Discrimination and Classification of SGA Births

Mette Langaas
Department of Mathematical Sciences
The Norwegian Institute of Technology
The University of Trondheim
mette@imf.unit.no

Geir Jacobsen
Department of Community Medicine and General Practice
The University of Trondheim
geir.jacobsen@ismut.unit.no

Abstract

This is an analysis of the Scandinavian part of the NICHD Study of Successive SGA Births, consisting of data from Trondheim, Bergen and Uppsala. The term SGA (Small-for-Gestational Age) means that the weight of a newborn child is below the 15th percentile of all children born at the gestational age in question. The aim of the present study is to construct methods for predicting an SGA birth as early as possible, early enough so that measures preventing and limiting the consequences of an SGA birth can be taken. The data is used for discrimination and classification with two classes, SGA and non-SGA births. The explanatory variables at study are grouped into three categories; prepregnancy variables, variational variables and growth variables. The variational and growth variables are analysed using random coefficients regression models. Classification rules are formed for available data at five different stages of the pregnancy; time of conception, mid second trimester, early third trimester, late third trimester and time of birth.

Some keywords: Discriminant Analysis; (Supervised) Classification; Random Coefficient Regression Models; Growth Curves.

Contents

1 Introduction

This presentation is based on the Scandinavian part of the NICHD Study of Successive SGA Births, see Bakketeig et al. (1993) and Jacobsen (1992). The Scandinavian part consists of data from Trondheim, Bergen and Uppsala.

The term SGA (Small-for-Gestational Age) means that the weight of a newborn child is below the 15th percentile of the children born at the gestational age in question (definition with 10th percentile also exists).

The SGA data set presented here consists of measurements of women expecting their second or third child. Some of the women are from a 10% random sample from the study population and some from a group of women fulfilling one or more defined risk criteria for SGA birth. [1] The study population is uniform in ethnic and socio-demographic terms both in Norway (Trondheim and Bergen) and Sweden (Uppsala). All the women are white Caucasians. The data analysed here is a subsample of the Scandinavian part of the NICHD Study of Successive SGA Births, the selection criteria are given in the next section. For a thorough presentation and further details and analyses about the Scandinavian part see Bakketeig et al. (1993) and Jacobsen (1992).

The aim of this presentation is to find methods for predicting an SGA birth as early as possible, early enough so that measures preventing and limiting the consequences of an SGA birth can be taken. The data is used for discrimination and classification with two classes, SGA and non-SGA births. Five different situations in time are studied. These five situations consist of available data at the time of conception, until the mid second trimester (day 154, week 22), until the early third trimester (day 231, week 33), until the late third trimester (day 273, week 39) and at time of birth.

The analyses in this presentation are done using the S language. General introductions to the S language are Becker et al. (1988), Chambers and Hastie (1992) and Venables and Ripley (1994).

2 Variables at Study

The explanatory variables at study are placed in three categories; prepregnancy variables, variational variables and growth variables. The variables are listed in table 1.

The response is whether the outcome is an SGA or non-SGA birth, defined from child weight and gestational age at birth.

The prepregnancy variables at study are the study site, the age of the pregnant woman, the body mass index (BMI) (defined as prepregnancy weight divided by squared height), an indicator variable denoting whether the woman has had a previous SGA birth and the number of cigarettes smoked at the time around conception. The body mass index (BMI) is also called Quetelet's index and is given in kg/m^2.

The variational variables are systolic and diastolic blood pressure, haemoglobin and the number of cigarettes smoked per day (in average at each measured point in time).

[1] These criteria are given in Bakketeig et al. (1993) and are based on previous low weight births, smoking, low prepregnancy weight, a previous pertinal death and chronic maternal disease.

Explanatory Variables for the Scandinavian SGA Data Set			
Category	Description	Name	Measures
Prepregnancy Variables	study site (Trondheim, Bergen, Uppsala)	site	1
	age	age	1
	body mass index BMI (kg/m^2)	bmi	1
	indicator for previous SGA births	p.sga	1
	number of cigarettes smoked around conception	cig.cons	1
Variational Variables	systolic blood pressure	BPA	5-25
	diastolic blood pressure	BPB	5-25
	haemoglobin	HB	5-25
	number of cigarettes smoked per day	CIG	5
Growth Variables	proportional weight gain	PWG	5-25
	symphysis fundus height (SF height)	SFH	5-25
	ultrasound mean abdominal diameter	MAD	4
	ultrasound biparietal diameter	BPD	4
	ultrasound femur length	FEM	4

Table 1: The selected explanatory variables from the Scandinavian SGA data set. The column denoted "Measures" give the standard number of measurements for each explanatory variable.

The growth variables are foetal ultrasound measurements, symphysis fundus (SF) height (an outer measurement on the pregnant woman) and the proportional weight gain. The proportional weight gain is defined as weight gain relative to the prepregnancy weight, (weight - prepregnancy weight)/(prepregnancy weight). There are three foetal ultrasound measurements; the mean abdominal diameter (MAD), the biparietal diameter (BPD) and the femur length (FEM). Gestational age in days are given with all variational and growth variables. [2]

The standard number of measurements for each variable is listed in table 1. There are missing values in the data set. The data set analysed here is a subsample of the total data set from the Scandinavian SGA study. The selection procedure required that all women present in the subsample had at least two measurements of all growth and variational variables and that no prepregnancy variables were missing. Since all growth and variational variables are measured repeatedly for each pregnancy, only a few cases had to be discarded because too many variables were missing. There were 1552 women who had completed the special study to delivery, Jacobsen (1992). Of these 1020 women had at least two measurements of all growth and variational variables. There were 1011 women who in addition had no missing prepregnancy variables. Of the 1011 women 422 are from Trondheim, 203 from Bergen and 386 from Uppsala. This data set will hereafter be called the TBU-SGA data set.

Based on these 1011 pregnancies the prepregnancy variables are studied further in section

[2] An expected date of delivery was recorded from the last menstrual period and the ultrasound scan around week 17. If the last menstrual period was unknown or uncertain or there was a discrepancy between the two methods, the ultrasound dating was used. Population birth weight standards are based on menstrual age.

2.1. And based on the 1020 prepregnancies the variational variables are studied in section 2.2 and the growth variables in section 2.3.

2.1 Analysis of Prepregnancy Variables

Summary statistics for the prepregnancy explanatory variables are found in table 2.

The prepregnancy variables are found to be nearly uncorrelated. The correlation matrix is given below.

	age	*bmi*	*cig.cons*	*p.sga*
age	1.0000	0.0339	−0.0938	−0.0155
bmi	0.0339	1.0000	0.0575	−0.0878
cig.cons	−0.0938	0.0575	1.0000	0.0151
p.sga	−0.0155	−0.0878	0.0151	1.0000

Using separate two-sample t-tests the expected value of all variables (age, body mass index, previous SGA births and cigarette consumption around conception) are found to be significantly different for the SGA and the non-SGA births. The P-values are; age (P-value of 0.0495), body mass index (P-value of 0.0009), previous SGA birth (P-value of 0) and cigarette consumption around conception (P-value of 0.0005).

Mean age is not found to be different for SGA and non-SGA births in any of the study sites Trondheim, Bergen or Uppsala (P-values of 0.1523, 0.8446 and 0.0749, respectively). The mean body mass index is found to be different for SGA and non-SGA births in Bergen and Uppsala (P-values for Trondheim, Bergen and Uppsala of 0.3189, 0.0199 and 0.002, respectively). The expected value for previous SGA births and for cigarette consumption around conception are found to be different for SGA and non-SGA births in Trondheim and in Uppsala (P-values for previous SGA births are 0, 0.1244 and 0 and P-values for cigarette consumption are 0.0109, 0.5287 and 0.0051 for Trondheim, Bergen and Uppsala, respectively).

2.2 Analysis of Variational Curves

For the variational variables medical theory suggests that there might be a slight increasing trend in time for the blood pressure and a slightly U-shaped form in time for the haemoglobin.

The variational variables are measured at different points in time and for different number of times for each individual, that is, we have unbalanced data. The four measured variational variables are studied using random coefficient regression models with unbalanced data. A short introduction to the relevant theory is found in appendix A.

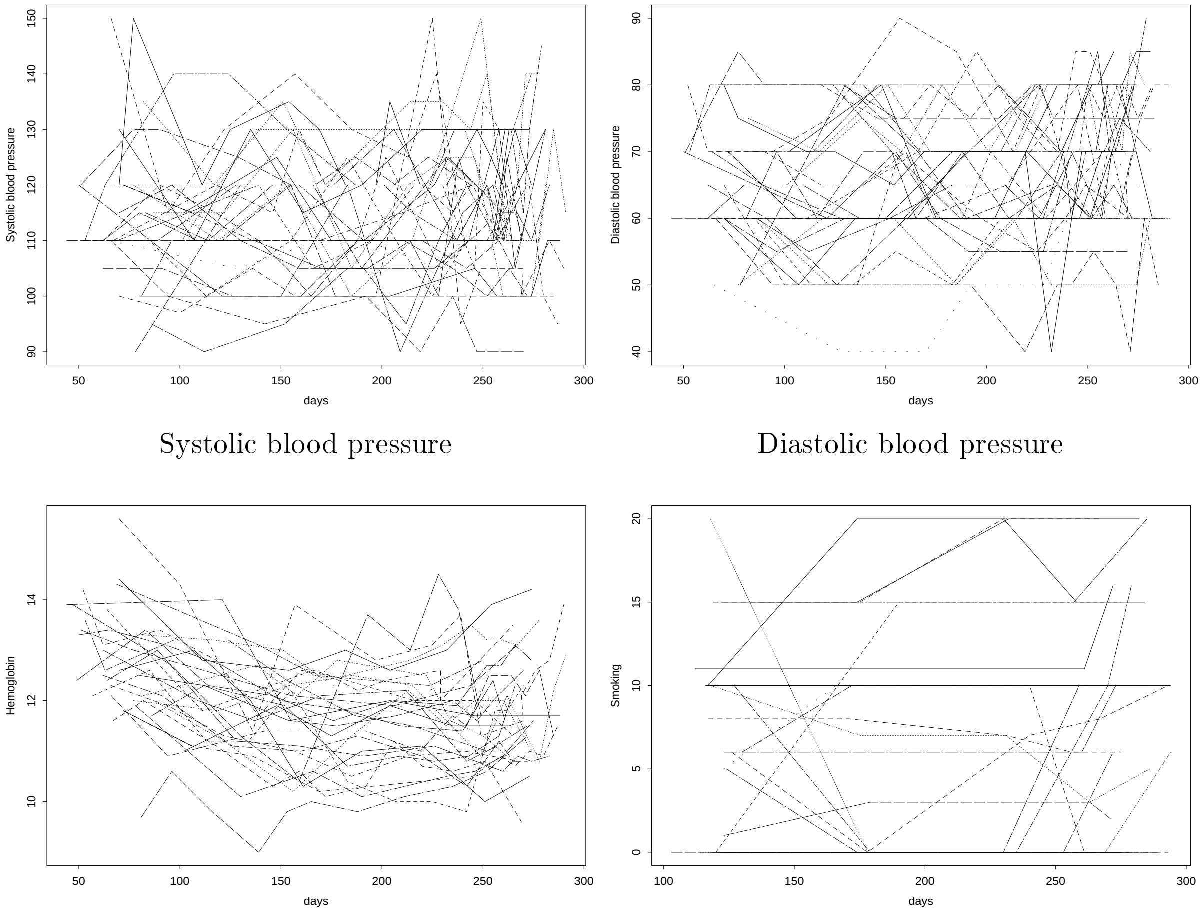
The four variational variables are shown in figure 1 for 10 randomly chosen individuals from each of the study sites.

From visual inspection the curves (for the 1020 individuals) were not found to have any apparent trend of quadratic, cubic, logarithmic and inverse form. Random coefficient regression models are fit for each variational curve *linear* in time using the generalized least squares estimators of section A.2 of the appendix. In addition for the haemoglobin also a quadratic curve in time was fitted. Estimates of the random coefficient regression model parameters, $\phi = (\boldsymbol{\Theta}, \sigma^2, \boldsymbol{\Lambda})$, are found in table 3. The estimated slope for all variational variables are close to zero, that is, the linear trend in time is small. This is also the case for the quadratic term for haemoglobin.

Summary Statistics for the Prepregnancy Explanatory Variables for the TBU-SGA Data Set													
Variable		Total				SGA				Non-SGA			
		Total	T	B	U	Total	T	B	U	Total	T	B	U
births	count	1011	422	203	386	152	53	28	71	859	369	175	315
	perc	100 %	41.7%	20.1 %	38.2 %	15.0 %	12.6 %	13.8 %	18.4 %	85.0 %	87.4 %	86.2 %	81.6 %
age	min	18	19	18	19	19	21	22	19	18	19	18	19
	medi	29	29	29	30	28	28	28.5	28	29	29	29	30
	mean	29.32	28.9	29.35	29.76	28.71	28.15	29.21	28.93	29.43	29.01	29.37	29.95
	max	41	41	38	41	40	40	37	40	41	41	38	41
	std	4.17	4.07	3.93	4.36	4.56	4.27	4.37	4.85	4.09	4.04	3.86	4.23
bmi	min	16.02	16.02	16.41	16.2	16.22	16.94	16.59	16.22	16.02	16.02	16.41	16.2
	medi	20.96	20.8	20.69	21.71	20.07	19.38	19.72	21.19	21.11	20.82	20.82	21.77
	mean	21.51	21.16	21.14	22.09	20.74	20.78	19.99	21.02	21.65	21.21	21.32	22.33
	max	38.83	36.16	33.09	38.83	36.16	36.16	24.77	27.05	38.83	33.22	33.09	38.83
	std	3.08	2.97	2.82	3.24	2.85	3.77	2.01	2.27	3.10	2.84	2.89	3.38
p.sga	perc	20.4 %	18.2 %	24.1 %	20.8 %	43.4 %	45.3 %	35.7 %	45.1 %	16.4 %	14.3 %	22.3 %	15.5 %
cig.	min	0	0	0	0	0	0	0	0	0	0	0	0
cons.	medi	5	8	6	4	10	10	9	10	5	6	5	1
	mean	6.92	7.36	7.29	6.26	8.82	9.68	8.11	8.45	6.59	7.02	7.16	5.77
	max	60	60	25	40	25	25	20	25	60	60	25	40
	std	7.25	7.11	7.36	7.31	6.97	6.73	6.80	7.25	7.25	7.12	7.46	7.24

Table 2: Summary statistics for the prepregnancy explanatory variables for the TBU-SGA data set. The columns marked "Total" include data from all sites and columns "T", "B", "U" denote data from Trondheim, Bergen and Uppsala, respectively. The summary statistics are the minimum ("min"), the median ("medi"), the mean ("mean"), the maximum ("max"), the count ("count"), the percent ("perc") and the standard deviation ("std").

Individual Variational Variable Curves



Systolic blood pressure

Diastolic blood pressure

Haemoglobin

Smoking (Number of cigarettes per day)

Figure 1: Individual curves for the variational variables for 10 randomly chosen individuals from each of the study sites. The x-axis give the gestational age in days. The standard number of measurements for the blood pressures and the haemoglobin is 5 to 25, and for smoking it is 5. Lines are drawn between successive measurements for each individual.

Generalized Least Squares Random Coefficient Regression Estimates for the Variational and Growth Explanatory Variables

Curve	$\hat{\mathbf{\Theta}}$ SGA			$\hat{\mathbf{\Theta}}$ non-SGA			$\hat{\sigma}^2$	$\hat{\mathbf{\Lambda}}$		
	Trondheim	Bergen	Uppsala	Trondheim	Bergen	Uppsala				
BPA	111.5018	112.10820	110.89872	109.90091	108.86362	114.91317	51.700	100.22926		-0.18794
(lin)	0.0077	-0.00623	0.02148	0.01713	0.02064	0.00932		[-0.56]		0.00114
BPB	61.68989	63.94418	62.24726	64.41853	58.91573	63.01865	37.272	70.79347		-0.16615
(lin)	0.02920	0.00313	0.04561	0.01998	0.03357	0.04245		[-0.64]		0.00094
HB	12.25491	12.32126	12.61865	12.39332	12.66811	12.51286	0.454	1.11494		-0.00382
(lin)	-0.00175	-0.00200	-0.00300	-0.00278	-0.00391	-0.00332		[-0.80]		2.043e-05
HB	14.64814	14.38726	14.22602	14.32082	14.81428	14.53226	0.00027	7.02990	-7.178e-02	1.823e-04
(quad)	-0.03273	-0.02996	-0.02618	-0.02846	-0.03171	-0.02975		[-0.91]	8.9104e-04	-2.4219e-06
	8.55e-05	7.93e-05	6.80e-05	7.21e-05	7.79e-05	7.54e-05		[0.83]	[-0.98]	6.9053e-09
CIG	7.66556	6.25779	8.20765	5.42395	5.08122	4.60826	2.829	41.92515		-0.04541
(lin)	-0.00034	0.00085	-0.00062	-0.00227	-0.00177	-0.00061		[-0.41]		0.00029
MAD	-13.844	-14.412	-17.676	-19.586	-21.185	-21.922	5.728	6.90079		-0.053896
(lin)	0.439	0.430	0.455	0.489	0.484	0.497		[-0.87]		0.000562
BPD	-285.190	-289.511	-282.815	-296.600	-299.045	-294.529	3.733	179.280		-36.587
(log)	67.927	68.701	66.842	70.359	70.826	69.501		[-0.99]		7.588
FEM	-249.326	-264.253	-267.709	-262.957	-272.725	-272.734	2.352	147.579		-28.312
(log)	57.314	59.858	60.688	60.283	61.778	61.933		[-0.99]		5.541
SFH	-91.822	-91.030	-95.379	-94.085	-97.635	-102.300	1.536	98.592		-18.556
(log)	22.330	22.105	22.948	22.960	23.570	24.559		[-0.99]		3.546
PWG	-0.04095	-0.04575	-0.04528	-0.04943	-0.06515	-0.04524	0.000427	2.243e-03		-9.099e-06
(lin)	0.00105	0.00100	0.00099	0.00115	0.00119	0.00108		[-0.54]		1.248e-07

Table 3: Parameter estimates (intercept, slope, curvature) for the variational and growth curves for the SGA data set of 1020 individuals. The abbreviations (lin),(log) and (quad) denote that the individual regression models are linear, logarithmic and quadratic in time, respectively. The abbreviations in the column denoted "Curve" are found in table 1. The generalized least squares parameter estimates are $\hat{\mathbf{\Theta}}$ (location parameter estimate), $\hat{\sigma}^2$ (variance estimate) and $\hat{\mathbf{\Lambda}}$ (the estimated covariance matrix of the coefficients). The elements in square brackets [ ] in $\hat{\mathbf{\Lambda}}$ are the correlations.

Results for Testing Hypotheses for SGA and non-SGA Births
$H_0 : \mathbf{\Theta}_{SGA} = \mathbf{\Theta}_{Non-SGA}$

	All Sites			Trondheim			Bergen			Uppsala		
	curve	int.	slope	curve	int.	slope	curve	int.	slope	curve	int.	slope
BPA	••	•	••	∘	∘	∘	•	∘	••	•	•	•
BPB	••	•	••	∘	∘	∘	••	•	••	∘	∘	∘
HB	∘	∘	∘	∘	∘	∘	∘	∘	∘	∘	∘	∘
CIG	••	••	∘	•	•	∘	∘	∘	∘	••	••	∘
MAD	••	••	••	••	••	••	••	••	••	••	••	••
BPD	••	••	••	••	••	••	••	•	•	••	••	••
FEM	••	••	••	••	••	••	••	•	••	••	•	••
SFH	••	••	••	••	∘	∘	••	•	••	••	••	••
PWG	•	∘	••	∘	∘	∘	•	∘	••	∘	∘	∘

Table 4: Results from testing the H_0 hypothesis of equal variational and growth curves for SGA and non-SGA births based on the test statistics for linear hypotheses of the location parameter of appendix A.3. The test is performed for all sites together and for each site separately and for both intercept and slope together (denoted "curve") and for intercept and slope separately (denoted "int." and "slope"). Rejection of the H_0 hypothesis at the 1% level is denoted by •• and at the 5% level by •. No rejection is denoted ∘. The abbreviations in the first column are found in table 1. The row denoted HB gives the result for both the linear and the quadratic haemoglobin curves.

The random coefficient regression location parameters have been tested using the approximate tests of section A.3 for testing linear hypotheses about the location parameter $\mathbf{\Theta}$. Results from the tests for equivalence between SGA and non-SGA births are found in table 4.

Based on testing the hypothesis $H_0 : \mathbf{\Theta}_{SGA} = \mathbf{\Theta}_{nonSGA}$, the variational curves for the systolic and diastolic blood pressure and for smoking are found to be different for the SGA and the non-SGA births (P-values below 0.01). For separate test for slope and intercept of SGA and non-SGA births, both slope and intercept are found to be different for the systolic and diastolic blood pressure (P-values below 0.003). In addition the intercept for smoking is found to be different (P-value below 0.0002). Separate tests are also conducted for equivalence of SGA and non-SGA curve parameters at the three study cites. More details are found in table 4.

The effect of study site (for SGA and non-SGA births together) has also been studied. The hypotheses $H_0 : \mathbf{\Theta}_{citeA} = \mathbf{\Theta}_{citeB}$ for all combinations of sites (Trondheim versus Bergen, Trondheim versus Uppsala and Bergen versus Uppsala) have also been tested. The mean haemoglobin and smoking curves are not found to be different in any of the study sites for all births. The same result is found for separate testing of slope and intercept. The mean systolic blood pressure curves are found to be different for Trondheim and Uppsala and for Bergen and Uppsala, but not for Trondheim and Bergen. Separate tests for mean slope and intercept gave different mean slopes in Bergen and Uppsala and different mean intercepts in Trondheim and Uppsala and in Bergen and Uppsala. The mean diastolic blood pressure curves are found to be different in all study sites, and are found to be different for the mean slopes for all study sites, and for mean intercept in Trondheim and Bergen and in Bergen and Uppsala.

2.3 Analysis of Growth Curves

There are a total of five growth variables, three foetal ultrasound measurements, SF height and proportional weight gain, see table 1. The number of ultrasound measurements are four, situated around gestational week 17, 25, 33 and 37. The number of SF height measurements is between 5 and 25. Early in the pregnancy the SF height is measured every second week and in the end of the pregnancy the measurements are taken every week. [3] The weight of the pregnant woman was measured together with the SF height. Since growth variables are measured at different points in time and for different number of times for each individual also here we have unbalanced data. There are missing data in the data set.

The five growth variables are shown in figure 2 for 10 randomly chosen individuals from each study site. These are the same individuals as for the variational curves in figure 1. To be able to trace one pregnancy, lines are drawn between successive measurements.

From the visual inspection of the growth curves (for the 1020 individuals) it is found that the trend of the femur length (FEM), the biparietal diameter (BPD) and the SF height curves are slightly concave and the mean abdominal diameter (MAD) and proportional weight gain (PWG) curves are relatively linear.

For the five individual growth variables regression using linear, quadratic, cubic, logarithmic and inverse terms has been tried. The best choices (measured by P-values and R^2) are found to be a linear model in time for the mean abdominal diameter (MAD) and the proportional weight gain (PWG) measurement and a linear model in the logarithm of time for the biparietal diameter (BPD), the femur length (FEM) and the SF height measurements. For a visual inspection of the estimated individual slopes and intercepts for the SGA and non-SGA births of the randomly chosen sample of 10 individuals from each study site, see figure 3. In appendix B.1 the estimated individual slopes and intercepts for the SGA and non-SGA births are shown for Trondheim, Bergen and Uppsala in figures 5, 6 and 7, respectively. Observe the strong correlation between the estimated intercept and slope for the femur length (FEM), the biparietal diameter (BPD) and the SF height growth curves. For the mean abdominal diameter (MAD) plot we see that in general the slopes are smaller and the intercepts are larger for SGA births than for non-SGA births. From this visual inspection the mean abdominal diameter (MAD) appears to be the best single discriminatory variable between the SGA and the non-SGA births.

The five measured growth variables are studied using random coefficient regression models with unbalanced data, see appendix A. Estimates of the random coefficient regression model parameters, $\boldsymbol{\phi} = (\boldsymbol{\Theta}, \sigma^2, \boldsymbol{\Lambda})$, are found using the generalized least squares estimators of section A.2 of the appendix. These estimates are given in table 3. The random coefficient regression location parameters have been tested using the approximate tests of section A.3 for testing linear hypotheses.

[3] The SF height measures were made by different health personnel categories in Trondheim, Bergen and Uppsala; general practitioners, obstetricians and midwifes respectively.

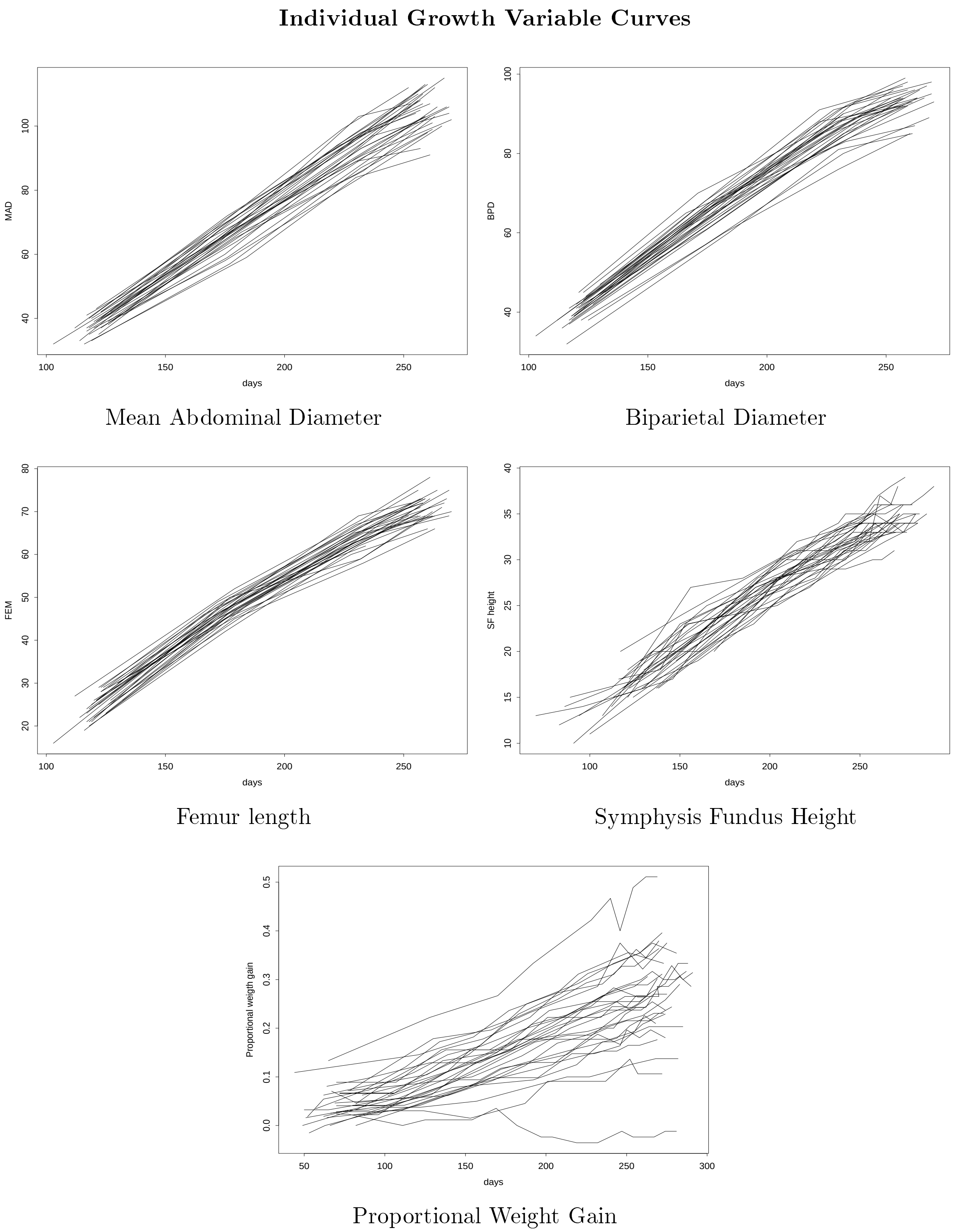


Figure 2: Individual curves for the growth variables for 10 randomly chosen individuals from each of the study sites. These are the same individuals that are shown in figure 1. The x-axis give the gestational age in days. Lines are drawn between successive measurements for each individual.

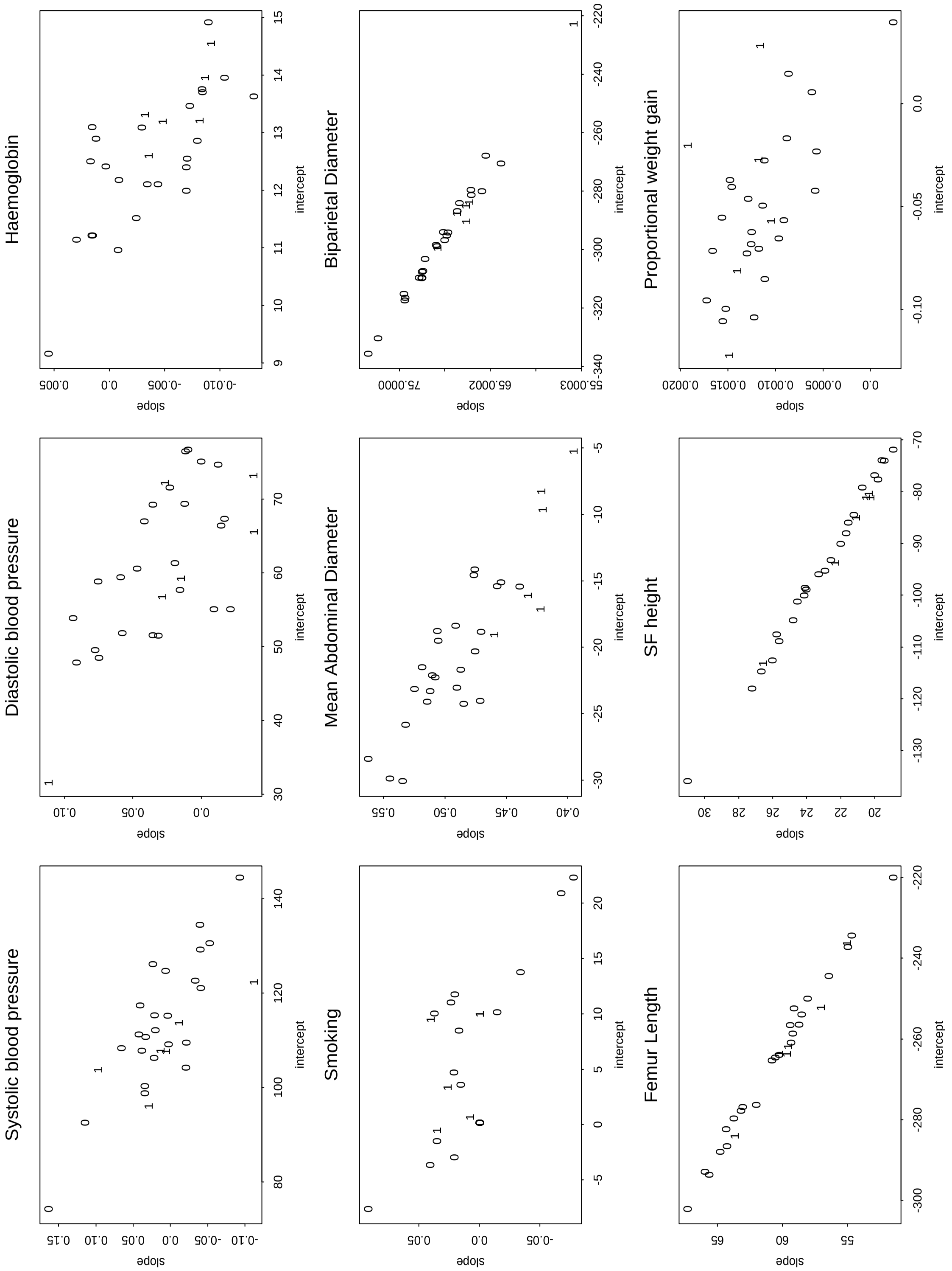


Figure 3: Individual regression coefficient estimates (least squares) for the variational and growth curves for 10 randomly chosen individuals from each study site. The numbers “0” denote non-SGA births and the numbers “1” denote SGA births.

Based on testing the hypothesis $H_0 : \mathbf{\Theta}_{SGA} = \mathbf{\Theta}_{nonSGA}$, all mean growth curves are found to be different for the SGA and the non-SGA births with P-values of 0 for the three foetal ultrasound measures and the SF height and below 0.02 for the proportional weight gain. For separate tests for slope and intercept of SGA and non-SGA births, both slope and intercept are found to be different for the three ultrasound measures and the SF height (P-values below 0.00005). In addition the slope for the proportional weight gain is found to be different for SGA and non-SGA births (P-value below 0.004). For more details and separate tests for each study site see table 4.

The effect of study site (for SGA and non-SGA births together) has also been studied using the hypothesis $H_0 : \mathbf{\Theta}_{citeA} = \mathbf{\Theta}_{citeB}$ for all combinations of the study sites. Using a significance level of 5% the mean mean abdominal diameter (MAD), femur length (FEM) and SF height curves are found to be different in all the study sites for all births. The mean biparietal diameter (BPD) curves are found to be different in Trondheim and Uppsala and in Bergen and Uppsala. The mean proportional weight gain (PWG) curves are found to be different in Trondheim and Bergen. Separate testing of mean slope and intercept have also been conducted. The SF height mean slopes and intercept are found to be different for all study sites. The mean biparietal diameter (BPD) intercept is not found to be different in any of the study sites, but the mean slopes in Trondheim and Uppsala and in Bergen and Uppsala are found to be different. The mean mean abdominal diameter (MAD) intercept is found to be different in all the study sites, and the slope is found to be different in Trondheim and Uppsala and in Bergen and Uppsala. The mean femur length (FEM) slope and intercept are found to be different in Trondheim and Bergen and in Trondheim and Uppsala. The mean proportional weight gain (PWG) intercept is found to be different for Trondheim and Bergen and for Bergen and Uppsala, and the slope is found to be different for Trondheim and Uppsala and for Bergen and Uppsala. This is consistent with the results found in Jacobsen (1992) using repeated measures analysis of variance with balanced data.

The variables of primary interest in Jacobsen (1992) were the SF height and the ultrasound measures because the overall objective of the analyses was to compare the efficacy in birth outcome prediction between serial SF height and ultrasound measurements. In Jacobsen (1992) each growth variable was presented in the classification using a projected measure. The projected measure was the predicted (extrapolated) value at birth found using a second order polynomial with equal slope and curvature but different intercept for each pregnancy. This means that the slope part of the regression model was regarded as fixed coefficients for all pregnancies (both SGA and non-SGA births) and was estimated as the average slope for all pregnancies at each study site, that is, different slopes for Trondheim, Bergen and Uppsala. The intercept was regarded as a random coefficient.

2.4 Conclusions

It is not convenient to include all the variational and growth variables as explanatory variables for the discrimination and classification. There are several reasons for this. The variational and growth variables are measured repeatedly for unequal number of times for the different pregnancies. The repeated measurements are highly correlated in time and would not provide efficient information. We therefore introduce derived variables in the further analyses.

From the random coefficient regression model analyses of the variational and growth variables some conclusions with respect to discriminatory information can be drawn.

From the analyses of the variational variables there do not seem to be much discriminatory information in the linear and quadratic haemoglobin curves. The intercept of the linear cigarette consumption curves have some discriminatory information, but this information is correlated with the prepregnancy variable of cigarette consumption around conception (correlation coefficient of 0.7). Both the slope and intercept of the systolic and diastolic blood pressure curves are found to contain discriminatory information. But since the slope and the intercept are correlated (see table 3) and the slope is an easier quantity to give a physical interpretation, the slope of the systolic and diastolic blood pressure are chosen to be part of the explanatory variables included in the further discrimination and classification studies.

From the analyses of the growth variables there seem to be discriminatory information in all of the growth variables. Since all the slopes and the intercepts are correlated (see table 3) and again the slope is an easier quantity to give a physical interpretation, the slopes of the growth curves are preferred in the further discrimination and classification studies

It is of interest to be able to predict an SGA birth as early in the pregnancy as possible and not be dependent on measurements taken late in the pregnancy. In the further analyses available data at five stages of the pregnancy is studied.

- At time of birth.
- At the late third trimester:
  Gestational age of at least 273 days (39 weeks).
- At the early third trimester:
  Gestational age of at least 231 days (33 weeks).
- At the second trimester:
  Gestational age of at least 154 days (22 weeks).
- At time of conception.

3 Discrimination and Classification

Based on the analyses of the variables at study we will here construct discrimination and classification rules for the TBU-SGA data set. Different methods for discrimination and classification are applied. These methods are classical and modern statistical method for discrimination and classification (linear and quadratic discriminant analysis as described in Johnson and Wichern (1982), classification trees of Breiman et al. (1984), classification using projection pursuit regression of Friedman and Stuetzle (1981) and flexible discriminant analysis of Hastie et al. (1994)), methods from artificial neural networks (multilayer perceptrons as described in Hertz et al. (1991)) and mixture codebook classification (MC^2) methods of Langaas (1995b and 1995c). Short presentations of the methods used here for discrimination and classification are found in Langaas (1995a, 1995b and 1995c) and details on performing the discrimination and classification analyses in this section are found in appendix B.2.

For the TBU-SGA data minimizing the total error rate is not an interesting solution to the classification problem. This is because the loss of misclassifying an SGA birth is greater than the loss of misclassifying a non-SGA birth. It is, however, difficult to quantify this difference in loss. We will consider the following loss function, $L(\omega, \delta)$, which defines the loss of classifying an individual from class ω to the class δ. That is, the true class is denoted ω and the predicted class is denoted δ. Let the two classes of SGA births and non-SGA births be denoted ω_{SGA} and ω_{nonSGA}, respectively.

$$L(\omega, \delta) = \begin{cases} 0 & \text{if } \delta = \omega \\ A & \text{if } \delta = \omega_{nonSGA} \text{ and } \omega = \omega_{SGA} \\ 1 & \text{if } \delta = \omega_{SGA} \text{ and } \omega = \omega_{nonSGA} \end{cases}$$

We define the cost of classifying an individual with explanatory variables $\boldsymbol{x}$ to class δ to be $C(\delta|\boldsymbol{x})$.

$$C(\delta|\boldsymbol{x}) = \sum_{\forall \omega} L(\omega, \delta) p(\omega|\boldsymbol{x})$$

where $p(\omega|\boldsymbol{x})$ denotes the posterior probability for class ω. Given an observation $\boldsymbol{x}$ we will here choose to classify to the class giving the lowest cost of misclassification. Since we only have two classes, ω_{SGA} and ω_{nonSGA}, this means that we classify to class ω_{SGA} if

$$C(\delta = \omega_{nonSGA}|\boldsymbol{x}) \geq C(\delta = \omega_{SGA}|\boldsymbol{x})$$

$$L(\omega = \omega_{SGA}, \delta = \omega_{nonSGA}) p(\omega_{SGA}|\boldsymbol{x}) \geq L(\omega = \omega_{nonSGA}, \delta = \omega_{SGA}) p(\omega_{nonSGA}|\boldsymbol{x})$$

$$p(\omega_{SGA}|\boldsymbol{x}) \geq \frac{1}{A+1}$$

This classification rule will minimize the expected cost of misclassification, see Johnson and Wichern (1982). Based on trial we will here, somewhat arbitrarily, choose $A = 5$ and classify as an SGA birth if $p(\omega_{SGA}|\boldsymbol{x}) \geq \frac{1}{6}$.

The data set of 1011 individuals has been randomly divided into a training set of size 611 and a test set of size 400. The training set contains 93 SGA births and 518 non-SGA births and the test set contains 59 SGA births and 341 non-SGA births. The classification rules are constructed based on the training set. The test set is used for comparing and evaluating the rules. For each discrimination and classification method only the best model choice (and training scheme) is quoted, that is, the model choice giving the lowest average cost (defined below in equation (1)) on the test set and a low average cost on the training set. Details on performing the model selection and on constructing the discrimination and classification rules are found in appendix B.2. The fact that the model selection is based on the test set slightly biases the analyses and must be kept in mind when comparing the performance of the different discrimination and classification rules. A better strategy for training and testing would be to use two layers of 10-fold cross-validation; the first layer for model selection and the second layer for method evaluation. We propose what this can be done in further analyses of the SGA data set.

The sensitivity of a classification method is the percent of correctly classified SGA births. The specificity is the percent of correctly classified non-SGA births. For each discrimination and classification method the error rates for the total training and test set are given together with the error rates for the group of SGA births and the group of non-SGA births. In addition the average cost and approximate standard deviations for the average cost are given. The average cost is defined as,

$$C_{ave} = 5 \cdot \frac{N(\text{SGA errors})}{N(\text{total})} + 1 \cdot \frac{N(\text{non-SGA errors})}{N(\text{total})} \tag{1}$$

where $N(\text{SGA errors})$ is the number of SGA births that has been misclassified, $N(\text{non-SGA errors})$ is the number of non-SGA births that has been misclassified and $N(\text{total})$ is the total number of observations (611 for the training set and 400 for the test set). The two expressions for the approximate standard deviation of the average cost are the following. The first, C_{sd1}, is based on assuming a trinomial distribution for the number of SGA errors, the number of non-SGA errors and the number of correct classifications,

$$\begin{aligned} C_{\text{sd1}} = \frac{1}{N(\text{total})} \{ & 5^2 \cdot N(\text{total}) \, \frac{N(\text{SGA errors})}{N(\text{total})}(1 - \frac{N(\text{SGA errors})}{N(\text{total})}) \\ & +1^2 \cdot N(\text{total}) \frac{N(\text{non-SGA errors})}{N(\text{total})}(1 - \frac{N(\text{non-SGA errors})}{N(\text{total})}) \\ & -2 \cdot 5 \cdot N(\text{total}) \frac{N(\text{SGA errors})}{N(\text{total})} \frac{N(\text{non-SGA errors})}{N(\text{total})} \}^{\frac{1}{2}} \end{aligned} \tag{2}$$

The second approximation for the standard deviation of the average cost, C_{sd2}, is based on using a binomial distribution for each class conditional on the number of SGA and non-SGA births.

$$\begin{aligned} C_{\text{sd2}} = \frac{1}{N(\text{total})} \{ & 5^2 \cdot N(\text{SGA}) \frac{N(\text{SGA errors})}{N(\text{SGA})}(1 - \frac{N(\text{SGA errors})}{N(\text{SGA})}) \\ & 1^2 \cdot N(\text{non-SGA}) \frac{N(\text{non-SGA errors})}{N(\text{non-SGA})}(1 - \frac{N(\text{non-SGA errors})}{N(\text{non-SGA})}) \}^{\frac{1}{2}} \end{aligned} \tag{3}$$

where $N(\text{SGA})$ is the number of SGA births in the data set (93 for the training set and 59 for the test set) and $N(\text{non-SGA})$ is the number of non-SGA births in the data set (511 for the training set and 341 for the test set).

We will now construct discrimination and classification rules at five stages of the pregnancy; at birth, at the late third trimester, at the early third trimester, at the second trimester and at time of conception.

3.1 At Birth

We will first look at the TBU-SGA data set at time of birth. As explanatory variables we use the prepregnancy variables and the slope of the growth and variational variables that were found to contain discriminatory information in the random coefficient regression analyses. The prepregnancy variables are the age, the body mass index (BMI), an indicator denoting previous SGA births, the number of cigarettes smoked around conception and indicator variables denoting study site (for Trondheim and Bergen, and therefore also indirectly for Uppsala). All growth variable slopes (ultrasound, SF height and proportional weight gain) and the slopes for blood pressure (systolic and diastolic) are used as explanatory variables.

For each pregnancy a regression model (linear or logarithmic in time) is estimated separately for each of the chosen growth and variational variables using all available data from this pregnancy and the estimated slopes are used in the further analyses. The individual regressions are linear in time for the blood pressures, the mean abdominal diameter (MAD) and the proportional weight gain (PWG). The individual regressions are linear in the logarithm of time for the biparietal diameter (BPD), the femur length (FEM) and the SF height.

Different statistical, neural network and mixture codebook classification (MC^2) methods for discrimination and classification are applied to the "at birth" data set and the results are given in table 5. Using the average cost on the test set as performance measure, the projection pursuit regression is found to give the best results of the statistical methods with a sensitivity of 71.2 % and a specificity of 83.9 %. It is closely followed by linear discrimination analysis with a sensitivity of 76.3 % and a specificity of 78.3 %. The performance of the three multilayer perceptrons are very similar and give sensitivities and specificities around 80%. For the MC^2 methods maximizing the joint likelihood for Student-t class-conditional densities with equal diagonal covariance matrices gives a sensitivity of 81.4 % and a specificity of 76.8 %.

The pruned classification tree produces a very simple classification rule; classify as an SGA birth if the slope of the mean abdominal diameter is less than 0.453635 or if the slope of the mean abdominal diameter lies in [0.454635, 0.513656] and the age of the pregnant woman is less than 23.5 years. This simple rule has a sensitivity of 62.7 % and a specificity of 82.7 % on the test set. Comparing this rule with each of the other classification rules gives no significant difference on a 5 % significance level when the comparisons are based on the average cost on the test set.

3.2 Late Third Trimester

We here look at available information in the late third trimester, gestational age of 273 days (39 weeks). At this stage of the pregnancy at least two measurements are present for all the variational and growth variables for all individuals in the TBU-SGA data set. Therefore, the same explanatory variables as for the analyses at birth are used. The only difference being that the individual growth and variational slopes are estimated from all the measurements available at the gestational age of 273 days.

Table 6 contains results from discrimination and classification based on data at the late third trimester. The average costs on the test set are in general only slightly higher than for the classification using available data at time of birth. The best results are, as for the data at birth, found for the projection pursuit regression, multilayer perceptron with sigmoid activation function and entropy loss and MC^2 maximizing the joint likelihood. These methods have a sensitivity and a specificity on the test set around 76-80 %.

Also here the pruned classification tree gives a simple classification rule, which is presented in figure 4. This rule is only dependent on the slope of the mean abdominal diameter, the slope of the diastolic blood pressure, the age of the pregnant woman and whether the pregnant women has had a previous SGA birth. The rule produced by the pruned classification tree has a sensitivity of 72.9 % and a specificity of 77.4 % on the test set. Comparing this rule with each of the other classification rules give no significant difference on a 5 % significance level when the comparisons are based on the average cost on the test set.

Classification of SGA Births Using Available Data at Birth												
Method	Training Set (611,93,518)						Test Set (400,59,341)					
	Error Rates (%)			Cost			Error Rates (%)			Cost		
	tot	SGA	non	ave	sd1	sd2	tot	SGA	non	ave	sd1	sd2
MAD slope <= 0.453635	14.2	28.0	11.8	31.3	4.2	3.7	16.5	42.4	12.0	41.5	6.1	5.0
Linear Discriminant Analysis	18.3	19.4	18.2	30.1	3.6	3.4	22.0	23.7	21.7	36.0	4.8	4.5
Quadratic Discriminant Analysis	15.9	11.8	16.6	23.1	3.0	2.9	25.2	28.8	24.6	42.2	5.2	4.8
Classification Tree (default)	9.2	2.2	10.4	10.5	1.6	1.6	24.2	45.8	20.5	51.2	6.3	5.1
ditto (pruned)	15.5	20.4	14.7	28.0	3.7	3.4	20.2	37.3	17.3	42.2	5.8	5.0
Projection Pursuit Regression (2)	14.4	18.3	13.7	25.5	3.5	3.3	18.0	28.8	16.1	35.0	5.2	4.7
Flexible DA with MARS (1)	14.4	28.0	12.0	31.4	4.2	3.7	16.0	35.6	12.6	37.0	5.7	4.8
ditto with MARS (2)	9.0	34.4	4.4	30.0	4.5	3.8	16.0	49.2	10.3	45.0	6.5	5.0
Multilayer Perceptron 12-1-1 (lin)	17.5	18.3	17.4	28.6	3.5	3.4	22.2	22.0	22.3	35.2	4.7	4.4
Multilayer Perceptron 12-4-1 (sig)	16.4	17.2	16.2	26.8	3.4	3.3	19.2	23.7	18.5	33.2	4.8	4.5
Multilayer Perceptron 12-5-1 (ent)	16.0	22.6	14.9	29.8	3.8	3.6	20.2	18.6	20.5	31.2	4.4	4.2
MC^2 JOINT Gaussian (c) K=(1,1)	21.3	23.7	20.8	35.7	3.9	3.7	21.8	20.3	22.0	33.8	4.5	4.3
MC^2 JOINT Gaussian (s) K=(1,1)	20.5	22.6	20.1	34.2	3.9	3.6	22.0	20.3	22.3	34.0	4.5	4.3
MC^2 JOINT Student-t (c) K=(1,1)	21.8	23.7	21.4	36.2	3.9	3.7	22.5	18.6	23.2	33.5	4.4	4.2
MC^2 JOINT Student-t (s) K=(1,1)	22.9	18.3	23.8	34.0	3.6	3.4	25.8	17.0	27.3	35.8	4.3	4.1
MC^2 COND Gaussian (c) K=(2,1)	26.8	44.1	23.8	53.7	5.1	4.2	25.8	27.1	25.5	41.8	5.1	4.7
MC^2 COND Student-t (c) K=(1,1)	20.5	21.5	20.3	33.6	3.8	3.6	23.2	32.2	21.7	42.2	5.5	4.9

Table 5: Error rates (%) and costs for classification rules using available data at birth. The number of observations of the training and the test sets are given in parenthesis in the heading (total number of observations, number of SGA births, number of non-SGA births). The columns denoted "tot", "SGA" and "non" give the error rates (%) for the total data, the SGA births and the non-SGA births, respectively. The columns "ave" give the average cost (1) and "sd1" and "sd2" give the approximate standard deviations of the average costs given in (2) and (3), respectively. The classification rule based on the slope of the mean abdominal diameter alone is constructed by fitting a classification tree using the mean abdominal diameter as the only explanatory variable and then using only the first tree split to construct the classification rule; classifying as an SGA birth if the slope of the mean abdominal diameter is less than or equal to 0.453635. The number in parenthesis for the projection pursuit regression gives the number of projections used, and for the flexible discriminant analyses the degrees of interaction for MARS. For the multilayer perceptrons the network architecture (for example 12-1-1) and the choice of output activation and loss function (for example "lin") is given, see appendix B.2. For the MC^2 method the following is given; the choice of criterion function (JOINT or COND), the choice of component distributions (Gaussian or Student-t), the use of common (c) or class specific (s) diagonal covariance matrices and the component architecture (for example K=(2,1)). Based on the average costs on the test set no comparisons of any two methods concludes with significant difference on a 5 % significance level. Pairwise comparisons based on the total error rate for the test set have a 95 % confidence interval of half-length about 1.3 %, assuming that all the errors made by the method with the lower error rate are made by the method with the higher error rate.

Classification of SGA Births Using Available Data at the Late Third Trimester												
Method	Training Set (611,93,518)						Test Set (400,59,341)					
	Error Rates (%)			Cost			Error Rates (%)			Cost		
	tot	SGA	non	ave	sd1	sd2	tot	SGA	non	ave	sd1	sd2
MAD slope <= 0.453635	14.1	29.0	11.4	31.8	4.2	3.8	16.5	42.4	12.0	41.5	6.1	5.0
Linear Discriminant Analysis	18.5	21.5	18.0	31.6	3.8	3.5	21.5	23.7	21.1	35.5	4.8	4.5
Quadratic Discriminant Analysis	15.2	9.7	16.2	21.1	2.7	2.7	25.8	30.5	24.9	43.8	5.4	4.9
Classification Tree (default)	10.0	0.0	11.8	10.0	1.2	1.2	24.5	42.4	21.4	49.5	6.1	5.1
ditto (pruned)	17.0	9.7	18.3	22.9	2.8	2.7	23.2	27.1	22.6	39.2	5.1	4.7
Projection Pursuit Regression (2)	20.1	16.1	20.8	30.0	3.4	3.3	23.2	20.3	23.8	35.2	4.6	4.3
Flexible DA with MARS (1)	14.7	29.0	12.2	32.4	4.2	3.8	16.2	35.6	12.9	37.2	5.7	4.9
ditto with MARS (2)	10.3	36.6	5.6	32.6	4.7	3.9	15.5	45.8	10.3	42.5	6.3	5.0
Multilayer Perceptron 12-1-1 (lin)	17.8	17.2	18.0	28.3	3.4	3.3	22.0	22.0	22.0	35.0	4.7	4.4
Multilayer Perceptron 12-1-1 (sig)	17.5	18.3	17.4	28.6	3.5	3.4	20.2	22.0	19.9	33.2	4.7	4.4
Multilayer Perceptron 12-4-1 (ent)	15.7	17.2	15.4	26.2	3.4	3.3	20.0	22.0	19.6	33.0	4.7	4.4
MC^2 JOINT Gaussian (c) K=(2,1)	18.7	21.5	18.2	31.8	3.8	3.5	20.5	23.7	19.9	34.5	4.8	4.5
MC^2 JOINT Gaussian (s) K=(1,1)	20.3	21.5	20.1	33.4	3.8	3.6	21.8	20.3	22.0	33.8	4.5	4.3
MC^2 JOINT Student-t (c) K=(1,1)	22.3	22.6	22.2	36.0	3.9	3.6	23.2	18.6	24.0	34.2	4.4	4.2
MC^2 JOINT Student-t (s) K=(1,1)	22.6	16.1	23.8	32.4	3.4	3.3	26.0	17.0	27.6	36.0	4.3	4.2
MC^2 COND Gaussian (c) K=(2,1)	21.9	26.9	21.0	38.3	4.2	3.8	23.2	25.4	22.9	38.2	5.0	4.6
MC^2 COND Student-t (c) K=(1,1)	25.5	38.7	23.2	49.1	4.8	4.2	26.8	37.3	24.9	48.8	5.8	5.1

Table 6: Error rates (%) and costs for classification rules using available data at the late third trimester (day 273, week 39). The number of observations of the training and the test sets are given in parenthesis in the heading (total number of observations, number of SGA births, number of non-SGA births). The columns denoted "tot", "SGA" and "non" give the error rates (%) for the total data, the SGA births and the non-SGA births, respectively. The columns "ave" give the average cost (1) and "sd1" and "sd2" give the approximate standard deviations of the average costs given in (2) and (3), respectively. The classification rule based on the slope of the mean abdominal diameter is identical to the one found using available data at birth, and is found in the same way. Based on the average costs on the test set no comparisons of any two methods concludes with significant difference on a 5 % significance level. Pairwise comparisons based on the total error rate for the test set have a 95 % confidence interval of half-length about 1.3 %, assuming that all the errors made by the method with the lower error rate are made by the method with the higher error rate.

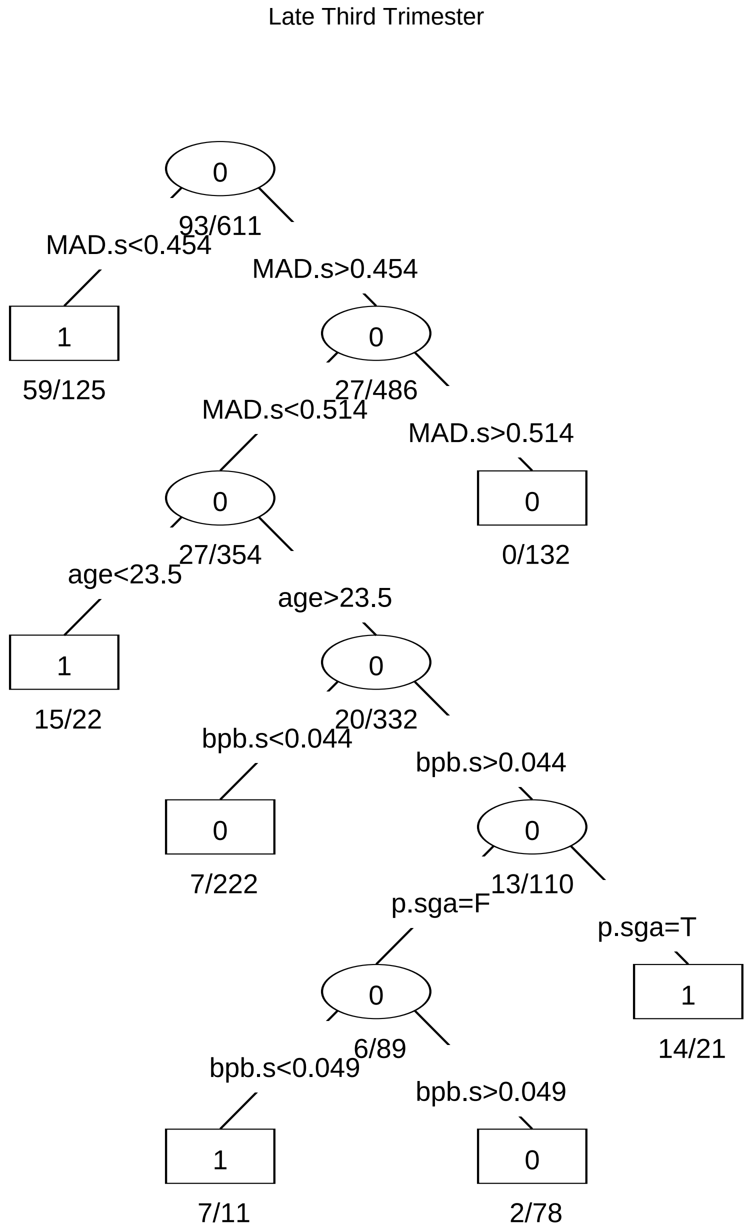


Figure 4: This is a classification tree for SGA and non-SGA births based on the explanatory variables at the late third trimester. Here "MAD.s" denotes the mean abdominal diameter (MAD) slope, "age" the age of the pregnant woman, "bpb.s" the slope of the diastolic blood pressure and "p.sga" whether the woman has had a previous SGA birth or not. At each node of the tree a question is asked, and the individuals for which the answer is "yes" are assigned to the left branch of the tree and the others to the right branch. The rectangular boxes are the terminal nodes of the tree, and the oval boxes are internal nodes. The number inside each node is the classification. A "0" denotes a non-SGA birth node and "1" an SGA birth node. This classification tree gives a sensitivity of 72.9 % and a specificity of 77.4 % on the test set. The fraction below each node is the misclassification rate for the training set at this node. For more details on classification trees see Clark and Pregibon (1992).

3.3 Early Third Trimester

At the gestational age of 231 days (33 weeks) there are many individuals for which only one (or none) measurement is available for some growth variables. In addition to the prepregnancy variables we therefore use as explanatory variables the average value for the growth variables (ultrasound, SF height and proportional weight gain) and the blood pressure (diastolic and systolic). For three individuals no SF height measurement was taken up to the gestational age of 231 days and for these three individuals the first SF height measurement taken after this day was included as an explanatory variable (these three measurements were taken at 238, 242 and 254 days).

Discrimination and classification results for data until the early third trimester are found in table 7. As expected, these results are not as good as those found at the late third trimester. Based on the average cost on the test set the best results are as follows. For the statistical methods the projection pursuit regression is the best, but has a low specificity (49.8 %) on the test set. The neural network multilayer perceptron with sigmoid activation function and entropy loss has produced slightly better results, with a sensitivity of 72.9 % and a specificity of 75.4 % on the test set. For the MC^2 methods good results are found maximizing the conditional likelihood with two Gaussian component densities with equal diagonal matrices of covariance for both classes. This MC^2 method has a sensitivity of 67.8 % and a specificity of 72.4 %.

3.4 Second Trimester

We will here look at available information from the pregnancy at the gestational age of 154 days (22 weeks). At this stage most of the variational and growth variables have been measured once. This is however not the case for the SF height were 83 individuals had no measured SF height at that stage. The SF height is therefore not included in the explanatory variables. For the blood pressure only one individual was without measurement and for the ultrasound (MAD, FEM and BPD) 12 individuals were without measurements. As explanatory variables we use the prepregnancy variables and the mean for the ultrasound, proportional weight gain and blood pressure measurements available at the gestational age of 154 days. For the individuals with missing measurements the first observation after 154 days is included (in the worst case at 178 days).

Table 8 contains results from discrimination and classification at the second trimester. These results are in general slightly worse than those found for the data at early third trimester. Based on the average cost on the test set again projection pursuit gives the best classification results of the statistical methods. The projection pursuit regression has a sensitivity of 72.9 % and a specificity of 59.5 %. For the neural network methods the best results are found using the multilayer perceptron with quadratic loss function and linear output activation. For the MC^2 methods the best result is found maximizing the joint likelihood with class-conditional Student-t distributions with class specific diagonal matrices of covariance. This MC^2 method has a sensitivity of 66.1 % and a specificity of 72.4 %.

Classification of SGA Births Using Available Data at the Early Third Trimester

Method	Training Set (611,93,518)						Test Set (400,59,341)					
	Error Rates (%)			Cost			Error Rates (%)			Cost		
	tot	SGA	non	ave	sd1	sd2	tot	SGA	non	ave	sd1	sd2
Average MAD <= 63.1667	45.8	26.9	49.2	62.2	4.2	4.0	44.0	18.6	48.4	55.0	4.5	4.4
Linear Discriminant Analysis	25.0	37.6	22.8	48.0	4.8	4.1	27.0	44.1	24.0	53.0	6.2	5.2
Quadratic Discriminant Analysis	23.4	24.7	23.2	38.5	4.0	3.8	30.0	44.1	27.6	56.0	6.2	5.2
Classification Tree (default)	12.3	2.2	14.1	13.6	1.7	1.7	31.5	52.5	27.9	62.5	6.7	5.2
ditto (pruned)	25.2	23.7	25.5	39.6	4.0	3.7	31.8	35.6	31.1	52.8	5.7	5.1
Projection Pursuit Regression (1)	39.0	9.7	44.2	44.8	3.0	3.0	44.5	11.9	50.2	51.5	3.9	3.9
Flexible DA with MARS (1)	23.4	35.5	21.2	45.0	4.7	4.1	30.5	42.4	28.4	55.5	6.1	5.2
ditto with MARS (2)	18.0	37.6	14.5	40.9	4.8	4.0	26.0	49.2	22.0	55.0	6.5	5.2
Multilayer Perceptron 12-7-1 (lin)	24.7	26.9	24.3	41.1	4.2	3.8	27.2	30.5	26.7	45.2	5.4	4.9
Multilayer Perceptron 12-6-1 (sig)	25.5	32.3	24.3	45.2	4.5	4.0	25.8	35.6	24.0	46.8	5.7	5.0
Multilayer Perceptron 12-2-1 (ent)	24.1	28.0	23.4	41.1	4.2	3.9	25.0	27.1	24.6	41.0	5.1	4.7
MC^2 JOINT Gaussian (c) K=(2,2)	27.8	34.4	26.6	48.8	4.6	4.1	27.2	35.6	25.8	48.2	5.7	5.0
MC^2 JOINT Gaussian (s) K=(2,2)	29.1	33.3	28.4	49.4	4.5	4.1	31.0	33.9	30.5	51.0	5.6	5.0
MC^2 JOINT Student-t (c) K=(2,2)	29.8	34.4	29.0	50.7	4.6	4.1	28.8	35.6	27.6	49.8	5.7	5.0
MC^2 JOINT Student-t (s) K=(2,2)	31.1	30.1	31.3	49.4	4.4	4.0	32.2	30.5	32.6	50.2	5.4	4.9
MC^2 COND Gaussian (c) K=(2,2)	28.3	34.4	27.2	49.3	4.6	4.1	28.2	32.2	27.6	47.2	5.5	4.9
MC^2 COND Student-t (c) K=(1,1)	18.2	68.8	9.1	60.1	6.2	3.8	18.2	59.3	11.1	53.2	7.1	4.9

Table 7: Error rates (%) and costs for classification rules using available data at the early third trimester (day 231, week 33). The number of observations of the training and the test sets are given in parenthesis in the heading (total number of observations, number of SGA births, number of non-SGA births). The columns denoted “tot”, “SGA” and “non” give the error rates (%) for the total data, the SGA births and the non-SGA births, respectively. The columns “ave” give the average cost (1) and “sd1” and “sd2” give the approximate standard deviations of the average costs given in (2) and (3), respectively. The classification rule based on the average of the available mean abdominal diameter measurements classifies as an SGA birth if the average mean abdominal diameter is less than or equal to 63.1667. This rule is constructed by fitting a classification tree to the average mean abdominal diameter alone and using only the first two splits. Based on the average costs on the test set no comparisons of any two methods concludes with significant difference on a 5 % significance level. Pairwise comparisons based on the total error rate for the test set have a 95 % confidence interval of half-length about 1.3 %, assuming that all the errors made by the method with the lower error rate are made by the method with the higher error rate.

Classification of SGA Births Using Available Data at the Second Trimester												
	Training Set (611,93,518)						Test Set (400,59,341)					
Method	Error Rates (%)			Cost			Error Rates (%)			Cost		
	tot	SGA	non	ave	sd1	sd2	tot	SGA	non	ave	sd1	sd2
Average MAD <= 39.5	54.2	32.3	58.1	73.8	4.4	4.1	49.8	27.1	53.7	65.8	5.1	4.8
Linear Discriminant Analysis	24.4	47.3	20.3	53.2	5.3	4.2	25.5	45.8	22.0	52.5	6.3	5.2
Quadratic Discriminant Analysis	24.7	31.2	23.6	43.7	4.4	4.0	28.5	42.4	26.1	53.5	6.1	5.2
Classification tree (default)	14.2	1.1	16.6	14.9	1.6	1.6	33.5	59.3	29.0	68.5	7.0	5.2
ditto (pruned)	33.4	25.8	34.7	49.1	4.1	3.9	41.0	32.2	42.5	60.0	5.5	5.0
Projection Pursuit Regression (1)	32.6	23.7	34.2	47.0	4.0	3.8	38.5	27.1	40.5	54.5	5.1	4.8
Flexible DA with MARS (1)	24.7	36.6	22.6	47.0	4.7	4.1	31.2	50.8	27.9	61.2	6.6	5.2
ditto with MARS (2)	19.8	32.3	17.6	39.4	4.5	4.0	30.0	61.0	24.6	66.0	7.1	5.1
Multilayer Perceptron 12-1-1 (lin)	28.6	34.4	27.6	49.6	4.6	4.1	29.8	32.2	29.3	48.8	5.5	5.0
Multilayer Perceptron 12-5-1 (sig)	24.2	29.0	23.4	41.9	4.3	3.9	29.0	33.9	28.2	49.0	5.6	5.0
Multilayer Perceptron 12-2-1 (ent)	26.8	43.0	23.9	53.0	5.1	4.2	28.5	40.7	26.4	52.5	6.0	5.1
MC^2 JOINT Gaussian (c) K=(1,1)	22.6	49.5	17.8	52.7	5.4	4.2	24.5	45.8	20.8	51.5	6.3	5.1
MC^2 JOINT Gaussian (s) K=(1,1)	23.2	51.6	18.2	54.7	5.5	4.2	27.5	49.2	23.8	56.5	6.5	5.2
MC^2 JOINT Student-t (c) K=(1,1)	28.6	45.2	25.7	56.1	5.2	4.3	28.5	33.9	27.6	48.5	5.6	5.0
MC^2 JOINT Student-t (s) K=(1,1)	27.8	38.7	25.9	51.4	4.8	4.2	29.8	32.2	29.3	48.8	5.5	5.0
MC^2 COND Gaussian (c) K=(1,1)	26.0	43.0	23.0	52.2	5.1	4.2	26.5	39.0	24.3	49.5	5.9	5.1
MC^2 COND Student-t (c) K=(1,1)	28.5	44.1	25.7	55.3	5.1	4.2	31.8	40.7	30.2	55.8	6.0	5.2

Table 8: Error rates (%) and costs for classification rules using available data at the mid second trimester (day 154, week 22). The number of observations of the training and the test sets are given in parenthesis in the heading (total number of observations, number of SGA births, number of non-SGA births). The columns denoted “tot”, “SGA” and “non” give the error rates (%) for the total data, the SGA births and the non-SGA births, respectively. The columns “ave” give the average cost (1) and “sd1” and “sd2” give the approximate standard deviations of the average costs given in (2) and (3), respectively. The classification rule based on the average of the available mean abdominal diameter measurements classifies as an SGA birth if the average mean abdominal diameter is less than or equal to 39.5. This rule is constructed by fitting a classification tree to the average mean abdominal diameter alone and using the first split only. Based on the average costs on the test set no comparisons of any two methods concludes with significant difference on a 5 % significance level. Pairwise comparisons based on the total error rate for the test set have a 95 % confidence interval of half-length about 1.3 %, assuming that all the errors made by the method with the lower error rate are made by the method with the higher error rate.

3.5 At Time of Conception

We will conclude the discrimination and classification analyses by regarding the available data at time of conception. The available data is the prepregnancy variables; age, the body mass index (BMI), an indicator denoting previous SGA births, the number of cigarettes smoked around conception and indicator variables denoting study site (for Trondheim and Bergen, and therefore also indirectly for Uppsala).

Results from the analyses of the prepregnancy variables are found in table 9. Based on the average cost on the test set also here projection pursuit regression is found to be the best statistical classification method. The projection pursuit regression has a sensitivity of 71.2 % and a specificity of 64.2 %. The best neural network is the multilayer perceptron with entropy loss function and sigmoid output activation. This multilayer perceptron has a sensitivity of 61.0 % and a specificity of 75.7 %. The MC^2 models with Student-t component distributions give good results maximizing both the conditional and the joint likelihood. The conditional Student-t MC^2 model has the lower average cost, but the joint Student-t MC^2 model with class specific diagonal covariance matrices for the classes and two components for the non-SGA class and three components for the SGA class give more balanced sensitivity and specificity on the test set. This classification rule has a sensitivity of 66.1 % and a specificity of 67.7 %.

The pruned classification tree give at time of conception a very simple classification rule; classify as an SGA birth if one of the following three criteria is true: the pregnant woman has had a previous SGA birth, her age is less than 23.5 years or she has a body mass index (BMI) less than 19.4. This simple rule has a sensitivity of 67.8 % and a specificity of 57.5 % on the test set. Comparing this rule with each of the other classification rules give no significant difference on a 5 % significance level when the comparisons are based on the average cost on the test set. This rule is also the rule of the pruned classification tree using available data at the second trimester.

4 Conclusions and Comments

In this study of the Scandinavian TBU-SGA data set discrimination and classification rules for predicting an SGA birth are formed at different stages of the pregnancy. The explanatory variables for the TBU-SGA data set have been grouped into prepregnancy variables, growth variables and variational variables. The prepregnancy variables have been studied using two sample t-tests and the growth and variational variables have been studied using random coefficient regression models.

The random coefficient regression analyses are performed using generalized least squares estimates of the parameters (appendix A.2). From these analyses the mean growth curves and the mean blood pressure and smoking curves are found to be different for the SGA and the non-SGA births. From examining the plots of the estimated location parameters of the random coefficient regression analyses, the ultrasound mean abdominal diameter is believed to have the strongest discriminatory effect for discriminating between SGA and non-SGA births. This is consistent with the results found in Jacobsen (1992). However, the ultrasound mean abdominal diameter (MAD) alone will not be able to classify more than 57.6 % of the SGA births correctly using a linear rule based on the slope of the available mean abdominal diameter measurements until time of birth. This and the following results are based on the

Classification of SGA Births Using Available Data at Time of Conception

Method	Training Set (611,93,518)						Test Set (400,59,341)					
	Error Rates (%)			Cost			Error Rates (%)			Cost		
	tot	SGA	non	ave	sd1	sd2	tot	SGA	non	ave	sd1	sd2
Linear Discriminant Analysis	23.1	55.9	17.2	57.1	5.6	4.2	24.2	47.5	20.2	52.2	6.4	5.1
Quadratic Discriminant Analysis	26.2	40.9	23.6	51.1	4.9	4.2	33.8	40.7	32.6	57.8	6.0	5.2
Classification Tree (default)	21.4	2.2	24.9	22.7	2.0	2.0	36.0	50.8	33.4	66.0	6.6	5.3
ditto (pruned)	33.4	25.8	34.7	49.1	4.1	3.9	41.0	32.2	42.5	60.0	5.5	5.0
Projection Pursuit Regression (1)	29.5	35.5	28.4	51.1	4.7	4.1	34.8	28.8	35.8	51.8	5.2	4.9
Flexible DA with MARS (1)	25.5	40.9	22.8	50.4	4.9	4.2	34.0	45.8	32.0	61.0	6.3	5.2
ditto with MARS (2)	23.2	47.3	18.9	52.0	5.3	4.2	28.8	49.2	25.2	57.8	6.5	5.2
Multilayer Perceptron 12-1-1 (lin)	24.1	52.7	18.9	56.1	5.5	4.2	24.2	45.8	20.5	51.2	6.3	5.1
Multilayer Perceptron 12-3-1 (sig)	25.5	45.2	22.0	53.0	5.2	4.2	28.5	40.7	26.4	52.5	6.0	5.1
Multilayer Perceptron 12-3-1 (ent)	25.0	47.3	21.0	53.8	5.3	4.2	26.5	39.0	24.3	49.5	5.9	5.1
MC^2 JOINT Gaussian (c) K=(1,1)	22.9	55.9	17.0	57.0	5.6	4.2	24.2	49.2	19.9	53.2	6.5	5.1
MC^2 JOINT Gaussian (s) K=(1,1)	23.1	55.9	17.2	57.1	5.6	4.2	25.8	50.8	21.4	55.8	6.6	5.2
MC^2 JOINT Student-t (c) K=(1,2)	30.6	34.4	29.9	51.6	4.6	4.1	35.0	30.5	35.8	53.0	5.4	4.9
MC^2 JOINT Student-t (s) K=(2,3)	27.8	36.6	26.2	50.1	4.7	4.1	32.5	33.9	32.3	52.5	5.6	5.0
MC^2 COND Gaussian (c) K=(1,1)	24.6	50.5	19.9	55.3	5.4	4.2	25.5	42.4	22.6	50.5	6.1	5.1
MC^2 COND Student-t (c) K=(2,1)	22.9	60.2	16.2	59.6	5.8	4.1	21.5	47.5	17.0	49.5	6.4	5.1

Table 9: Error rates (%) and costs for classification rules using available data at time of conception. The number of observations of the training and the test sets are given in parenthesis in the heading (total number of observations, number of SGA births, number of non-SGA births). The columns denoted "tot", "SGA" and "non" give the error rates (%) for the total data, the SGA births and the non-SGA births, respectively. The columns "ave" give the average cost (1) and "sd1" and "sd2" give the approximate standard deviations of the average costs given in (2) and (3), respectively. Based on the average costs on the test set no comparisons of any two methods concludes with significant difference on a 5 % significance level. Pairwise comparisons based on the total error rate for the test set have a 95 % confidence interval of half-length about 1.3 %, assuming that all the errors made by the method with the lower error rate are made by the method with the higher error rate.

performance when the classifiers are constructed using a training set of size 611 and model choice and evaluation are done on a test set of size 400. The use of the test set for both model choice and evaluation slightly biases the comparisons.

Using available data at time of conception, the prepregnancy variables, gives a sensitivity of around 60 % and a specificity of around 70 %. Including data from the first 33 weeks of the pregnancy do not improve substantially upon the classification. Based on available data at the second trimester, after 22 weeks, a good classification rule has a sensitivity of 66.1 % and a specificity of 72.4 %. At the early third trimester, after 33 weeks, a good classification rule gives a sensitivity of 72.9 % and a specificity of 75.4 %. At the second and early third trimester mean values for the growth variables and the blood pressure are included in addition to the prepregnancy variables.

The classification is improved when available data at the late third trimester, after 39 weeks, is included. At this stage the explanatory variables used are the pregnancy variables together with the individual (linear) regression slopes for the growth variables and the blood pressure measurements. The best classification rule has a sensitivity and a specificity of around 80 %. Including additional data after 39 weeks until the time of birth do not improve substantially on the classification.

In the present study methods for predicting an SGA birth are constructed based on data collected at different stages of the pregnancy. Hopefully, this has provided valuable insight into the study of SGA births. The conclusions drawn in this study may also be used as starting points for future analyses.

Acknowledgements

The first author would like to thank her supervisor Bo Lindqvist for helpful discussions and suggestions. This report was written while the first author was funded by a Dr.Ing. grant from the Norwegian Research Council. The study was supported by the U.S. National Institute of Child Health and Human Development, NIH (NICHD Contract No 1-HD-4-2803).

A Random Coefficient Regression Models

This is a short presentation of the theory of random coefficient regression models with unbalanced data. Random coefficient regression models are often also called growth curve models and random effects models for longitudinal data.

Both direct (noniterative) and iterative methods for estimation of model parameters are presented. These are based on generalized least squares estimates and maximum likelihood estimates via the Expectation-Maximization (EM) algorithm, respectively. Methods for testing hypotheses about the model parameters are also presented.

An early review of random coefficients models is found in Spjøtvoll (1977). Other references to random coefficients models are Swamy (1970), Johansen (1984), Laird and Ware (1982), Jennrich and Schluchter (1986), Carter and Yang (1986) and Vonesh and Carter (1987).

A.1 The Random Coefficient Regression Model

The random coefficient regression model with unbalanced data can be written in two stages; a "within-individual" and a "between-individual" model, see Vonesh and Carter (1987).

The first stage defines a linear regression model for each individual.

$$\boldsymbol{y}_j = \boldsymbol{X}_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \qquad j = 1, ..., n \tag{4}$$

where $\boldsymbol{y}_j$ is a $[m_j \times 1]$ vector of repeated measurements, $\boldsymbol{X}_j$ is a $[m_j \times p]$ matrix of fixed explanatory variables, $\boldsymbol{\beta}_j$ is a $[p \times 1]$ vector of unobserved random regression coefficients and $\boldsymbol{\varepsilon}_j$ is a $[m_j \times 1]$ unobserved vector of random errors. The number of observations for each individual equals m_j and may be different for each individual, and therefore the data is called unbalanced.

Further, the errors $\boldsymbol{\varepsilon}_j$ are assumed to be i.i.d. Gaussian with mean zero and nonsingular covariance matrix $\boldsymbol{\Sigma}_j = \sigma^2 \boldsymbol{I}_{m_j}$, where $\boldsymbol{I}_{m_j}$ denotes the $m_j \times m_j$ identity matrix.

$$\boldsymbol{\varepsilon}_j \sim N_{m_j}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{m_j}) \qquad j = 1, ..., n \tag{5}$$

With this covariance structure the model is called a conditional-independence model since it implies that the m_j responses for individual j are independent, conditional on the coefficients $\boldsymbol{\beta}_j$, Laird and Ware (1982). Other covariance structures than the conditional-independence can be found in Jennrich and Schluchter (1986) for unbalanced repeated-measurements models.

At the second stage we have a "between-individual" multivariate linear model.

$$\boldsymbol{\beta}_j = \boldsymbol{\Theta} \boldsymbol{Z}_j + \boldsymbol{\xi}_j \qquad j = 1, ..., n \tag{6}$$

where $\boldsymbol{\Theta}$ is a $[p \times q]$ matrix of fixed location parameters, $\boldsymbol{Z}_j$ is a $[q \times 1]$ vector of "between-individual" one-time explanatory variables ($q < n$) and $\boldsymbol{\xi}_j$ is a $[p \times 1]$ vector of random errors.

Further, the errors $\boldsymbol{\xi}_j$ are assumed to be i.i.d. Gaussian with mean zero and nonsingular covariance matrix $\boldsymbol{\Lambda}$.

$$\boldsymbol{\xi}_j \sim N_p(\boldsymbol{0}, \boldsymbol{\Lambda}) \qquad j = 1, ..., n \tag{7}$$

For simplicity we will here restrict $\boldsymbol{Z}_j$ to be a 0-1 indicator vector denoting group membership. This is done both for computational and practical reasons, and is the situation for the SGA variational and growth data analysed in this report. Let z_{ji} be the ith element of vector $\boldsymbol{Z}_j$.

$$z_{ji} = \begin{cases} 1 & \text{if individ } j \text{ belongs to group } i \\ 0 & \text{else} \end{cases}$$

That is, $\boldsymbol{Z}_j = (1, 0, 0, ..., 0)^T$ if individual j belongs to group one, and so forth. The matrix $\boldsymbol{\Theta}$ consists of column vectors for each group. That is, $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2 | \cdots | \boldsymbol{\theta}_q)$. Here $\boldsymbol{\theta}_i$ denotes the fixed location parameters for group i. The situation with only one group, that is, $q = 1$ and $\boldsymbol{Z}_j = 1$ for $\forall j$, is a simplification and is included in the above model. The more general case then $\boldsymbol{Z}_j$ is a vector of one-time covariates representing either group indicator variables and/or fixed covariates is treated in Vonesh and Carter (1987). Time dependent "between-individual" covariates are not treated here, but are included in Laird and Ware (1982).

We see that population parameters, individual effects and "within-individual" variation are introduced at stage one and "between-individual" variation at stage two.

A.2 Estimation

The parameters of interest are $\boldsymbol{\phi} = (\boldsymbol{\Theta}, \boldsymbol{\Lambda}, \sigma^2)$.

Generalized Least Squares Estimation

The main references for this section are Swamy (1970), Spjøtvoll (1977), Carter and Yang (1986) and Vonesh and Carter (1987).

The individual regression coefficients, $\boldsymbol{\beta}_j$ can be estimated by the least squares estimator, $\hat{\boldsymbol{\beta}}_j$,

$$\hat{\boldsymbol{\beta}}_j = (\boldsymbol{X}_j^T \boldsymbol{X}_j)^{-1} \boldsymbol{X}_j^T \boldsymbol{y}_j$$

having unconditional distribution

$$\hat{\boldsymbol{\beta}}_j \sim N_p(\boldsymbol{\Theta} \boldsymbol{Z}_j, \Lambda + \sigma^2 (\boldsymbol{X}_j^T \boldsymbol{X}_j)^{-1})$$

The variances σ^2 can be estimated from the individual regression residual sum of squares

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^n (\boldsymbol{y}_j - \boldsymbol{X}_j \hat{\boldsymbol{\beta}}_j)^T (\boldsymbol{y}_j - \boldsymbol{X}_j \hat{\boldsymbol{\beta}}_j)}{(\sum_{j=1}^n m_j) - np} \tag{8}$$

The problem of estimating the parameters $\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$ remains. If $\boldsymbol{\Lambda}$ and σ^2 were known then $\boldsymbol{\theta}_i$ could be estimated using the generalized least squares estimate $\boldsymbol{\theta}_i^*$.

$$\boldsymbol{\theta}_i^* = [\sum_{j=1}^n z_{ji} (\boldsymbol{\Lambda} + \sigma^2 (\boldsymbol{X}_j^T \boldsymbol{X}_j)^{-1})^{-1}]^{-1} \sum_{j=1}^n z_{ji} (\boldsymbol{\Lambda} + \sigma^2 (\boldsymbol{X}_j^T \boldsymbol{X}_j)^{-1})^{-1} \hat{\boldsymbol{\beta}}_j \qquad i = 1, ..., q$$

$$\boldsymbol{\theta}_i^* \sim N_p(\boldsymbol{\theta}_i, [\sum_{j=1}^n z_{ji} (\boldsymbol{\Lambda} + \sigma^2 (\boldsymbol{X}_j^T \boldsymbol{X}_j)^{-1})^{-1}]^{-1})$$

This is the maximum likelihood estimator, Swamy (1970), and also the unbiased estimate with the uniformly minimum variance (UMVU) when the data is balanced.

However, $\boldsymbol{\Lambda}$ and σ^2 are unknown. A more straightforward estimate of $\boldsymbol{\theta}_i$ is the average of the individual regression coefficients, $\hat{\boldsymbol{\theta}}_i$.

$$\hat{\boldsymbol{\theta}}_i = \frac{1}{\sum_{j=1}^n z_{ji}} \sum_{j=1}^n z_{ji} \hat{\boldsymbol{\beta}}_j = \frac{1}{n_i} \sum_{j=1}^n z_{ji} \hat{\boldsymbol{\beta}}_j$$

$$\hat{\boldsymbol{\theta}}_i \sim N_p(\boldsymbol{\theta}_i, \frac{1}{n_i} \boldsymbol{\Lambda} + \frac{\sigma^2}{n_i^2} \sum_{j=1}^n z_{ji} (\boldsymbol{X}_j^T \boldsymbol{X}_j)^{-1})$$

Here n_i is the number of individuals belonging to group i. This is the best estimator of $\boldsymbol{\theta}_i$ when $\sigma^2 = 0$, that is, the unbiased estimator with uniformly minimum variance (UMVU), Spjøtvoll (1977). We then use the fact that $Cov(\hat{\boldsymbol{\beta}}_j) = \Lambda + \sigma^2(\boldsymbol{X}_j^T\boldsymbol{X}_j)^{-1}$ and the estimates $\hat{\sigma}^2$ and $\hat{\boldsymbol{\theta}}_i$ of σ^2 and $\boldsymbol{\theta}_i$ to construct the method of moments estimator of $\boldsymbol{\Lambda}$, Vonesh and Carter (1987).

$$\hat{\boldsymbol{\Lambda}} = \frac{1}{n-q}\sum_{j=1}^{n}\sum_{i=1}^{q} z_{ji}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\theta}}_i)(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\theta}}_i)^T - \hat{\sigma}^2\frac{1}{n-q}\sum_{j=1}^{n}(1 - \frac{1}{n_{Z_j}})(\boldsymbol{X}_j^T\boldsymbol{X}_j)^{-1} \equiv \boldsymbol{S} - \hat{\sigma}^2\boldsymbol{T} \quad (9)$$

where n_{Z_j} denotes the number of individuals belonging to the group of individual j. Notice that for $\hat{\boldsymbol{\theta}}_i$ given above $\sum_{j=1}^{n}\sum_{i=1}^{q} z_{ji}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\theta}}_i)\hat{\boldsymbol{\theta}}_i^T = 0$. The estimate $\hat{\boldsymbol{\Lambda}}$ might turn out to be nonpositive semidefinite (and can even be negative on the diagonal), for example if a linear regression model gives a bad fit to the individual data the term $\hat{\sigma}^2$ will be large compared to the first term of $\hat{\boldsymbol{\Lambda}}$. This can be overcome by considering the modified estimator $\tilde{\boldsymbol{\Lambda}}$ of Carter and Yang (1986) and Vonesh and Carter (1987). First, let $\hat{\zeta}$ be the smallest root of

$$|\boldsymbol{S} - \zeta\boldsymbol{T}| = |\frac{1}{n-q}\sum_{j=1}^{n}\sum_{i=1}^{q} z_{ji}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\theta}}_i)(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\theta}}_i)^T - \zeta\frac{1}{n-q}\sum_{j=1}^{n}(1 - \frac{1}{n_{Z_j}})(\boldsymbol{X}_j^T\boldsymbol{X}_j)^{-1}|$$

If the matrix $\boldsymbol{S}$ is nonsingular then $1/\hat{\zeta}$ can be found as the largest eigenvalue of $\boldsymbol{S}^{-1}\boldsymbol{T}$. The new estimator $\tilde{\boldsymbol{\Lambda}}$ depends on the value of $\hat{\zeta}$ compared to the value of $\hat{\sigma}^2$.

$$\tilde{\boldsymbol{\Lambda}} = \begin{cases} \hat{\boldsymbol{\Lambda}} & \text{if } \hat{\zeta} > \hat{\sigma}^2 \\ \boldsymbol{S} - \hat{\zeta}\boldsymbol{T} & \text{if } \hat{\zeta} \leq \hat{\sigma}^2 \end{cases} \quad (10)$$

It is then possible to define a new estimator for $\boldsymbol{\theta}_i$ based on the estimator $\boldsymbol{\theta}_i^*$ inserted the estimators of $\tilde{\boldsymbol{\Lambda}}$ and $\hat{\sigma}^2$ for $\boldsymbol{\Lambda}$ and σ^2.

$$\tilde{\boldsymbol{\theta}}_i = [\sum_{j=1}^{n} z_{ji}(\tilde{\boldsymbol{\Lambda}} + \hat{\sigma}^2(\boldsymbol{X}_j^T\boldsymbol{X}_j)^{-1})^{-1}]^{-1}\sum_{j=1}^{n} z_{ji}(\tilde{\boldsymbol{\Lambda}} + \hat{\sigma}^2(\boldsymbol{X}_j^T\boldsymbol{X}_j)^{-1})^{-1}\hat{\boldsymbol{\beta}}_j \quad (11)$$

An estimator of the covariance of $\tilde{\boldsymbol{\theta}}_i$ is $\widehat{Cov}(\tilde{\boldsymbol{\theta}}_i)$.

$$\widehat{Cov}(\tilde{\boldsymbol{\theta}}_i) = [\sum_{j=1}^{n} z_{ji}(\tilde{\boldsymbol{\Lambda}} + \hat{\sigma}^2(\boldsymbol{X}_j^T\boldsymbol{X}_j)^{-1})^{-1}]^{-1} \quad (12)$$

Maximum Likelihood Estimation

From equation (4) and (6) we have the marginal distribution for the responses.

$$\boldsymbol{y}_j \sim N_{m_j}(\boldsymbol{X}_j\boldsymbol{\Theta}\boldsymbol{Z}_j, \boldsymbol{X}_j\boldsymbol{\Lambda}\boldsymbol{X}_j^T + \sigma^2\boldsymbol{I}_{m_j})$$

The loglikelihood of the observed data is denoted $l(\boldsymbol{\phi}|(\boldsymbol{y}_j)_{j=1}^n)$.

$$l(\boldsymbol{\phi}|(\boldsymbol{y}_j)_{j=1}^n) = \sum_{j=1}^{n}\{\log|\boldsymbol{X}_j\boldsymbol{\Lambda}\boldsymbol{X}_j^T + \sigma^2\boldsymbol{I}_{m_j}|^{-\frac{1}{2}} - \frac{m_j}{2}\log 2\pi$$

$$-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}_j\boldsymbol{\Theta}\boldsymbol{Z}_j)^T(\boldsymbol{X}_j\boldsymbol{\Lambda}\boldsymbol{X}_j^T + \sigma^2\boldsymbol{I}_{m_j})^{-1}(\boldsymbol{y} - \boldsymbol{X}_j\boldsymbol{\Theta}\boldsymbol{Z}_j)\}$$

The maximum likelihood estimates of the parameters can be found by maximizing the total likelihood of the observed data. This can be done by various optimization techniques, among others the Newton-Raphson and Fisher scoring algorithms. The maximum likelihood estimates can also be found using the EM algorithm of Dempster et al. (1977). The EM algorithm augments the observed data by unobserved data and proceeds by creating a complete data loglikelihood based on the augmented data. The expectation step (E-step) of the algorithm calculates the expected complete data loglikelihood given the observed data and the current estimate of the parameters. The maximization step (M-step) maximizes this likelihood with respect to the parameters. The E- and M-steps are alternated until convergence. We will here let the observed data $(\boldsymbol{y}_j)_{j=1}^n$ be augmented by the unobserved $(\boldsymbol{\beta}_j)_{j=1}^n$ to yield the complete data $(\boldsymbol{y}_j, \boldsymbol{\beta}_j)_{j=1}^n$. The complete data loglikelihood is denoted $l(\boldsymbol{\phi}|(\boldsymbol{y}_j, \boldsymbol{\beta}_j)_{j=1}^n)$,

$$
\begin{aligned}
l(\boldsymbol{\phi}|(\boldsymbol{y}_j, \boldsymbol{\beta}_j)_{j=1}^n) &= \log \prod_{j=1}^n f(\boldsymbol{y}_j, \boldsymbol{\beta}_j|\boldsymbol{\phi}) = \log \prod_{j=1}^n f(\boldsymbol{y}_j|\boldsymbol{\beta}_j, \boldsymbol{\phi}) f(\boldsymbol{\beta}_j|\boldsymbol{\phi}) \\
&= n \log |\boldsymbol{\Lambda}^{-1}|^{\frac{1}{2}} - \frac{1}{2}\sum_{j=1}^n (\boldsymbol{\beta}_j - \boldsymbol{\Theta}\boldsymbol{Z}_j)^T \boldsymbol{\Lambda}^{-1} (\boldsymbol{\beta}_j - \boldsymbol{\Theta}\boldsymbol{Z}_j) - \frac{1}{2}\sum_{j=1}^n m_j \log \sigma^2 \\
&\quad - \frac{1}{2\sigma^2} \sum_{j=1}^n (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}_j)^T (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}_j) + C
\end{aligned}
$$

where C is a constant. From the random coefficient regression model we find the distribution of the unobserved $\boldsymbol{\beta}_j$ conditional on the observed data $\boldsymbol{y}_j$.

$$
\begin{aligned}
\boldsymbol{\beta}_j|\boldsymbol{y}_j &\sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\gamma}_j) \\
\boldsymbol{\mu}_j &= \boldsymbol{\gamma}_j[\boldsymbol{\Lambda}^{-1}\boldsymbol{\Theta}\boldsymbol{Z}_j + \frac{1}{\sigma^2}\boldsymbol{X}_j^T\boldsymbol{y}_j] \\
\boldsymbol{\gamma}_j &= (\boldsymbol{\Lambda}^{-1} + \frac{1}{\sigma^2}\boldsymbol{X}_j^T\boldsymbol{X}_j)^{-1}
\end{aligned}
$$

We can then find the expected complete data loglikelihood conditional on the observed data $(\boldsymbol{X}_j, \boldsymbol{Z}_j, \boldsymbol{y}_j)_{j=1}^n$ and the estimated parameter at the sth step of the algorithm, $\boldsymbol{\phi}^{(s)}$.

$$
E(l|(\boldsymbol{y})_{j=1}^n, \boldsymbol{\phi}^{(s)}) \propto -2n \log |\boldsymbol{\Lambda}^{-1}|^{\frac{1}{2}} + \sum_{j=1}^n [tr(\boldsymbol{\Lambda}^{-1}\boldsymbol{\gamma}_j^{(s)}) + (\boldsymbol{\mu}_j^{(s)})^T \boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}_j^{(s)}]
$$

$$
-2\sum_{j=1}^n (\boldsymbol{\mu}_j^{(s)})^T \boldsymbol{\Lambda}^{-1}\boldsymbol{\Theta}\boldsymbol{Z}_j + \sum_{j=1}^n \boldsymbol{Z}_j^T\boldsymbol{\Theta}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{\Theta}\boldsymbol{Z}_j + (\sum_{j=1}^n m_j)\log\sigma^2 + \frac{1}{\sigma^2}\sum_{j=1}^n \boldsymbol{y}_j^T\boldsymbol{y}_j - \frac{2}{\sigma^2}\sum_{j=1}^n \boldsymbol{y}_j^T\boldsymbol{X}_j\boldsymbol{\mu}_j^{(s)}
$$

$$
+\frac{1}{\sigma^2}\sum_{j=1}^n [tr(\boldsymbol{X}_j^T\boldsymbol{X}_j\boldsymbol{\gamma}_j^{(s)}) + (\boldsymbol{\mu}_j^{(s)})^T\boldsymbol{X}_j^T\boldsymbol{X}_j\boldsymbol{\mu}_j^{(s)}]
$$

The M-step equations can be found by differentiating the expected loglikelihood with respect to the parameters $\boldsymbol{\phi}$. [4] This leads to the following M-step equations.

$$
\sigma^{2(s+1)} = \frac{1}{\sum_{j=1}^n m_j}\sum_{j=1}^n [tr(\boldsymbol{X}_j^T\boldsymbol{X}_j\boldsymbol{\gamma}_j^{(s)}) + (\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\mu}_j^{(s)})^T(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\mu}_j^{(s)})] \tag{13}
$$

[4] The differentiation is straightforward using the formulas in the appendix A.9 of Mardia et al. (1979).

$$\mathbf{\Theta}_i^{(s+1)} = \frac{1}{n_i}\sum_{j=1}^{n} z_{ji}\boldsymbol{\mu}_j^{(s)} \qquad i = 1, ..., q \tag{14}$$

$$\mathbf{\Lambda}^{(s+1)} = \frac{1}{n}\sum_{j=1}^{n}[\boldsymbol{\gamma}_j^{(s)} + (\boldsymbol{\mu}_j^{(s)} - \mathbf{\Theta}^{(s+1)}\boldsymbol{Z}_j)(\boldsymbol{\mu}_j^{(s)} - \mathbf{\Theta}^{(s+1)}\boldsymbol{Z}_j)^T] \tag{15}$$

In Laird and Ware (1982) the structure of the EM algorithm is schematically outlined for time-dependent between-individual covariates, regarding $(\boldsymbol{\beta}_j, \boldsymbol{\varepsilon}_j)_{j=1}^n$ as unobserved instead of just the $(\boldsymbol{\beta}_j)_{j=1}^n$ as in our approach. Their approach has also been tried here, but has led to singularity (or near singularity) problems with respect to $(\boldsymbol{X}_j\mathbf{\Lambda}\boldsymbol{X}_j^T)$. Numerically this may be overcome using generalized inverses.

In Laird and Ware (1982) concerns were raised about the slow convergence of the EM algorithm, its sensitivity to starting values and convergence to local maxima. The slow convergence is found to be a substantial problem in the applications in this report.

A.3 Testing Hypotheses

Testing Linear Hypotheses

We will first look into testing the multivariate general linear hypothesis $H_0 : \boldsymbol{C}\mathbf{\Theta}\boldsymbol{A} = \mathbf{0}$ where $\boldsymbol{C}$ is a $[\nu_1 \times q]$ matrix of rank $\nu_1 \leq q$ and $\boldsymbol{A}$ is a $[p \times \nu_2]$ matrix of rank $\nu_2 \leq p$, see Vonesh and Carter (1987).

The hypothesis can be written as $H_0^* : (\boldsymbol{A}^T \otimes \boldsymbol{C})\boldsymbol{\vartheta} = \mathbf{0}$ where $\otimes$ denotes the tensor (direct) product and $\boldsymbol{\vartheta}$ is the $[pq \times 1]$ vector obtained by arranging the row vectors of $\mathbf{\Theta}$ below one another, see Vonesh and Carter (1987). Let $\tilde{\boldsymbol{\vartheta}}$ be the estimated generalized least squares estimate of $\boldsymbol{\vartheta}$ defined from (11) and $\tilde{\mathbf{\Omega}}$ the $[pq \times pq]$ estimated covariance matrix of $\tilde{\boldsymbol{\vartheta}}$ constructed by arranging the elements of $\widehat{Cov}(\tilde{\boldsymbol{\theta}}_i)$ of equation (12) for $i = 1, ..., q$ appropriately.

$$T^2(\hat{\boldsymbol{\phi}}) = [(\boldsymbol{A}^T \otimes \boldsymbol{C})\tilde{\boldsymbol{\vartheta}}]^T[(\boldsymbol{A}^T \otimes \boldsymbol{C})\hat{\mathbf{\Omega}}(\boldsymbol{A}^T \otimes \boldsymbol{C})^T]^{-1}[(\boldsymbol{A}^T \otimes \boldsymbol{C})\tilde{\boldsymbol{\vartheta}}]$$

Under H_0^* the statistic $T^2(\hat{\boldsymbol{\phi}})$ will have a approximate χ^2-distribution with $\nu_1\nu_2$ degrees of freedom when the sample size, n, is large.

It might also be of interest to test whether the coefficients $\boldsymbol{\beta}_j$ are random (as assumed) or fixed. The coefficients are fixed and not random if $\mathbf{\Lambda} = \mathbf{0}$. This problem is not addressed here, but is mentioned in Spjøtvoll (1977).

Likelihood Ratio Tests

More general hypotheses can be tested using likelihood ratio tests. Let $\mathbf{\Phi}$ denote the entire parameter space. Consider testing the general hypothesis $H_0 : \boldsymbol{\phi} \in \mathbf{\Phi}_0$ versus $H_1 : \boldsymbol{\phi} \in \mathbf{\Phi}_0^C$. This can be done using a likelihood ratio test, see for example Casella and Berger (1990). Let $\boldsymbol{y} = (\boldsymbol{Y}_j)_{j=1}^n$ denote a random sample from the distribution $f(\boldsymbol{y}|\boldsymbol{\phi})$.

$$\lambda(\boldsymbol{y}) = \frac{\sup_{\mathbf{\Phi}_0} L(\boldsymbol{\phi}|\boldsymbol{y})}{\sup_{\mathbf{\Phi}} L(\boldsymbol{\phi}|\boldsymbol{y})} = \frac{L(\hat{\boldsymbol{\phi}}_0|\boldsymbol{y})}{L(\hat{\boldsymbol{\phi}}|\boldsymbol{y})}$$

Here $\hat{\boldsymbol{\phi}}_0$ is the maximum likelihood estimator of $\boldsymbol{\phi}$ obtained by doing a restricted maximization in the parameter space $\boldsymbol{\Phi}_0$ and $\hat{\boldsymbol{\phi}}$ by doing a unrestricted maximization.

Under some regularity conditions, if H_0 holds then the distribution of the statistic $-2\log\lambda(\boldsymbol{y})$ converges to a χ^2-distribution as the sample size increases, $n \to \infty$. The number of degrees of freedom of this limiting distribution is the difference between the number of free parameters specified by $\boldsymbol{\phi} \in \boldsymbol{\Phi}_0$ and the number of free parameters specified by $\boldsymbol{\phi} \in \boldsymbol{\Phi}$.

A.4 Comments on the two Estimation Procedures

In the random coefficient regression model each individual is analysed using a linear regression model and the individual estimates are then analysed using multivariate techniques.

In this presentation the "between-individual" one-time explanatory variables are restricted to be group indicators for two main reasons. The application of the random coefficient regression model to the variational and growth variables of the SGA data is used as a preprocessing of the data, and therefore only requires group indicators for SGA and site as one-time explanatory variables. For the generalized least squares estimation with a general matrix of one-time covariates, it is required to compute the $n \times n$ matrix $\boldsymbol{Z}^T\boldsymbol{Z}$, where $\boldsymbol{Z}^T = [\boldsymbol{Z}_1^T, \boldsymbol{Z}_2^T, ..., \boldsymbol{Z}_n^T]$. In this application n equals 1020 and a 1020×1020 matrix causes computational difficulties with respect to storage and speed, which is avoided by using group indicators.

Analyses are conducted for the variational and growth variables of the SGA data set of 1020 women using the random coefficient regression models in section 2.2 and 2.3. For these applications each step of the EM algorithm takes about the same time as the generalized least squares estimation. In addition the convergence of the EM algorithm is found to be very slow. For the ultrasound femur length (FEM) random coefficient regression the EM was started from the generalized least squares random coefficient regression estimates. The EM algorithm needed approximately 50 iterations to converge. The convergence criterion used was that the average change in the estimated parameters from one iteration to the next was less than 0.005. The estimates based on the generalized least squares estimation and the maximum likelihood estimation using the EM algorithm are given in table 10 together with values for the loglikelihood. There is a difference in the estimates of the parameters in $\boldsymbol{\Lambda}$, but the other estimates do not differ much for the two estimation procedures. The ratio for the loglikelihoods for the two procedures are 1.000057. In the study of the variational and growth variables of the SGA data set in section 2.2 and 2.3 the generalized least squares estimation procedure is used.

B Some Details on the Analyses

B.1 Random Coefficient Regression Models

Plots of individual regression coefficient estimates for the variational and growth curves are given in figures 5,6 and 7 for data from Trondheim, Bergen and Uppsala, respectively.

Parameter Estimates for Ultrasound Femur Length		
Parameters	Generalized LS	Maximum Likelihood via EM
$\boldsymbol{\theta}_{SGA,T}$	(-249.326, 57.314)	(-249.333, 57.315)
$\boldsymbol{\theta}_{SGA,B}$	(-262.957, 60.283)	(-262.970, 60.285)
$\boldsymbol{\theta}_{SGA,U}$	(-264.253, 59.858)	(-264.308, 59.869)
$\boldsymbol{\theta}_{nonSGA,T}$	(-272.725, 61.778)	(-272.776, 61.787)
$\boldsymbol{\theta}_{nonSGA,B}$	(-267.709, 60.688)	(-267.727, 60.691)
$\boldsymbol{\theta}_{nonSGA,U}$	(-272.734, 61.933)	(-272.780, 61.941)
σ^2	2.352474	2.322854
$\Lambda_{intercept}$	147.5792	161.5316
$\Lambda_{intercept,slope}$	-28.3119	-30.8911
Λ_{slope}	5.5410	6.0174
scaled loglikelihood	-8682.659	-8682.161
ditto ratio	1.0	1.000057

Table 10: Parameters estimates for the ultrasound femur length using all available data for the total TBU-SGA data set. The femur length is modelled to be linear in the logarithm of the gestational age in days. The random coefficient regression parameter estimates in the second column are from using the generalized least squares estimation and the third column from using maximum likelihood estimates using the EM algorithm, see A.2. The location parameters are denoted $\boldsymbol{\theta}$ with a subindex for SGA or non-SGA births and for site (T=Trondheim, B=Bergen and U=Uppsala). The variance is denoted σ^2 and the covariance matrix of the location parameters $\boldsymbol{\Lambda}$. The ratio of the loglikelihoods inserted the two sets of parameter estimates show a that the difference in the sets of parameter estimates is small.

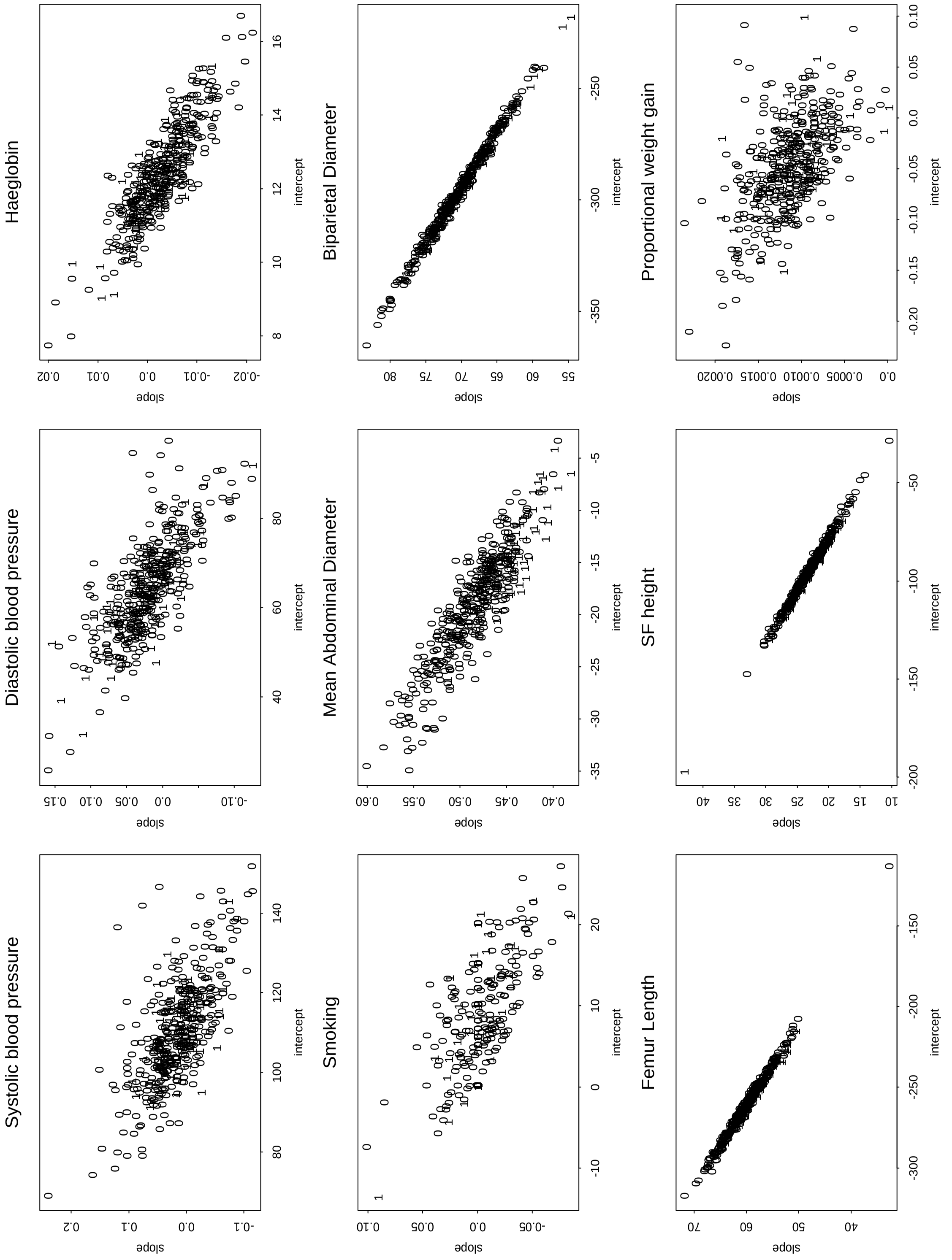


Figure 5: Individual regression coefficient estimates (least squares) for the variational and growth curves in Trondheim. The numbers “0” denote non-SGA births and the numbers “1” denote SGA births.

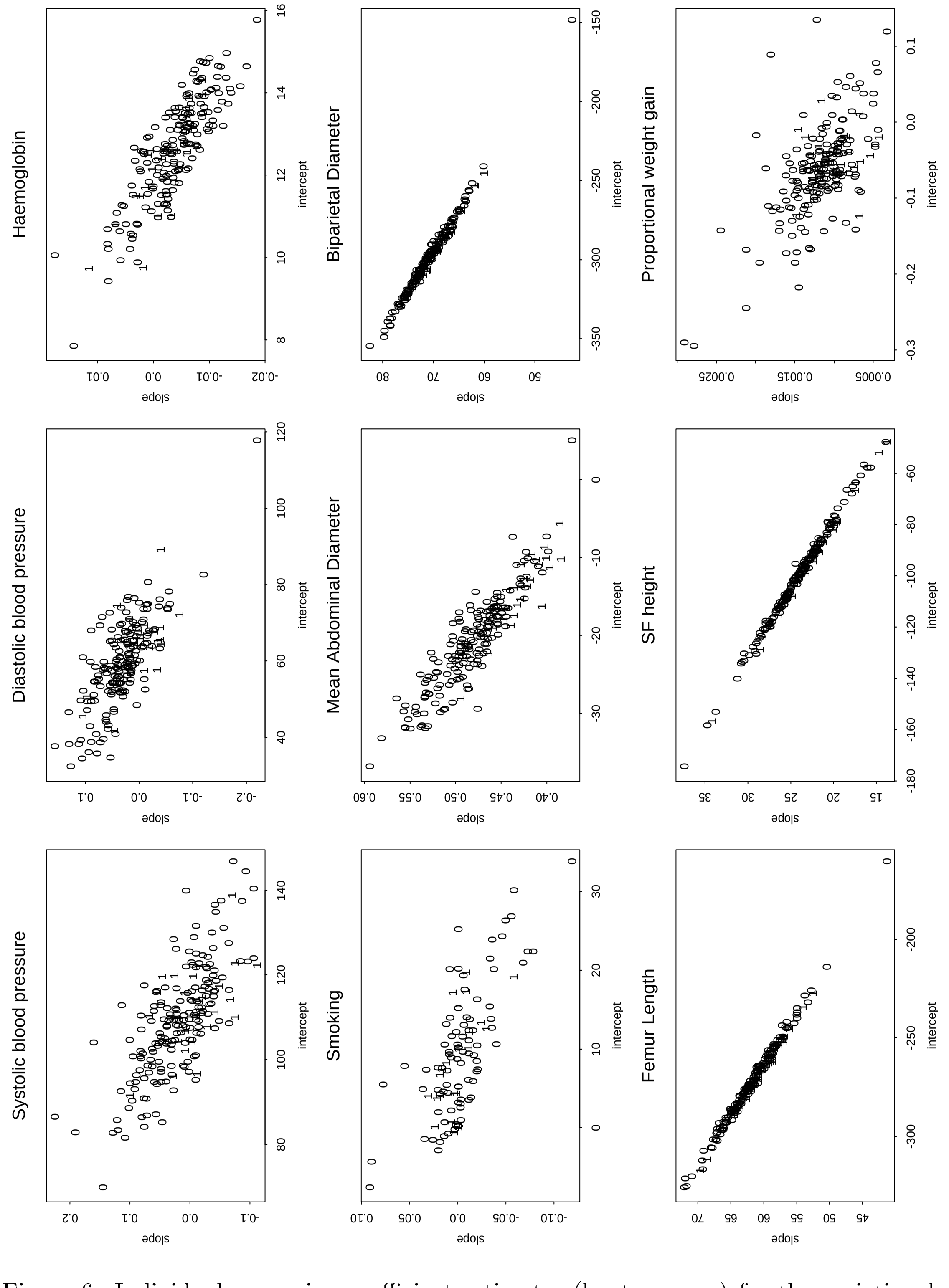


Figure 6: Individual regression coefficient estimates (least squares) for the variational and growth curves in Bergen. The numbers "0" denote non-SGA births and the numbers "1" denote SGA births.

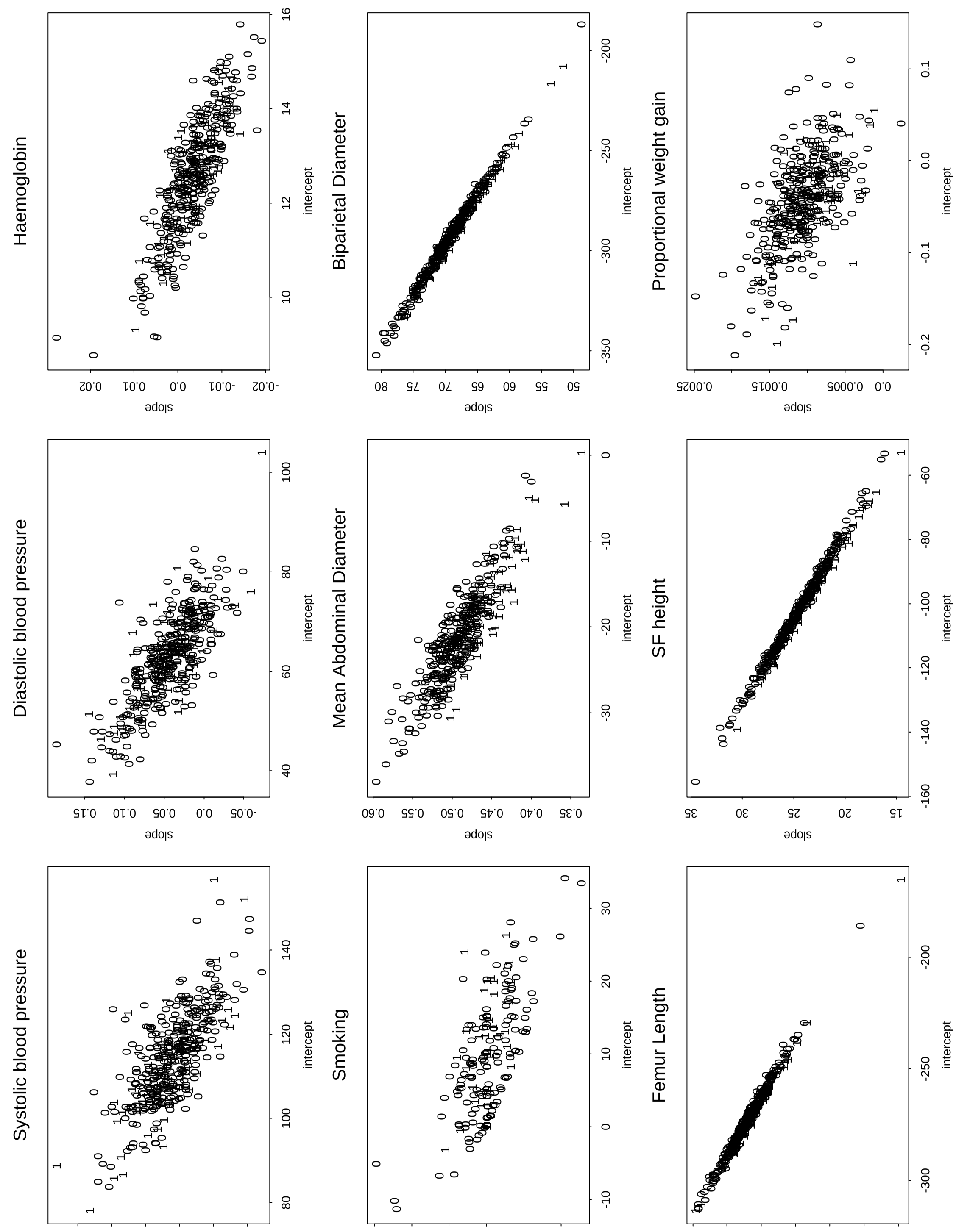


Figure 7: Individual regression coefficient estimates (least squares) for the variational and growth curves in Uppsala. The numbers “0” denote non-SGA births and the numbers “1” denote SGA births.

B.2 Discrimination and Classification

We here provide details on performing the analyses in section 3. We will look into model choices and the use of observational weights during training.

All analyses are done using the S language. The S-Plus libraries used here are available from the statistics archives at Carnegie-Mellon University (`statlib@lib.stat.cmu.edu`).

Altering the Priors of the Training Set Using Observational Weights Altering the priors of the training set is a way to incorporate variable costs of misclassification for the classes. This can be done by weighting the observations in the training set or replicating some of the observations in the training set.

We have set the cut-off point for classification to the class of SGA births at $p(\omega_{SGA}|\boldsymbol{x}) \geq \frac{1}{A+1}$, with $A = 5$ based on regarding the misclassification costs for SGA and non-SGA births. In addition we have true prior probabilities for the classes of $p(\omega_{SGA}) = 0.15$ and $p(\omega_{nonSGA}) = 0.85$.

$$p(\omega_{SGA}|\boldsymbol{x}) = \frac{p(\omega_{SGA})f(\boldsymbol{x}|\omega_{SGA})}{p(\omega_{SGA})f(\boldsymbol{x}|\omega_{SGA}) + p(\omega_{nonSGA})f(\boldsymbol{x}|\omega_{nonSGA})} =$$

$$\frac{1}{1 + \frac{1-p(\omega_{SGA})}{p(\omega_{SGA})}\frac{f(\boldsymbol{x}|\omega_{nonSGA})}{f(\boldsymbol{x}|\omega_{SGA})}} \geq \frac{1}{A+1}$$

Here $f(\boldsymbol{x}|\omega_{SGA})$ and $f(\boldsymbol{x}|\omega_{nonSGA})$ denote the class-conditional densities of the class of SGA and non-SGA births. Assume that we now alter the priors of the training set to $q(\omega_{SGA})$ and $q(\omega_{nonSGA})$. What should then the cut-off point for classification to the class of SGA births be in this altered setting? Let us denote the posterior probability for the class of SGA birth in this altered setting by $q(\omega_{SGA}|\boldsymbol{x})$.

$$q(\omega_{SGA}|\boldsymbol{x}) = \frac{q(\omega_{SGA})f(\boldsymbol{x}|\omega_{SGA})}{q(\omega_{SGA})f(\boldsymbol{x}|\omega_{SGA}) + q(\omega_{nonSGA})f(\boldsymbol{x}|\omega_{nonSGA})} = \frac{1}{1 + \frac{1-q(\omega_{SGA})}{q(\omega_{SGA})}\frac{f(\boldsymbol{x}|\omega_{nonSGA})}{f(\boldsymbol{x}|\omega_{SGA})}}$$

We can rewrite this to get an expression for the fraction of the two class-conditional distributions,

$$\frac{f(\boldsymbol{x}|\omega_{nonSGA})}{f(\boldsymbol{x}|\omega_{SGA})} = \frac{1 - q(\omega_{SGA}|\boldsymbol{x})}{\frac{1-q(\omega_{SGA})}{q(\omega_{SGA})}q(\omega_{SGA}|\boldsymbol{x})}$$

and then insert this into the expression for the cut-off point in the situation with the correct priors,

$$p(\omega_{SGA}|\boldsymbol{x}) = \frac{1}{1 + \frac{1-p(\omega_{SGA})}{p(\omega_{SGA})}\frac{1-q(\omega_{SGA}|\boldsymbol{x})}{\frac{1-q(\omega_{SGA})}{q(\omega_{SGA})}q(\omega_{SGA}|\boldsymbol{x})}} \geq \frac{1}{A+1}$$

solving for $q(\omega_{SGA}|\boldsymbol{x})$ we then find the cut-off point as,

$$q(\omega_{SGA}|\boldsymbol{x}) \geq \frac{1}{1 + A \cdot \frac{p(\omega_{SGA})}{1-p(\omega_{SGA})}\frac{1-q(\omega_{SGA})}{q(\omega_{SGA})}}$$

Replicating (twice, three times or more) all the SGA births will be the same as giving them higher weight. Weighting the SGA births by a factor of 2 leads to a cut-off point of 0.29. Weighting by a factor of 3 leads to a cut-off point of 0.38, by a factor of 4 to a cut-off point of 0.45 and by a factor of 5 to a cut-off point of 0.50. For some of the methods analysed here, the SGA births have been weighted by 3 or 5 during training, which we will refer to as *using observational weights during training.*

Details on Training and Model Choices For a given method the result presented in tables 5 to 9 of section 3 is for the model choice and training scheme giving the lowest average cost on the test set.

The linear and quadratic discriminant analysis are performed using the `classif` library of B.D. Ripley and require no model selection.

For the classification tree a default tree is fitted and pruned to optimal size with respect to minimum average cost on the test set. Observational weights during training have been tried. The minimum average cost on the test set was in all cases found using no weighting.

The flexible discriminant analysis is performed using the `fda` library of T. Hastie and R. Tibshirani using MARS with first and second degree interactions. Observational weights during training have been tried. The best results were obtained using no weighting.

For the projection pursuit regression one binary response denoting an SGA birth of not, is used. Different number of projections (1, 2 or 3) are tried. Observational weights during training have been used.

The multilayer perceptrons are trained with the `nnet` library of B.D. Ripley and have one hidden layer with sigmoid activation and skip-layer connections from the input to the output layer. Three combinations of output activation functions and loss functions are used; sigmoid output activation function and quadratic loss (sig), sigmoid output activation function and entropy loss (ent) and linear output activation function with quadratic loss (lin). Network architectures with one to eight hidden nodes are fitted. Each multilayer perceptron architecture is run 5 times, and the result with minimum of the chosen minimization criterion is selected. Observational weights during training have been tried. No weight-decay is used.

The method of mixture codebook classification (MC^2) is presented in Langaas (1995b and 1995c). Different model and method choices are investigated systematically; Gaussian and Student-t component densities, equal and class specific covariance matrices (full, diagonal and diagonal with equal variances). For each of these model choices using the joint likelihood optimization, the stepwise component architecture selection routine of Langaas (1995c) is used to find a good component architecture. A heuristic grid search is used on the basis of the above chosen component architecture for conditional likelihood optimization. For the conditional likelihood criterion only diagonal covariance matrices common to the two classes are tried. Moderate scale penalties (0.001) have been tried. Observational weights during training have been used for the conditional likelihood optimization criterion. The joint likelihood routines have used one to three starting points and the conditional likelihood routines have used five starting points. The use of common components has not been investigated in these analyses.

References

Bakketeig, L. S., Jacobsen, G., Hoffman, H. J., Lindmark, G., Bergsjø, P., Molne, K., and Rødsten, J. (1993), "Pre-pregnancy risk factors of small-for-gestational age births among parous women in Scandinavia.", *Acta Obstetricia et Gynecologica Scandinavica*, 72, 273–279.

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988), *The NEW S Language*, Pacific Grove, California: Wadsworth & Brooks/Cole.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984), *Classification and Regression Trees*, Montery: Wadsworth and Brooks/Cole.

Carter, R. L. and Yang, M. C. K. (1986), "Large sample inference in random coefficient regression models", *Communications in Statistics, Theory and Methods*, 15, 2507–2525.

Casella, G. and Berger, R. L. (1990), *Statistical Inference*, Pacific Grove, California: Wadsworth & Brooks/Cole.

Chambers, J. M. and Hastie, T. J., editors (1992), *Statistical Models in S*, Pacific Grove, California: Wadsworth & Brooks/Cole.

Clark, L. A. and Pregibon, D. (1992), *Tree-Based Models*, chapter 9, in Chambers and Hastie (1992).

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm (with discussion)", *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Friedman, J. H. and Stuetzle, W. (1981), "Projection pursuit regression", *Journal of the American Statistical Association*, 76, 817–823.

Hastie, T., Tibshirani, R., and Buja, A. (1994), "Flexible discriminant analysis by optimal scoring", *Journal of the American Statistical Association*, 89, 1255–1270.

Hertz, J., Krogh, A., and Palmer, R. G. (1991), *Introduction to the Theory of Neural Computation*, Redwood City, California: Addison-Wesley.

Jacobsen, G. (1992), "Detection of Intrauterine Growth Deviations, A Comparison Between Serial Symphysis Fundus Height and Ultrasound Measurements", Thesis for Doctor of Public Health, Yale University, Department of Epidemiology and Public Health and University of Trondheim, Department of Commumity Medicine and General Practice.

Jennrich, R. I. and Schluchter, M. D. (1986), "Unbalanced repeated-measures models with structured covariance matrices", *Biometrics*, 42, 805–820.

Johansen, S. (1984), *Functional relations, random coefficients, and nonlinear regression with application to kinetic data*, New York: Springer Verlag.

Johnson, R. A. and Wichern, D. W. (1982), *Applied Multivariate Statistical Analysis*, Englewood Cliffs, New Jersey: Prentice-Hall, 2nd edition.

Laird, N. M. and Ware, J. H. (1982), "Random-effects models for longitudinal data", *Biometrics*, 38, 963–974.

Langaas, M. (1995a), "Discrimination and classification", Technical Report Statistics 1/95, Department of Mathematical Sciences, The Norwegian Institute of Technology, The University of Trondheim.

Langaas, M. (1995b), "Mixture Codebook Classification, Part 1: Method outline", Technical Report Statistics 3/95, Department of Mathematical Sciences, The Norwegian Institute of Technology, The University of Trondheim.

Langaas, M. (1995c), "Mixture Codebook Classification, Part 2: Model selection and method evaluation", Technical Report Statistics 6/95, Department of Mathematical Sciences, The Norwegian Institute of Technology, The University of Trondheim.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, London: Academic Press.

Spjøtvoll, E. (1977), "Random coefficients regression models. A review.", *Mathematische Operationsforschung und Statistik, Series Statistics*, 8, 69–93.

Swamy, P. A. V. B. (1970), "Efficient inference in a random coefficient regression model.", *Econometrica*, 38, 311–323.

Venables, W. N. and Ripley, B. D. (1994), *Modern Applied Statistics with S-Plus*, New York: Springer-Verlag.

Vonesh, E. F. and Carter, R. L. (1987), "Efficient inference for random-coefficient growth curve models with unbalanced data", *Biometrics*, 43, 617–628.