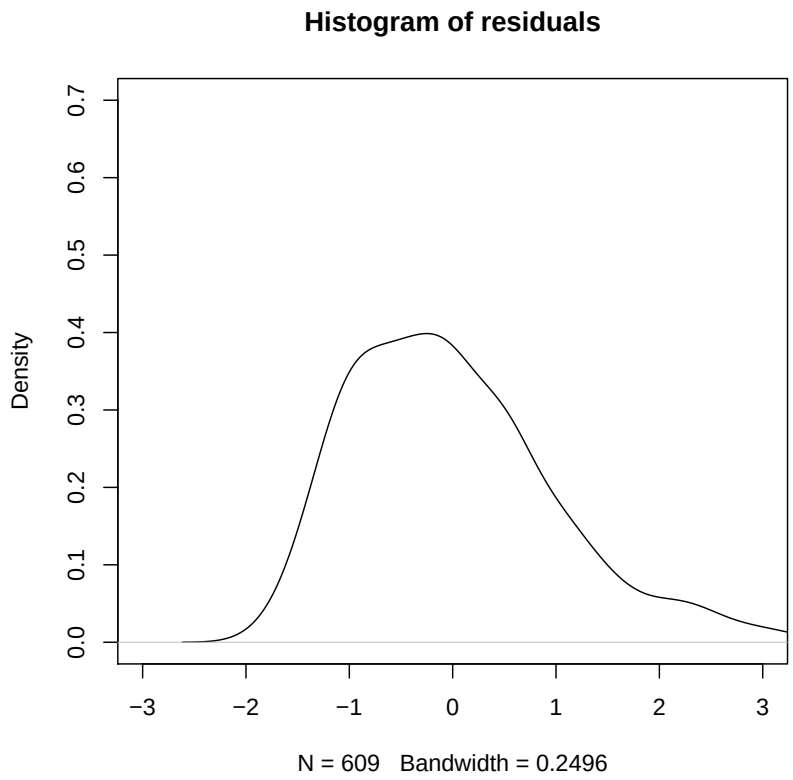
CD4 data example

- CD4 levels in children of HIV positive mothers



Non-normality of CD4 residuals





Location, Scale and Shape regression

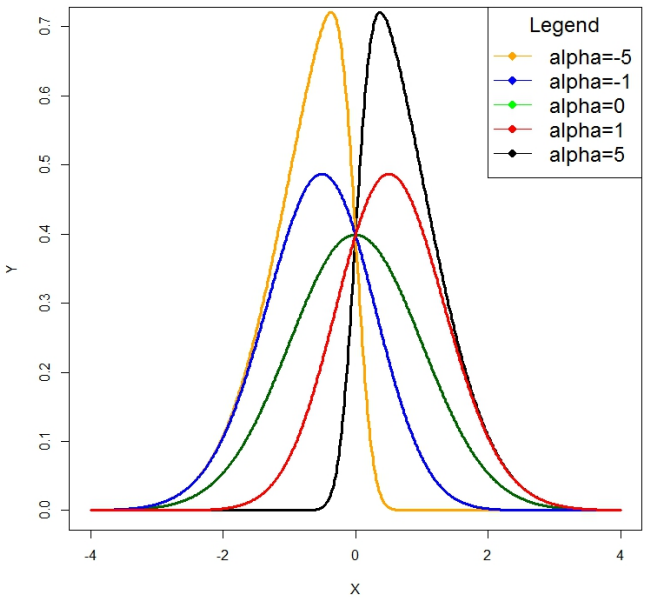
The proposed EM algorithm can be extended to allow shape regression (skewness) using the skew-normal distribution. Skew normal distribution is a 3 parameter distribution that incorporates location ε, scale ω and shape α.

$$f(x) = \frac{2}{\omega}\phi\left(\frac{x-\varepsilon}{\omega}\right)\Phi\left(\alpha\left(\frac{x-\varepsilon}{\omega}\right)\right)$$

$$\mu = \varepsilon + \omega\delta\sqrt{\frac{2}{\pi}} \qquad \text{where } \delta = \frac{\alpha}{\sqrt{1+\alpha}}$$

$$\sigma^2 = \omega^2\left(1 - \frac{2\delta^2}{\pi}\right)$$

$$\gamma = \frac{4-\pi}{2}\frac{\left(\delta\sqrt{2/\pi}\right)^3}{(1-2\delta^2/\pi)^{3/2}}$$



Comparison of models allowing skew

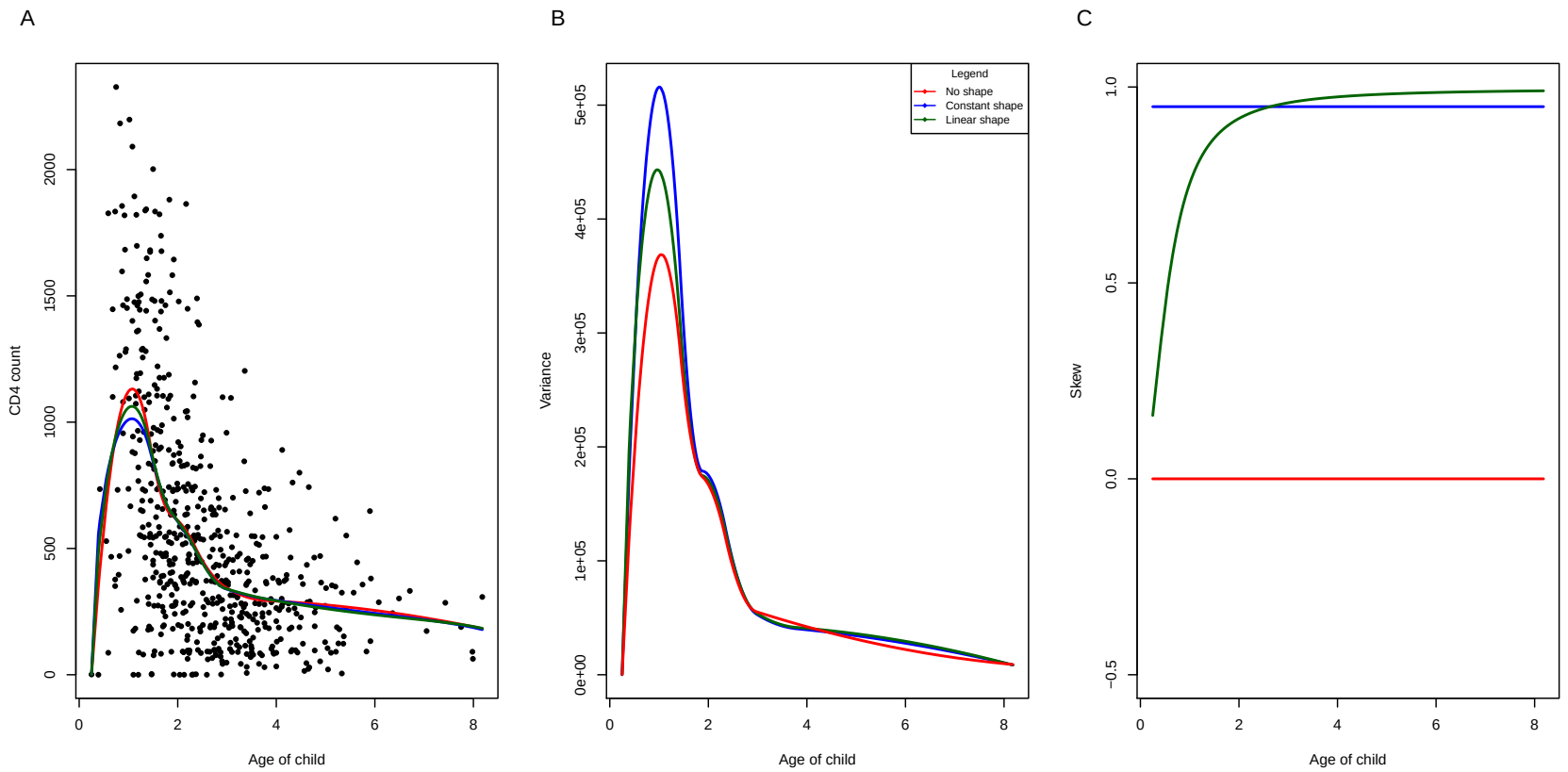


Figure: A: mean models, B: variance models, C: skew models

Comparison of models allowing skew

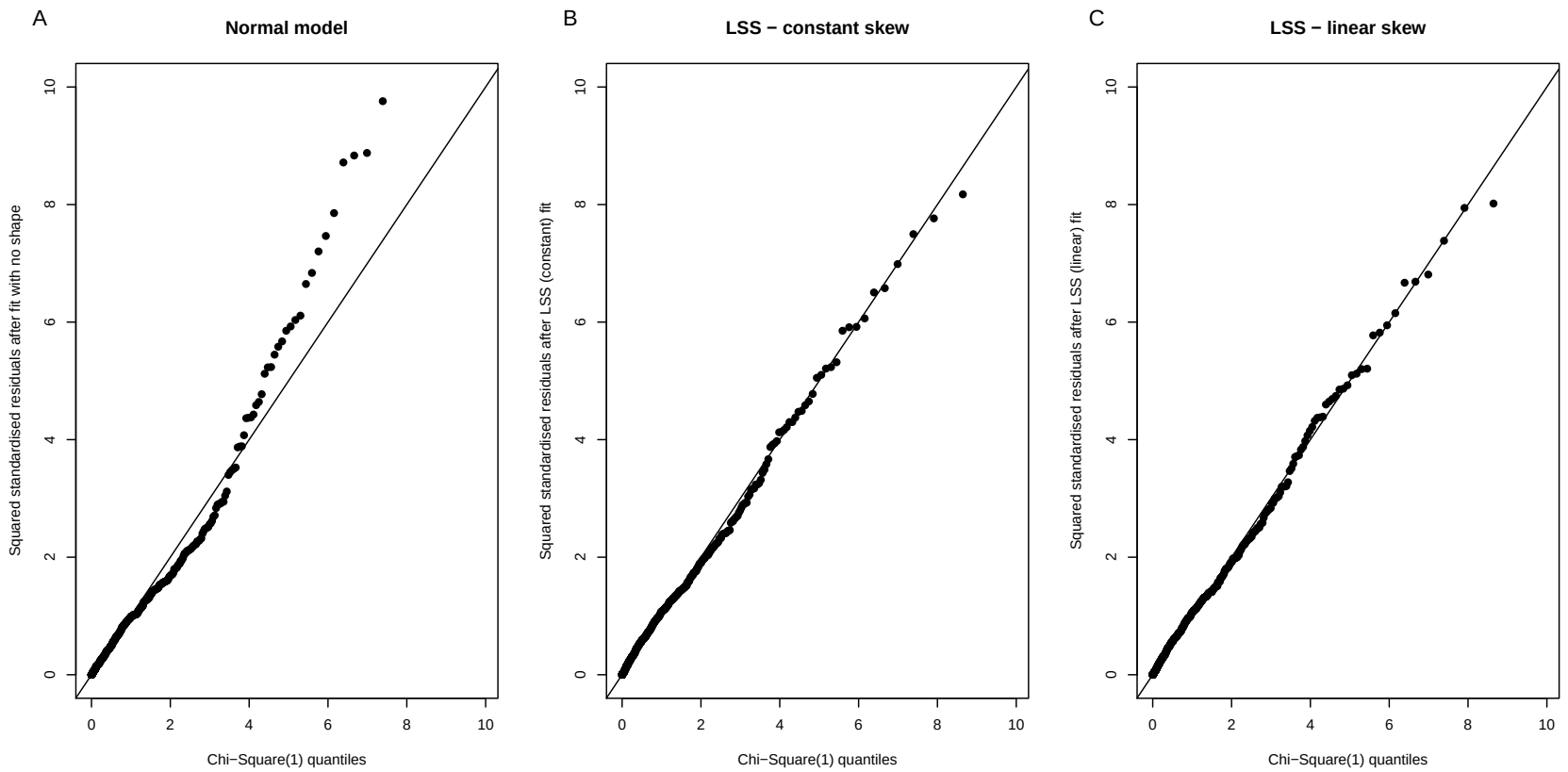


Figure: Residuals from the three models

Distribution over ages

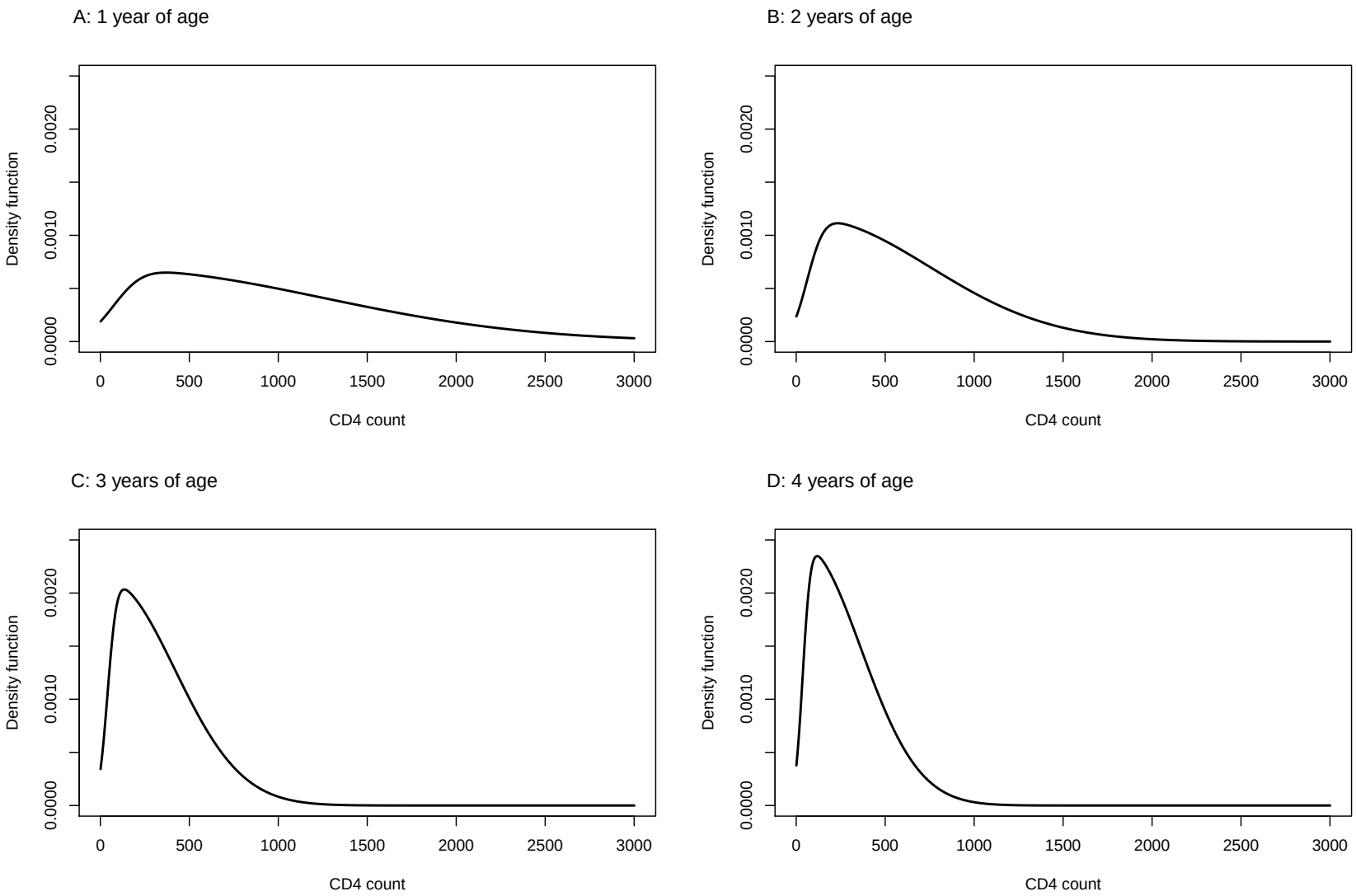


Figure: Comparison of distributions over age for constant skew model.

Conclusions

Our approach is a useful complement to existing methods for semi-parametric variance regression (eg GAMLSS, MFVB) and has good performance. It also can be easily extended.

Additional work:

- monotonicity restrictions
- censored data
- truncated data
- R package under construction

Questions?