1995

STATISTICAL ASPECTS OF FETAL SCREENING

DONOVAN, CHRISTINE M

STATISTICAL ASPECTS OF FETAL SCREENING

by

CHRISTINE M DONOVAN

A thesis submitted to the University of Plymouth
in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

School of Mathematics and Statistics
Faculty of Technology

In collaboration with
Johnson & Johnson
Clinical Diagnostics Ltd.

September 1995

ABSTRACT

STATISTICAL ASPECTS OF FETAL SCREENING

CHRISTINE M DONOVAN

This thesis discusses the current screening algorithm that is used to detect fetal Down's syndrome. The algorithm combines a model for predicting age related risks and a model for appropriately transformed serum concentrations to produce estimates of risks. A discriminant analysis is used to classify pregnancies as either unaffected or Down's syndrome.

The serum concentrations vary with gestational age and the relationship between serum concentrations and gestational age is modelled using regression. These models are discussed and alternative models for these relationships are offered. Concentration values are generally expressed in terms of multiples of the medians for unaffected pregnancies, or MoM values, which involves grouping the concentrations into weekly bins. Transformations of the MoM values are used in the model for predicting risks. The transformed values are equivalent to the residuals of the fitted regression models. This thesis directly models the residuals rather than converting the data to MoM values. This approach avoids the need to group gestational dates into completed weeks.

The performance of the algorithm is assessed in terms the detection rates and false positive rates. The performance rates are prone to considerable sampling error. Simulation methods are used to calculate standard errors for reported detection rates. The bias in the rates is also investigated using bootstrapping techniques.

The algorithm often fails to recognize abnormalities other than Down's syndrome and frequently associates them with low risks. A solution to the problem is offered that assigns an index of atypicality to each pregnancy, to identify those pregnancies that are atypical of unaffected pregnancies, but are also unlike Down's syndrome pregnancies.

Nonparametric techniques for estimating the class conditional densities of transformed serum values are used as an alternative to the conventional parametric techniques of estimation. High quality density estimates are illustrated and these are used to compute nonparametric likelihood ratios that can be used in the probability model to predict risks.

The effect of errors in the methods of recording gestational dates on the parameter estimates that are used in the discriminant analysis is also considered.

ACKNOWLEDGMENT

I express my gratitude to my supervisor, Dr. David Wright, for his advice and guidance and to Prof. Phil Dyke, Dr. Chris Davies and Dr. Charlie Martin for their help and support. I acknowledge Johnson & Johnson Clinical Diagnostics Ltd. for sponsoring the project and supplying data. I also acknowledge EPSRC and the University of Plymouth for funding the project. I thank Jenny for her help and continued friendship throughout. I thank my Mum and Dad for their valued help and the rest of my family for their support over the years. I also thank Michael for his words of encouragement. I express my dearest thanks to my son, Martin, for coping so well in my absence as a working mum.

AUTHOR'S DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

This study was financed with the aid of a studentship from EPSRC, and was carried out in collaboration with Johnson & Johnson Clinical Diagnostics Ltd.

Conferences and workshops attended:

Focus 92: Screening for fetal abnormalities, international conference, Blackpool, 1992.

Screening for Down's syndrome, international conference, Royal London Hospital of Obstetrics and Gynecology, 1993.

Bootstrapping and Data Analysis Workshop, Essex University Summer School, Colchester, 1994.

Published papers:

D.E. Wright, T.M. Reynolds and C.M. Donovan. *Assessment of Atypicality: an adjunct to prenatal screening for Down's syndrome that facilitates detection of other abnormalities.* Annals of Clinical Biochemistry. 1993 30: pp. 578-583.

D.E. Wright, C.M. Donovan and C. Davies. *Uncertainty in reported detection rates for Down's syndrome: Mathematical modelling techniques to obtain confidence intervals.* In Proceedings of the 4th conference on Endocrinology and Metabolism in Human Reproduction: Screening for Down's Syndrome, Eds. J. Grudzinskas, T. Chard, M. Chapman and H. Cuckle. Cambridge University Press 1993.

External Contacts:

Dr. Chris Davies and Dr. Charlie Martin,
Johnson & Johnson Clinical Diagnostics Ltd.
Cardiff

Mr. Kevin Spencer
Oldchurch Hospital
Romford
Essex

Signed.....

Date.....19/2/96.....

LIST OF FIGURES

LIST OF TABLES

To Martin

CONTENTS

Chapter 1

Introduction

1.1 Background

Antenatal screening for genetic disorders has become a topical issue within the medical literature over the last decade. Particular attention has been given to the chromosomal abnormality, Down's syndrome. With an incidence rate of 1:700 live births, Down's syndrome is the most frequently occurring genetic disorder during pregnancy (Van Lith (1994)). The disorder is the most common cause of severe mental retardation and it is characterized by short phalanges, a short and flat bridged nose and a flattening of the skull giving the appearance of a 'facial roundness'. There is clinical evidence of premature aging and pre-senile dementia, along with an increased risk of congenital heart disease (Martin (1978), Burger and Vogel (1973)).

Séguin (1846) originally described the disorder as a 'true multiple congenital anomalies/mental retardation syndrome'. John Langdon Down (1866) declared the disorder to be a reversion to the Mongoloid type caused by maternal tuberculosis. Following a laboratory error that consequently enabled clear resolution of all 46 chromosomes resident in each human cell, Dr Lejeune (1979) made clinical history by discovering the presence of a third chromosome, attached to the twenty first pair of chromosomes, in subjects with the abnormality.

Down's syndrome, (trisomy 21), and indeed other chromosomal disorders, can be detected reliably from the results of a trans-abdominal amniocentesis which is generally performed during the second trimester of pregnancy (weeks 15-21). Mid trimester

amniocentesis was first developed during the late nineteen sixties (Jacobson and Barter (1967)), and has since been credited as a highly sensitive procedure. However, there can be severe complications of process-related fetal loss through miscarriage. Such an invasive surgical procedure would severely compromise the health of the fetus if the procedure was performed before a gestation of sixteen weeks. Moreover, karyotyping of the amniotic fluid cells requires a further 2-3 weeks of culturing before diagnosis. Intrusions at such a late stage of pregnancy may impose psychological complications on the mother as well as interrupting fetal development.

Initial indications of successful first trimester genetic diagnosis were reported by Brambati and Simoni (1983). The publication discusses the various approaches to chorionic villus sampling. Their findings demonstrate that the most reliable method of sampling uncultured chorionic villi is by a transcervical insertion of a catheter coupled with ultrasonic guidance. Diagnostic karyotyping by this technique was considered successful enough at the time to move mid trimester cytogenetic diagnosis to the first trimester (weeks 0-14). Unlike amniocentesis, a result can be obtained within a few hours of sampling (Brambati and Simoni (1983)). Chorionic villus sampling permits an early diagnosis which has psychological and ethical advantages. However, more recent studies have reported the risk of process-induced miscarriage to be 1-2% (Lynch and Berkowitz (1992)).

According to Wald *et al* (1988) the best estimate of the fetal loss rate through amniocentesis was determined in the randomised trial reported by Ann Tabor *et al* (1986) indicating that the risk is about 1%. Amniocentesis is a well established procedure that imposes minimal risk on the fetus. This is probably responsible for the poor uptake of chorionic villus sampling in developed countries where the technological and economical necessities to cater for amniocentesis are widely available. The potential threat to fetal life along with economic factors restricts the relative utility of amniocentesis. The procedure is

only offered to women considered to be at high risk of delivering a Down's syndrome child.

Penrose (1934) ascertained the link between increasing incidences of Down's syndrome births and advancing maternal age, this association has since been well recognized. Until relatively recently, women have been classified as high risk on the basis of maternal age alone. Typically, women over 35 have been regarded as high risk and offered an amniocentesis. The age criteria has been adjusted for the changes in the mean maternal age of the pregnant population during the last two decades and in some regions the age cut-off has been increased to 37 (Van Lith (1994)). This procedure has led to approximately 5% of pregnant women being offered an amniocentesis.

The best estimate of a model for predicting risk from maternal age is given in Cuckle *et al* (1987). This model was fitted using the results derived from a combination of eight published surveys that monitored live births. The risk for each maternal age, in years, was specified. The estimated risk, c, is considered in terms of the probability of a Down's syndrome outcome for a given maternal age m, or is described as an odds ratio of the probability that the pregnancy is unaffected, $p(N/m)$, to the probability that it has Down's syndrome, $p(D/m)$, such that

$$\frac{p(N/m)}{p(D/m)} = c \Rightarrow odds = 1{:}c$$

A maternal age cut-off of 35-37 isolates less than 10% of the pregnant population, which amounts to 20-30% of all Down's syndrome pregnancies (Zeitune *et al* (1991)). Moreover, the uptake of amniocentesis is often as low as 50% reducing sensitivity to about 20% (Youings *et al* (1991)).

The problem of low detection for the system as a whole indicates the need for a more effective preliminary screen for identifying a high risk group. Although the mean maternal age of the pregnant population has elevated in recent years, the frequency of

Down's syndrome births has not increased since the introduction of genetic screening. However, the past twenty years of prenatal screening has had no real impact on the incidences of Down's syndrome. Current age related screening programs of this type are inefficient in terms of sensitivity. This emphasises the need for a broader screen that includes the whole population of pregnant women so that the number of unnecessary amniocenteses performed is reduced.

The possibility of an advanced screening procedure, introducing biochemical diagnostic testing methods for the detection of Down's syndrome, was first discussed by Merkatz (1984). The paper concerns the events surrounding a 28 year old woman who, not having been offered an amniocentesis, delivered a female infant with multiple congenital anomalies. The mother's levels of maternal serum concentrations of alpha-feto proteins, (AFP), drawn from blood sampling techniques, were available, as routine checks had been administered for neural tube defects. The analysis conducted by Merkatz (1984) shows that most autosomal trisomic births occur in women below 34 years. Evidence connecting fetal congenital abnormalities and low maternal serum concentrations of alpha-feto protein is also provided. Confirmation of this followed in a subsequent report by Cuckle *et al* (1987). In 1987 Bogart *et al* announced that concentrations of the placental protein human chorionic gonadotrophin, (HCG), could be used as a biochemical screening variable for Down's syndrome. Canick *et al* (1988) announced that low levels of maternal serum concentrations of unconjugated oestriol, (UE3), were associated with Down's syndrome pregnancies.

A multitude of similar studies have since been conducted from many regions with varying populations being studied. Most have shared the same objective, to examine possible connections between Down's syndrome pregnancies and maternal serum markers.

1.2 Current screening algorithms

Wald *et al* (1988) published a seminal article in the British Medical Journal which became a model for most subsequent work. Wald *et al* (1988) announced an achievable detection rate of 60% at an amniocentesis rate of 5%, by screening according to a composite risk derived from three analytes, AFP, UE3 and HCG in combination with maternal age.

Wald *et al* (1988) performed a case control study in which 77 singleton pregnancies associated with Down's syndrome were selected as cases and for each of these, 5 unaffected singleton pregnancies were chosen as controls, that were matched for serum sample duration, maternal age and gestational age. As with most of the published data of this type the serum concentrations were standardized to units known as multiples of medians or MoMs. The dependency between serum concentrations and gestational age is well recognized and unaffected median concentrations are often regressed against gestational age to establish smoothed weekly medians. Most workers use weighted least squares to fit loglinear or exponential models (Knight (1991)). Frequently, medians for completed weeks of gestation are simply specified or taken from previous studies. Each serum concentration is expressed as a multiple of the unaffected median for the same gestation to give a MoM value. For example, 2 MoM indicates the serum concentration is twice the unaffected median concentration. It is considered that MoM values have the advantage of removing the variation between the screening centres and also provide a means of standardizing measurements across the gestational age distribution. Other factors have been shown to influence analyte serum levels such as maternal weight and smoking habits. Some workers adjust median values for these explanatory variables (Reynolds *et al* (1991), Wald *et al*

(1991), Wald *et al* (1992)).

It is also assumed by Wald *et al* (1988) that the conditional probability density functions of appropriately transformed MoM values are multivariate Gaussian, with differing mean vectors and covariance matrices. In their paper, Wald *et al* (1988) highlight evidence of non-normality in the marginal distributions of transformed MoM analyte values. They therefore modify their algorithm so that it truncates excessively high or low values. Figures 1.1 to 1.3 illustrate the normal probability plots of the class conditional distributions of log transformed analyte values. Summaries of the data used to produce these figures are given in Chapter three.

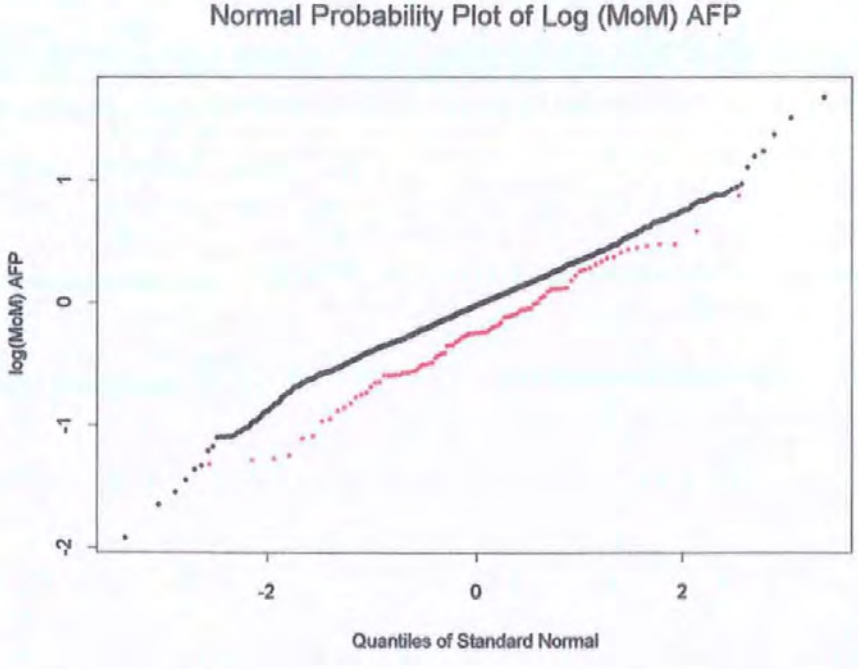

Figure 1.1: Normal probability plot of log(MoM) AFP values for unaffected pregnancies, (black) and Down's syndrome pregnancies, (red).

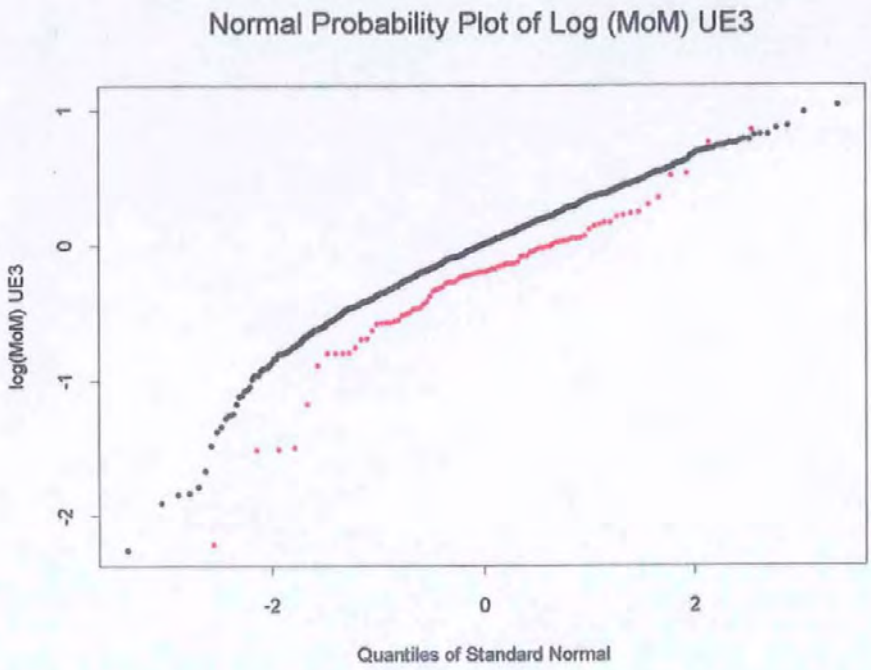

Figure 1.2: Normal probability plot of log(MoM) UE3 values for unaffected pregnancies, (black) and Down's syndrome pregnancies, (red).

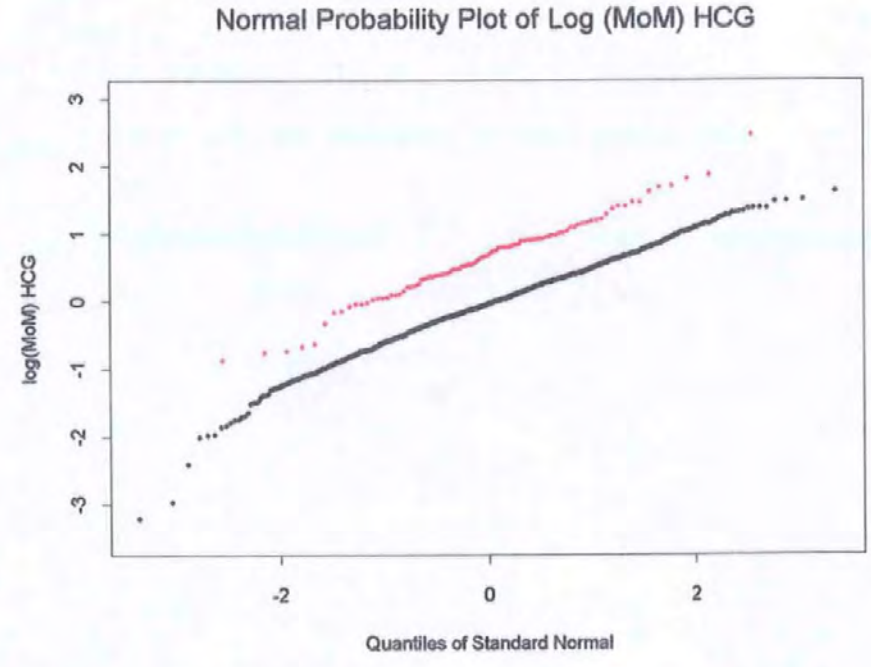

Figure 1.3: Normal probability plot of log(MoM) HCG values for unaffected pregnancies, (black) and Down's syndrome pregnancies, (red).

Having fitted these distributions to their data, Wald's algorithm then uses the ratio of the densities for unaffected and Down's syndrome distributions to modify the maternal age related odds, which produces a posterior odds ratio that is then used to identify a high risk group. Figures 1.4 to 1.6 plot the log(MoM) analyte values with the fitted Gaussian distributions. The plots demonstrate the degree of overlap between the marginal distributions for each outcome.

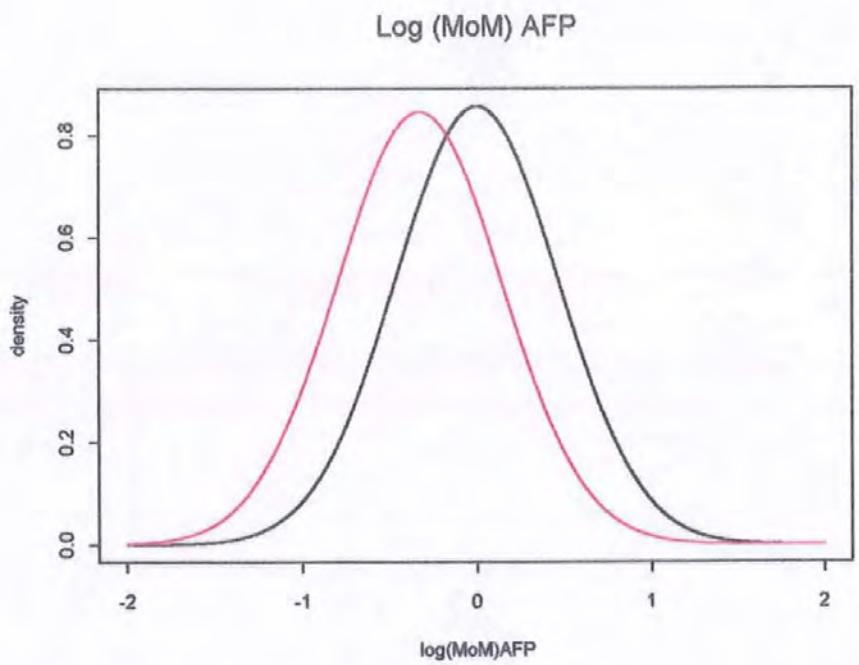

Figure 1.4: Fitted Gaussian distributions of log(MoM) AFP for unaffected pregnancies, (black) and Down's syndrome pregnancies, (red).

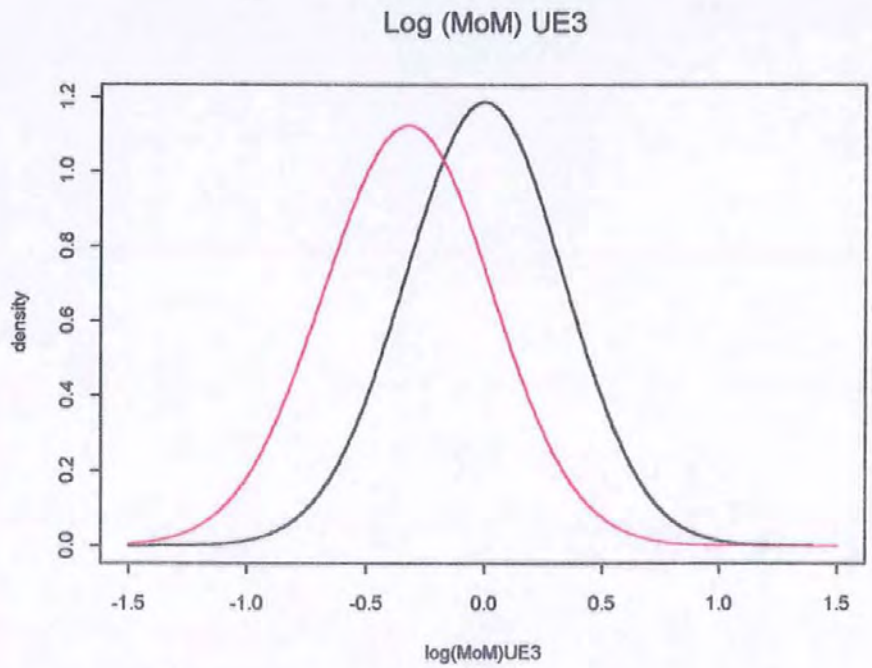

Figure 1.5: Fitted Gaussian distributions of log(MoM) UE3 for unaffected pregnancies, (black) and Down's syndrome pregnancies, (red).

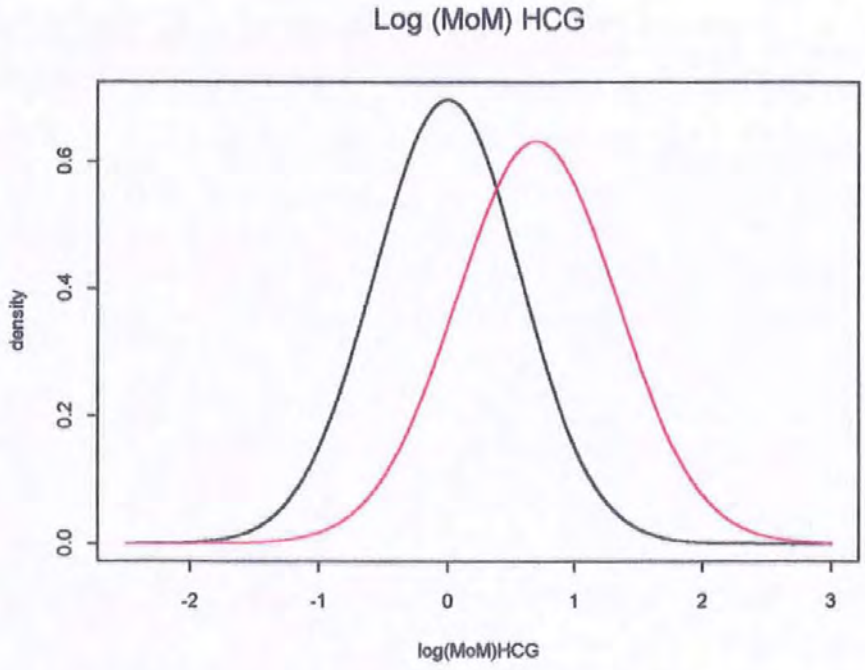

Figure 1.6: Fitted Gaussian distributions of log(MoM) HCG for unaffected pregnancies, (black) and Down's syndrome pregnancies, (red).

The posterior odds are used to assign each pregnancy with a risk of Down's syndrome. A risk cut-off level is selected and those pregnancies associated with a higher risk of Down's syndrome are screened positive. This particular group of women are offered an amniocentesis. A number of factors contribute to the selection of the risk cut-off. The rate of fetal loss through procedure-related miscarriage is necessarily considered. A high rate of amniocentesis would naturally lead to an unacceptable number of miscarriages and the procedure would be deemed as socially intolerable. Also, attention is paid to the detection rates and false positive rates, which are the respective proportions of pregnancies correctly and incorrectly diagnosed as having the abnormality by the screen. A risk cut-off is selected that maintains a desirable detection rate whilst considering the balance with the false positive rate. A false positive rate of 5% is generally classed as acceptable. The risk is, therefore, considered as a screening variable since the cut-off level depends on the sampling distribution of risks and is not fixed over different centres.

Most of the studies involving Down's syndrome screening that are presented in the medical literature follow this general model and they essentially utilize the same algorithm as that described by Wald *et al* (1988). Chapter two of this thesis sets out the statistical methodology of the screening algorithm given by Wald *et al* (1988), and discusses the assumptions it involves. The use of MoM values is reviewed and discussed. Chapter three offers an alternative approach to modelling, that avoids the need to convert concentrations to MoM values. A regression analysis is conducted and the forms of the fitted models are discussed.

The remainder of this chapter discusses the statistical theory that sets the foundations of the screening algorithms. The following sections discuss the application of discriminant analysis to screening, the methods of estimating performance measures, the meaning of risk and the current use of modelling techniques. Each item is dealt with

individually in subsequent chapters of this thesis.

1.3 Discriminant analysis

The problem of Down's syndrome screening is in essence a problem of discrimination and there is a wealth of relevant literature in this statistical field (Lachenbruch (1975), Aitchison and Dunsmore (1975), Hand (1981)). A short overview of discriminant analysis follows. The intention is simply to fix the basic notation and terminology for subsequent discussions rather than to provide a detailed summary. The notation adopted is in the style of McLachlan (1992).

Formally, consider a set of individuals divided into g mutually exclusive and exhaustive groups, or classes G_1, G_2 ,...,G_g, $g \geq 2$. For each set of subjects originating from a distinct group, G_i, $i = 1, 2, ..., g$, there exists a measurement space of recorded values from which a feature space of predictive variables can be selected. Values from the feature space provide each subject with a p-dimensional feature vector $\underline{x} = (x_1, x_2, ..., x_p)^T$ with a unique density $p(\underline{x} / G_i)$, $i = 1, 2, ..., g$.

Discriminant analysis is concerned with deriving an allocation rule, $r(\underline{x})$, from the feature vectors of subjects with known origins. Subjects of unknown origin can then be classified to one of g groups via the allocation rule. The allocation rule, or discriminant rule divides the measurement space into disjoint regions, R_i, $i = 1, 2, ..., g$, of postulated group separation, that are bounded by decision surfaces. The decision surfaces are derived from Bayes principle of minimizing error (Hand (1981)). Hence, $r(\underline{x}) = i$ implies the subject with feature vector $\underline{x}$ is allocated to group G_i since the rule classifies $\underline{x}$ in region

R_i. The decision theoretical approach minimizes error by allocating a subject to a group that maximizes the posterior probability of ownership. Thus the optimal allocation rule, $r_0(\underline{x})$, is such that

$$p(G_i/\underline{x}) > p(G_j/\underline{x}) \;\; \forall j \neq i,\; i,j = 1, 2,\ldots,g \text{ then } r_0(\underline{x}) = i$$

Equivalently, since the posterior probabilities are rarely known, $r_0(\underline{x})$ can be redefined using Bayes formula, to give

$$p(G_i/\underline{x}) = \frac{p(\underline{x}/G_i)p(G_i)}{p(\underline{x})}$$

yielding

$$p(\underline{x}/G_i)p(G_i) > p(\underline{x}/G_j)p(G_j) \quad \forall j \neq i,\; i,j = 1, 2,\ldots,g \text{ then } r_0(\underline{x}) = i \tag{1.1}$$

where $p(G_i)$ are the prior probabilities, or arrival rates of a random value belonging to each group in the classification space, and $p(\underline{x}/G_i)$ denote the class conditional densities, which are assumed to be known. With a space divided into two possible classes, as with the current screening algorithm, the minimum error decision rule reduces to a rule of the form

$$\text{if } \frac{p(\underline{x}/G_1)}{p(\underline{x}/G_2)} > \frac{p(G_2)}{p(G_1)} \text{ allocate to } G_1 \text{ and if } \frac{p(\underline{x}/G_1)}{p(\underline{x}/G_2)} < \frac{p(G_2)}{p(G_1)} \text{ allocate to } G_2 \tag{1.2}$$

By letting G_1 be the class of unaffected pregnancies, N, and G_2 be the Down's syndrome pregnancies, D, equation (1.2) can be rearranged to form the odds ratio used in the screening algorithm given by Wald *et al* (1988) to discriminate between the two groups

$$\frac{p(N/m)}{p(D/m)}\frac{p(\underline{x}/N)}{p(\underline{x}/D)} > c \text{ then allocate to } N \qquad (1.3)$$

where $p(N/m)$, $p(D/m)$ are the prior odds of each outcome specific to a maternal age m, $p(\underline{x}/N)$, $p(\underline{x}/D)$ are the class conditional distributions of MoM, or transformed MoM analyte values, $\underline{x}$, and c denotes the risk cut-off associated with Down's syndrome.

The rates of allocation of the decision rule are given below

$$e_{i,j}(r) = \Pr\{r(\underline{x}) = j/G_i\} \quad \forall \quad i, j = 1, 2, \ldots, g \qquad (1.4)$$

and these are evaluated through

$$e_{i,j}(r) = \int_{R_j} p(\underline{x}/G_i)\, d\underline{x} \quad \forall \quad i, j = 1, 2, \ldots, g \qquad (1.5)$$

The correct allocation rates for the biochemical screen for Down's syndrome, and similarly for other diagnostic tests that require an outright allocation to either a diseased or unaffected category, are defined as the sensitivity and specificity of the test. The sensitivity and specificity relevant to Down's syndrome screening are defined as

$$e_{D,D}(r) = 1 - e_{D,N}(r) \text{ and } e_{N,N}(r) = 1 - e_{N,D}(r)$$

respectively, where $e_{N,D}(r)$ is the false positive rate of the test, that is the proportion of unaffected pregnancies incorrectly screened positive and $e_{D,N}(r)$ is the false negative rate of the test, that is the proportion of Down's syndrome pregnancies screened negative by the test.

The decision rule of the form of equation (1.3) is necessarily based on

precise knowledge of the form of the class conditional densities. The feature vectors can be directly 'plugged in' to the formulas to give the rates of allocation. However, it is rare that the class conditional densities are known. One solution to this problem is to estimate the class conditional densities from a training set of data yielding a sample based discriminant rule that serves to estimate the optimal Bayes minimum error rule. From the sample based discriminant rule, estimates of the true error rates of allocation are computed. Such estimative procedures require extreme caution since the estimated allocation rates may be prone to bias from many sources. The possibility of bias in the estimated error rates is discussed in section 1.4.1 of this chapter.

1.4 Prenatal screening and discriminant analysis

With most current screening algorithms for Down's syndrome, feature vectors of transformed MoM AFP, UE3 and HCG analyte values are selected as predictive variables. Risk estimates are calculated and a discriminant rule is applied to the risks. The rule classifies those pregnancies with risks in excess of a selected cut-off level as Down's syndrome, and those pregnancies associated with lower risks as unaffected.

Parametric techniques are used to estimate the class conditional densities of transformed MoM values. The parameters are estimated from training data that is obtained from a retrospective study. Retrospective studies use samples whose classifications are already known. This gives rise to a sample based discriminant rule. The validity of the model is assessed through the detection rate and false positive rate of the algorithm.

1.4.1 Detection rates and their standard errors

In general the medical literature reports the performance statistics of the algorithm in the form of point estimates (Wald *et al* (1988), (1992)). Little or no attention is paid to sampling error or bias in the detection rates. It is well known that the discriminant rule is optimised for the design set so the estimated performance statistics are overrated. The parameter estimates will not be optimal for another random sample from the same distribution.

Unrepresentative maternal age distributions can create another source of bias. Since the risk of a Down's syndrome pregnancy increases with advancing maternal age, samples that over represent more mature women will raise the sensitivity of the screen.

Failure to consider the sampling error and bias in the estimated detection rates has led to unnecessary controversy over differences in the reported rates. Attention has focused on the benefits in screening with UE3 in addition to AFP and HCG (Crossley *et al* (1993)). Recent studies have indicated that free-β HCG is a more useful marker than Intact HCG (Spencer (1991)). A commercial interest has accelerated the race to achieve increased performance levels and combinations of markers have been patented in North America.

Chapter four of this thesis deals with the problem of sampling error and the possible bias in the estimated performance rates. Standard errors are calculated for published detection rates and for the detection rates associated with the parameter estimates under the models fitted in Chapter three. Also, parametric and nonparametric methods of calculating bias corrected error rates are discussed. A nonparametric method of bias correction is applied to the screening algorithm given by Wald *et al* (1988) to derive bias corrected detection rates and false positive rates.

1.4.2 The incorporation of a non-specific screen

The screening algorithm used by Wald *et al* (1988) is designed to classify pregnancies as either unaffected or Down's syndrome. However, the algorithm often fails to recognize other abnormalities. Heyl *et al* (1990) highlight cases in which other abnormalities, such as trisomy 18, are assigned low risks. He emphasised that the algorithm cannot be used legitimately to reassure a women that her pregnancy is 'normal'. Some abnormalities have analyte MoM values that are dissimilar to a Down's syndrome outcome but are atypical of an unaffected outcome. Frequently, these pregnancies are assigned a low risk of abnormality and are subsequently classified as unaffected even though they may be highly atypical of this outcome.

Some studies have made attempts to overcome the problem of low sensitivity in abnormalities other than Down's syndrome by incorporating other classifications into the screen. However, as Down's syndrome is the most common chromosomal fetal disorder, data involving other karyotypes are limited. It would therefore be impractical to screen for many fetal disorders simultaneously since any distributional assumptions would be unreliable due to small sample sizes. Some exceptions are other trisomy aneuploides, such as trisomy 13 and trisomy 18 whose occurrences are frequent enough to prompt researchers such as Heyl *et al* (1990) and Staples *et al* (1991) to screen for these disorders.

The question of atypical events in discriminant analysis has been addressed in many statistical publications (Aitchison and Dunsmore (1975)). Wright *et al* (1993) offers a simple approach based on the Mahanolobis distance to deal with the problem of low sensitivity when screening with other abnormalities. The problem can be reduced by incorporating a non-specific classification into the current screening algorithm. An index of

atypicality relative to all pregnancies that are classified as unaffected can be constructed using the Mahalanobis distance and these pregnancies associated with sufficiently large atypicality indices can be screened as non-specific. Chapter five of this thesis describes formally the methodology used to compute atypicality indices and illustrates the benefits of incorporating a non-specific classification into the existing algorithm.

1.4.3 Estimation of the class conditional densities

Wald *et al* (1988) adopt a parametric approach to the problem of estimating the class conditional densities of appropriately transformed MoM analyte values. Truncation limits are used to trim values that fall outside a linear range on a normal probability plot. This issue is addressed in Chapter two.

The problem of estimating the class conditional probability density functions of the form given in equation (1.3) is an equivalent problem to estimating the discriminant function. Hand (1981) points out that for a sample based discriminant rule to give a good approximation of the Bayes minimum error rule, the decision surfaces must be precisely defined to ensure the correct allocation of observations that fall within the tails of these distributions. Accurate estimation in the extremes of the class conditional densities is therefore of the utmost importance.

The lack of fit of a Gaussian form suggests the parametric approach to the problem is inadequate. In addition to this, the use of truncation limits coupled with the sparsity of data from affected pregnancies casts further doubt on the reliability of the estimation.

Several nonparametric methods of density estimation exist, such as kernel and nearest neighbour techniques. These methods are well documented in the literature (Hand (1981), Silverman (1986)). Implementing such robust procedures of density estimation into the

screening algorithm should add flexibility and serve to increase detection. Chapter six of this thesis discusses and illustrates the use of nonparametric density estimation in screening. A concise review of publications involving kernel methods of density estimation is also provided.

1.5 The effect of errors recorded in the gestational dating methods

It is now well established that analyte concentration levels vary with fetal age and concentrations are usually recorded during the second trimester of pregnancy. The methods of dating fetal growth rates vary. According to DiPietro and Allen (1991) the most reliable method uses an abdominal ultrasound scan, otherwise known as sonography, that dates pregnancies on the basis of standard fetal measurements. A commonly used method of dating, known as LMP dating, is based on the last menstrual period. The estimated date of delivery is expected to be 40 weeks from the first day of the last menstrual period. A less frequently used method of dating involves a clinical assessment of the uterus. Wald *et al* (1992a) discuss the performance of the screening algorithm when different dating methods are used. The paper concludes that obtaining gestational ages from sonography offers substantial advantages to screening.

It is, however, well known that each method of recording fetal age is subject to error. Linear regression models are currently used to model the analyte concentration values against gestational age. These models are based on the assumption that the explanatory variables are observed without error. No consideration is given to the errors in gestational dates that frequently occur. The errors in the recorded dates may affect the distributions of MoM values and also the calculation of risks.

Chapter seven of this thesis discusses the gestational dating methods set out by

DiPietro and Allen (1991). This Chapter investigates an alternative approach to modelling which replaces the regression models with functional models that assume a random error is present in the explanatory variable (Fuller (1980)). The functional models are used to illustrate the effects of errors in the recorded fetal dates on the parameter estimates for the linear models.

1.6 Conclusion

This thesis aims to highlight the main problems in the screening algorithm that is most commonly used in Britain to detect fetal Down's syndrome. An attempt is made to find solutions to these problems. Chapter eight of this thesis provides a summary of the research and discusses areas of possible future work. This section introduces these areas.

Chapter six of this thesis deals with the problem of estimating the class conditional densities for unaffected and affected pregnancies. The densities are estimated from the training data which usually comprises of thousands of unaffected controls but only tens of cases. Chapter six demonstrates how the large samples of controls can be used to produce high quality nonparametric estimates of the density for unaffected pregnancies. However the scarcity of data for affected pregnancies questions whether nonparametric methods of density estimation can be used to construct reliable estimates of the densities for affected pregnancies in higher dimensions.

Wright (1995) addresses the problem of estimating class conditional densities, for the purposes of discriminating between classes, when only relatively small samples are available from a class. The report investigates models in which the class conditional distributions are assumed to have a common distibutional form which is modelled using the nonparametric methods of density estimation described in Chapter six. Wright (1995) uses

parametric shifts in location and dispersion to model the differences between classes. The form of the model given in Wright (1995) is defined in Chapter eight and its application to screening is discussed.

Chapter 2

Current Methodology

2.1 Introduction

The aim of this chapter is to give an overview of the current methodology that is used to quantify the risk of Down's syndrome from a statistical perspective. Section 2.2 gives a brief historical review of screening methods. Section 2.3 describes the basic principles and assumptions involved in the model which is used to calculate risks. Section 2.4 presents a formal derivation of the risk algorithm that is based on the assumptions given in section 2.3. The algorithm is viewed as having two components, one is a model for predicting age related prior risks and the other is a model for the appropriately transformed serum concentrations which provides the likelihood ratio. The form of these models are described in sections 2.5 and 2.6 respectively. Section 2.7 discusses the use of MoMs in screening and questions their use as a standardized measure. Section 2.8 assesses the effects of truncation limits on the class conditional distributions of MoM values and on the calculation of risks in general. A brief summary of the chapter is given in section 2.9.

2.2 Historical background

Over the last twenty years, the relationship between an increased risk of Down's syndrome and maternal age has been well established and methods of selecting women for a diagnostic amniocentesis have been on the basis of maternal age. Typically, women aged 35 or above have been offered the test. It is now, however, well recognized that selective maternal age screening leads to poor performance levels and has little impact on the birth

incidence of Down's syndrome. Since the pregnant population consists mostly of women younger than 36 years, only a small proportion are offered the screen (Snijders (1993)). The detection rate, based on a 5% amniocentesis rate, using age as a criteria for screening is about 20%-30%, assuming a 100% uptake of amniocentesis (Wald *et al* (1988)).

Screening by maternal serum sampling was introduced in the mid 1980's. This brought about the opportunity to improve detection by using a screening program that combined the maternal age related risk with risks derived from maternal serum concentrations that were known predictors of Down's syndrome. The most widely used screen was introduced by Wald *et al* (1988). This combines the maternal age related risk with a risk derived from the analytes AFP, UE3, and HCG. Pregnancies are classified as screened positive if the risk is greater than a selected cut-off value and screened negative otherwise. The performance of the risk algorithm is monitored by the detection rate and false positive rate, which are described in Chapter one.

The screening test that is based on the three analytes has more commonly become known as 'the triple test'. The test provides a broader screen and it gives rise to a detection rate of about 65%. Although some regions of Britain still only use maternal age related risks to screen for Down's syndrome, most of the currently used screening algorithms are based on the methods of Wald *et al* (1988). However, developments in biochemical screening have led to differences in the combination of analytes and in the models used to calculate risks. Medical research publications debate the efficiency of various combinations of maternal serum markers as predictors of Down's syndrome. There is much controversy over the benefit in screening with different combinations of markers. Considerable attention is given to the use of UE3 with AFP and HCG. Different workers report different incremental benefits in detection from the addition of UE3 (Crossley *et al* (1993), Macri *et al* (1990)). The continuing argument over the reported performance statistics often ignores the

presence of sampling error and bias in the estimated detection rates. These issues are among those dealt with in Chapter four of this thesis. Variations in the choice of modelling techniques used to calculate estimates of risk also contribute to the argument over performance. Bishop (1994) discusses the findings reported by Ellis (1993) that highlight differences in the risk calculations between screening centres. Some centres also adjust the serum concentration levels for maternal weight, smoking habits, and ethnic origin (Reynolds *et al* 1992).

2.3 The basic model and its assumptions

Let N denote unaffected pregnancy outcomes and D denote Down's syndrome outcomes. m and g denote maternal age and gestational age respectively. X is the sample space of all possible feature vectors of appropriately transformed analyte concentrations $\underline{x}$.

The risk algorithm uses a posterior odds ratio of the form of equation (2.1)

$$\frac{1-[p(D/m)]}{p(D/m)} \frac{p(\underline{x}/g,N)}{p(\underline{x}/g,D)} \qquad (2.1)$$

where $p(D/m)$ is the maternal age related risk of Down's syndrome and both $p(\underline{x}/g,N)$ and $p(\underline{x}/g,D)$ are the class conditional distributions. A variety of approaches are described in the medical literature for estimating these parametric distributions. Leaving aside methods of estimation, this section considers the fundamental assumptions behind the form of equation (2.1) and the form of the class conditional distributions. The assumptions are given below.

Assumption (I)

It is assumed that each outcome of the procedure belongs to one of two types, unaffected or Down's syndrome.

(I) correctly states the mutually exclusiveness of the two outcomes, but also assumes exhaustiveness. Since abnormalities other than Down's syndrome may occur the classification scheme is not exhaustive. An important consequence of this is that low risks may be assigned to pregnancies with abnormalities other than Down's syndrome. As previously noted, this point has been discussed in the medical literature, see for example Heyl (1990), and a simple solution is offered in Wright *et al* (1993).

Assumption (II)

$$p(D/m,g) = p(D/m) \text{ and } p(N/m,g) = p(N/m)$$

(II) states formally that given maternal age pregnancy outcome is independent of gestational age. It has been established that Down's syndrome pregnancies abort more readily than unaffected pregnancies, particularly in early pregnancy (Snijders (1993)).

It is considered by some workers that spontaneous abortions are more frequent at later gestations with more mature women (Kratzer *et al* (1992)). Younger women may sustain abnormal pregnancies for longer durations. This issue is discussed in Snijders (1993).

Assumption (III)

$$p(\underline{x} / N, g, m) = p(\underline{x} / N, g) \text{ and } p(\underline{x} / D, g, m) = p(\underline{x} / D, g)$$

(III) states that given outcome and gestational age, the analyte concentration values are independent of maternal age. There is reported evidence of a slight negative correlation between UE3 values, when converted to MoMs, and maternal age (Davies *et al* (1991)). Although the correlation is statistically significant the dependency only explains 0.2% of the variation in transformed UE3 MoM values.

Assumption(IV)

$$p(\underline{x} / N, g) \sim N_3\left(\underline{\mu}(g), \Sigma_N\right) \text{ and } p(\underline{x} / D, g) \sim N_3\left(\underline{\mu}(g) + \underline{\Delta}, \Sigma_D\right)$$

The conditional distribution of appropriately transformed analyte values, given the outcome and gestational age are assumed to have a multivariate Gaussian distribution with unknown mean vector $\underline{\mu}(g)$ and covariance matrix Σ_N for an unaffected outcome and mean vector $\underline{\mu}(g) + \underline{\Delta}$ and covariance matrix Σ_D for a Down's syndrome outcome (IV). Essentially, the mean vector of appropriately transformed analyte concentrations for both outcomes are expressed as functions of gestational age. The mean vector for Down's syndrome pregnancies is assumed to differ from the unaffected mean vector by an additive shift, $\underline{\Delta}$. The magnitude of the shift remains constant over the range of gestation. (IV) also includes the assumption that the dispersion about the mean is constant over the range of gestation. Consequences of violating this assumption are discussed in Chapter three.

There is some debate over the differences in the Gaussian fit of the distributions of log(MoM)UE3 and MoM(UE3) (Wald *et al* (1993)). Both distributions demonstrate

marked deviations from a Gaussian form. Wald *et al* (1988) use untransformed UE3 MoMs. However, Wald *et al* (1992) conclude that a log Gaussian distribution fit the UE3 levels more efficiently. Crossley *et al* (1993) report better results with the untransformed data. Such arguments prompted Wald *et al* (1993) to reassess the issue using their original trials data, (Wald *et al* (1988)), and the trials data used in the 1992 publication. The results conclude that although the distributions of MoM UE3, with and without the logarithmic transformations show deviations from a Gaussian form for both unaffected and affected pregnancies, the transformed distributions provide a better fit. Neither of these papers mention the assumption of homogeneity of variance which is as critical as the assumption of normality. These issues are considered in Chapter three of this thesis.

The evidence of non-normality in the distributions of both MoM, and log(MoM) analyte values is more pronounced in the tails of the distributions. Truncation limits are applied to the transformed analyte values so that values that fall outside a linear range on a normal probability plot are trimmed to the nearest end limit. This is considered necessary in order to apply a Gaussian model which can be used in the calculation of risk. The application of truncation can be questioned over its inefficient use of data and its effect on the distributions, particularly in higher dimensions. There seems to be some confusion over the interpretation of the normal probability plots and the way they are used to determine truncating limits. This issue addressed in section 2.8.

2.4 Risk calculation and screening

With the (albeit false) assumption that only two categories, Down's syndrome and unaffected pregnancies are applicable, and by applying Bayes Theorem, the conditional probability of a Down's syndrome pregnancy, for each individual, given the maternal age, gestational age and maternal serum concentrations is given by

$$p(D / \underline{x}, m, g) = \frac{p(\underline{x} / D, m, g)p(D / m, g)}{p(x / D, m, g)p(D / m, g) + p(\underline{x} / N, m, g)p(N / m, g)} \tag{2.2}$$

By assumption (II), equation (2.2) can be written as

$$= \frac{p(\underline{x} / D, m, g)p(D / m)}{p(\underline{x} / D, m, g)p(D / m) + p(\underline{x} / N, m, g)p(N / m)} \tag{2.3}$$

and by assumption (III), equation (2.3) can be written as

$$= \frac{p(\underline{x} / D, g)p(D / m)}{p(\underline{x} / D, g)p(D / m) + p(\underline{x} / N, g)p(N / m)} \tag{2.4}$$

The discriminant rule forms the posterior odds ratio of an unaffected pregnancy to a Down's syndrome pregnancy and is given as a product of the likelihood ratio and the prior odds according to maternal age, equation (2.5).

$$\frac{p(N / \underline{x}, m, g)}{p(D / \underline{x}, m, g)} = \frac{p(\underline{x} / N, g)}{p(\underline{x} / D, g)} \frac{p(N / m)}{p(D / m)} \tag{2.5}$$

The algorithm produces a sample based distribution of odds or risks. A risk cut-off is determined such that 5% of the sample receive risks $\geq 1{:}c$. For any pregnancy, if

$$\frac{p(\underline{x}/N,g)}{p(\underline{x}/D,g)}\ \frac{p(N/m)}{p(D/m)} < c \qquad (2.6)$$

or equivalently $$LR(\underline{x}) \times \frac{P(N/m)}{P(D/m)} < c$$

where $LR(\underline{x})$ is the likelihood ratio, the pregnancy receives a positive result from the screen and is classified as screened positive. These women are counselled and offered an amniocentesis. In general, current screening policies state that no further action is necessary if a pregnancy is assigned a risk smaller than the risk cut-off. The shortcomings of this assumption are discussed in Chapter five of this thesis.

2.5 Models for predicting age related risks

The model given by Cuckle *et al* (1987) is now the most widely accepted model for predicting maternal age specific risks of Down's syndrome. Cuckle *et al* (1987) combined the results from eight published surveys monitoring live births. Random error associated with the combined estimated probabilities was reduced using a constant plus an exponential function of age model as described by Lamson and Hook (1981). The fitted model is given below.

$$p(D/m)=0.000627+\exp(-16.2395+0.286m) \qquad (2.7)$$

p(N/m) is found by calculating [1-p(D/m)], (Assumption(I)).

The model illustrates the increase in the prior probability of a Down's syndrome pregnancy with advancing maternal age. The model is shown in Figure 2.1.

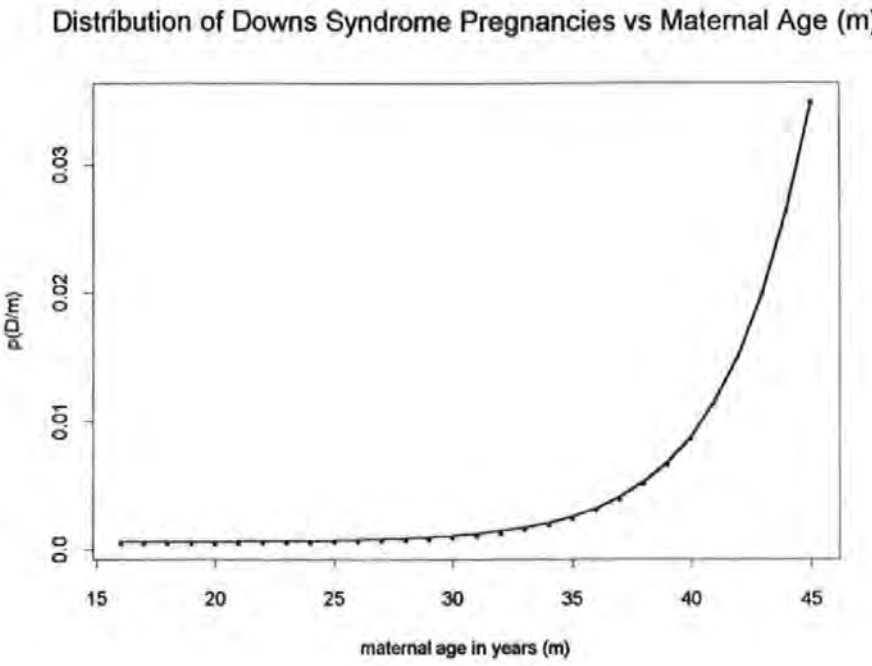

Figure 2.1: The estimated prior probability of a Down's syndrome birth at term as a function of maternal age (m). The fitted model given by Cuckle *et al* is $p(D/m) = 0.000627 + \exp(-16.2395 + 0.286m)$

Hecht and Hook (1994) recognize the extensive use of this model and therefore re-examine the data and analyses of Cuckle *et al* (1987). Hecht and Hook (1994) attach confidence intervals to the maternal age specific rates calculated from this model. They report that the rates must be viewed with relative uncertainty when used in conjunction with results from biochemical screening tests.

2.6 The likelihood

The risk of Down's syndrome by blood serum sampling is obtained from an application of Bayes Theorem. Maternal serum concentrations from pregnancies with known outcomes are used to model class conditional distributions of analyte values.

An estimate of $\underline{\mu}(g)$ is obtained by grouping the analyte concentrations into completed weeks of gestation. Median concentrations for unaffected pregnancies are then calculated for each week. Smoothed medians are established for unaffected pregnancies by applying median regression against completed weeks. As previously described, the concentration levels for each outcome are expressed in terms of MoM values. The MoM is a standard measure that is widely accepted in the medical literature. Concentration levels for both unaffected and affected pregnancies are expressed as some multiple of the unaffected median concentrations. A normalizing transformation is applied to the data which is usually of the logarithmic form to produce features that can be used in the discriminant analysis. Gaussian distributions are fitted to the appropriately transformed multiples of medians. The standard deviations for the class conditional distributions are estimated from the values that fall between a linear range on a normal probability plot. Wald *et al* (1988) recommend estimating between the $10^{th} - 90^{th}$ centiles of the plot. Wald *et al* (1993) admit that this is not a suitable range to estimate the standard deviation of log(MoM) UE3 for affected pregnancies and suggests that the range between the $25^{th} - 90^{th}$ centile is more appropriate. Revised estimates of these standard deviations are provided. $\underline{\mu}(g)$ and $\underline{\mu}(g) + \underline{\Delta}$, or some transformed versions, provide robust estimates of the means for the class conditional distributions. Truncation limits are applied to the transformed analyte values to remove excessive values. Gaussian densities are fitted to the truncated data and the quotient of the fitted densities for unaffected pregnancies and Down's syndrome pregnancies provides a likelihood ratio, $LR(\underline{x})$, which is then used to modify the age related risk.

The screening algorithms differ and the differences depend on the information used to compute MoM values. In Wald *et al* (1992) the transformed analyte concentrations are modelled separately for gestational age dates recorded by LMP dating methods and sonography. It is accepted that dating by sonography has less random error than dating by

LMP methods, therefore, the MoMs produced from gestational ages obtained by sonography show less dispersion. Wald *et al* (1992) also adjust the median analyte concentrations for maternal weight. This also reduces the dispersion about the mean concentrations which will inevitably effect the correlations and standard deviations of the concentrations. Thus, four algorithms are currently used. These comprise of LMP dating with and without weight corrections and sonography, also with and without weight corrections. The effects of recording concentrations using different dating methods on the calculated parameter estimates and performance statistics is discussed in Chapter seven.

2.7 The use of MoMs and median regression

As already described, analyte concentrations are conventionally expressed as multiples of the median value for unaffected pregnancies. It is considered that the use of MoMs provides a means of standardizing measures across the range of gestation whilst removing the effects of variation between centres. Thus data from different centres can be pooled. This approach has practical appeal since data for affected pregnancies are scarce. However the use of MoMs as an efficient standardized measure has been questioned (Bishop (1994)).

Recent developments in screening have led to differences in the models used to calculate risk between screening centres. Some centres adjust normal median concentrations for attributes such as smoking, maternal weight, ethnicity and gravidity (Reynolds *et al* (1991), Wald *et al* (1992)). Bishop reports the findings of Parvin *et al* (1991) that show that such methodological differences between screening centres result in a change in the distributional properties of MoM analyte values.

It is generally believed that the MoM is a standardized measure which removes variation between centres. Often, reported threshold values of MoMs are used as standard

indicators of Down's syndrome in studies conducted by different centres. Previously published parameter estimates are used as reference values for other independent trials. Such estimates may not perform efficiently as part of the calculation of risk if long term changes cause population shifts in the distributions of analytes or if the data is widely spread. Also differences in the calculations of MoM values will invariably lead to differences in the distributional features. Bishop (1994) discusses the findings of Parvin *et al* (1991) that conclude that likelihood ratio based risk calculations will be significantly affected unless centre specific reference distributions of MoM values are computed. Bishop (1994) investigates the statistical properties of MoM AFP values and discusses the consequences of pooling AFP data from different centres. Bishop (1994) notes that the analysis is generalized and can be applied to any data presented as MoM values. The study shows that standardized threshold MoM values relate to different percentiles of the gestational age dependent distributions. This is because the distributions of MoM AFP have lognormal parameters that depend on gestational age and the methodology specific to each centre. Therefore, standarized threshold MoM values cannot be reliably used as reference values as these relate to percentiles that are gestational age dependent. Bishop (1994) further illustrates that the combined distributions of centre specific MoMs over the range of gestational age are, in fact, a mixture of Gaussian distributions with mixtures from different gestational ages. It is therefore unreasonable to pool data from different centres. Chapter three of this thesis offers an alternative approach to modelling the distributions of analyte concentrations that avoids the need to standardize measures to MoM values.

2.8 Truncation limits

It has become standard practice to apply boundary or truncation limits to the analyte values outside the range over which, after transformation, a Gaussian distribution is deemed valid. Normal probability plots are generally used by Wald *et al* (1988) and (1992) as a criteria for identifying the range over which a Gaussian model seems appropriate. The transformed MoM values that fall outside this range are replaced with the nearest end limit of that range.

In the original article by Wald *et al* (1988) the truncation limits corresponding to each analyte used in the screening algorithm were $0.4 \leq MoM(AFP) \leq 2.5$, $0.4 \leq MoM(UE3) \leq 1.4$ and $0.2 \leq MoM(HCG) \leq 5$. In their 1992 publication , Wald *et al* revised the boundary limits for AFP and UE3 to $0.3 \leq MoM(AFP) \leq 3.3$, and $0.4 \leq MoM(UE3) \leq 2.5$ respectively. Crossley *et al* (1993) prompted Wald *et al* to reconsider their use of UE3 in screening. As a result revised truncation limits for UE3 distributions were given in Wald *et al* (1993). The revised values are $0.5 \leq MoM(UE3) \leq 2$. One consequence of applying truncation limits is that all pregnancies with log(MoM) analyte values outside the boundary limits are assigned a risk associated with the log(MoM) value at the nearest end limit.

It is generally assumed that the distribution is normal over the range where the points follow a straight line and it is non-normal elsewhere. However, even a small number of outliers can cause the plot to depart from a straight line well beyond the position of the outliers. This effect is illustrated below. Figure 2.2 plots the histogram and fitted Gaussian density for a random sample of 1000 observations drawn from a standard normal distribution. Figure 2.3 shows the normal probability plot of these observations. The observations in excess of ± 2 are outside a linear range. Figure 2.4 plots the histogram and

the fitted Gaussian density of the original data with 20 outliers each taking the value of -3.5. The plot illustrates that the distribution can be considered as Gaussian over the range of values above -3. However, the normal probability plot for the contaminated data, shown in Figure 2.5, is affected much further into the distribution then -3. The observations that fall below -1 depart from a straight line. If truncation limits were applied on the basis of this normal probability plot the lower limit would be equal to -1 and a large proportion of the data would be unnecessarily truncated. Therefore, such plots should not be used as a basis for setting limits within which the distribution can be assumed Gaussian.

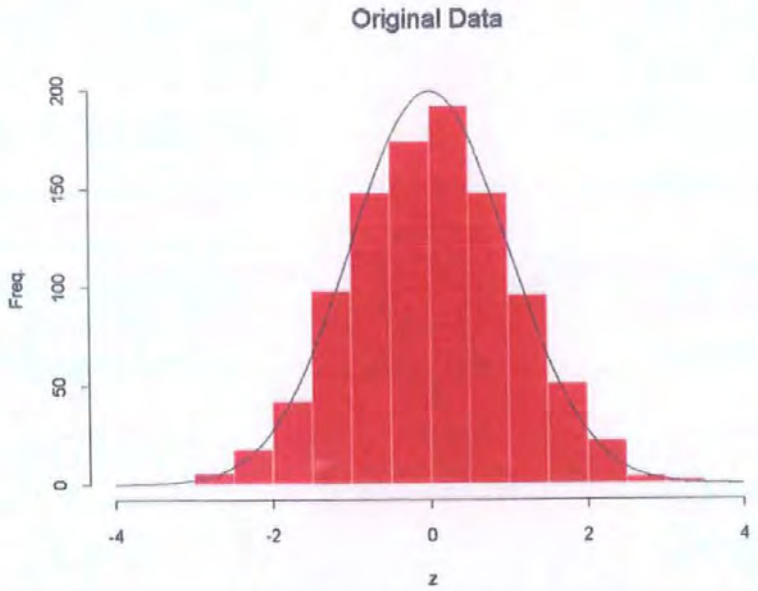

Figure 2.2: Histogram and fitted Gaussian density to a random sample of 1000 observations from a standard normal distribution.

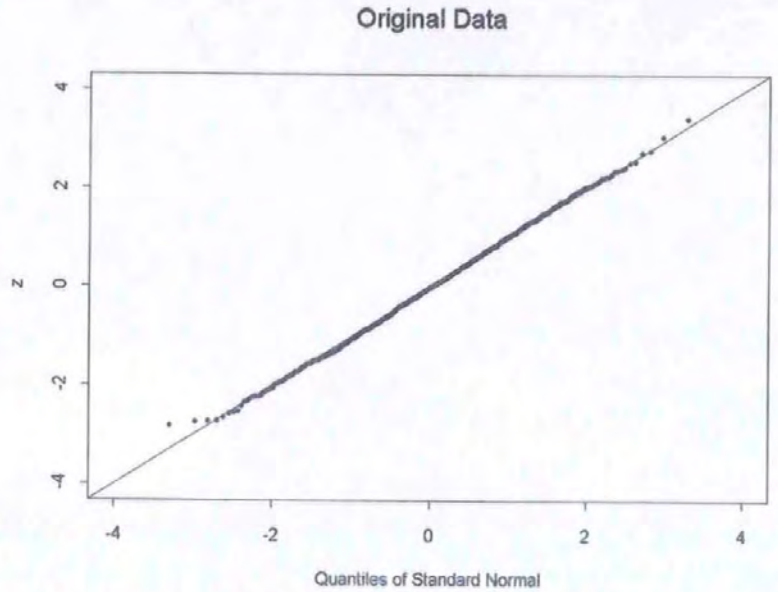

Figure 2.3: Normal probability plot of original data.

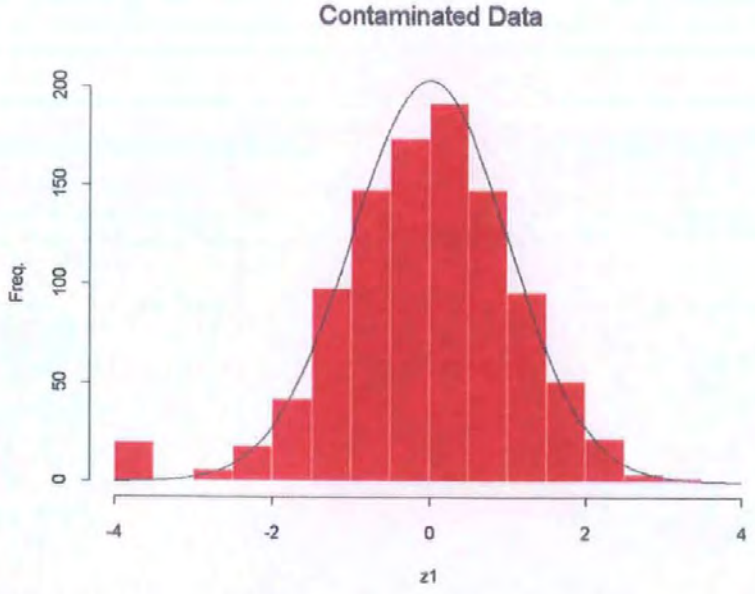

Figure 2.4: Histogram and fitted Gaussian density to a random sample of 1000 observations from a standard normal distribution plus 20 observations each taking the value -3.5.

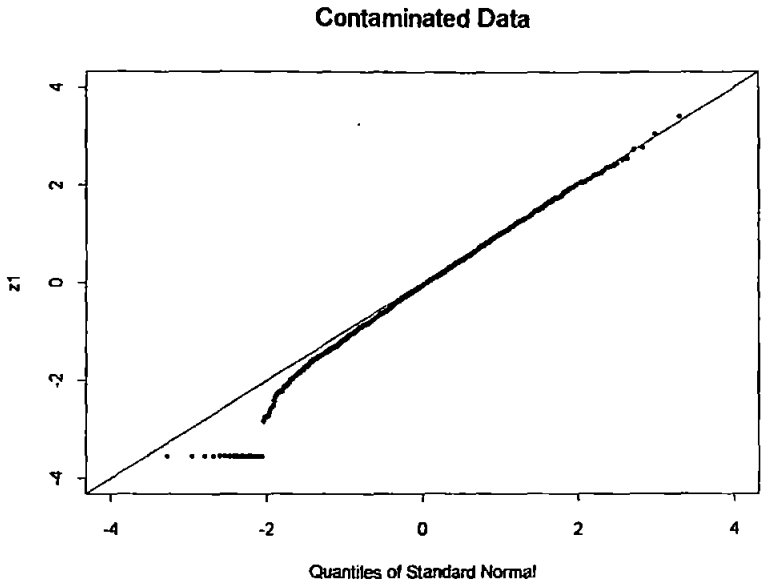

Figure 2.5: Normal probability plot of contaminated data.

2.9 Conclusion

This chapter has focused on the form of the probability models that are used to produce risk estimates of Down's syndrome. The assumptions involved in the probability models have also been discussed. The disadvantages of using the conventional method of modelling the MoM analyte values have been highlighted and an alternative method that directly models the residuals of the fitted regression equations have been proposed. This method of modelling is applied in Chapter six. The interpretation of the normal probability plots of the transformed MoM values and the use of truncation limits has been questioned.

Chapter 3

Analysis of clinical trials data

3.1 Introduction

This chapter presents an alternative to the conventional approach of using median regression to model analyte values. The approach uses standard linear and non-linear least squares for model fitting. Full summaries of the data used in the analysis are given in section 3.2. Section 3.3 describes the methods used to model the data and discusses models for location and variation. Appropriate methods to deal with outliers are also introduced in section 3.3. Sections 3.4, 3.5 and 3.6 report the findings of the analysis based on AFP, UE3 and HCG data respectively. Section 3.7 summarizes the results.

The parameter estimates of the fitted models quoted in this chapter are compared with those given in Wald *et al* (1988, 1992 and 1993) for LMP dating methods with no adjustment for maternal age. These are shown in Table 3.1. The date in brackets gives the most recent year of update.

PARAMETER	ANALYTE	UNAFFECTED	DOWN'S
MEANS	AFP	0.0000 (92)	-0.3286 (88)
	UE3	0.0000 (92)	-0.3249 (88)
	HCG	0.0000 (92)	0.6961 (88)
SD	AFP	0.4656 (92)	0.4720 (88)
	UE3	0.3362 (92)	0.3551 (93)
	HCG	0.5720 (92)	0.6309 (88)
R	AFP-UE3	0.2755 (92)	0.2708 (93)
	AFP-HCG	0.0723 (92)	0.1703 (88)
	HCG-UE3	-0.1752 (92)	-0.3204 (93)

Table 3.1: Means, standard deviations (SD), and correlation coefficients (R) in affected and unaffected pregnancies reported by Wald *et al* (1988), (1992) and (1993). All estimates are in logarithms to the base e. The method of recording gestational dates is LMP.

3.2 Data

The data used for analysis throughout this chapter and subsequent chapters are taken from the databases of six screening centres from the trial described by Davies *et al* (1991). The maternal serum concentrations AFP, UE3 and HCG are recorded. The centre indicators are shown below.

NAME	(1)	(2)	(3)	(4)	(5)	(6)
CENTRE	Amersham Trials	Bonn	Gottingen	Nottingham	Romford	Glasgow

Table 3.2: Summary of the screening centres from which data is obtained along with their indicator number.

Centres 2, 3, 5, and 6 have gestational ages recorded in days whilst centre 1 and 4 are in weeks. Summary statistics of the analyte concentrations are provided in Tables 3.3 to 3.5.

(I) AFP

	UNAFFECTED PREGNANCIES				DOWN'S SYNDROME PREGNANCIES			
CENTRE	FREQ.	MIN.	MAX.	MEDIAN	FREQ.	MIN.	MAX.	MEDIAN
1	60	13.95	98.75	33.05	N/A	N/A	N/A	N/A
2	259	13.2	232.5	40.51	16	13.2	62.45	32.08
3	279	9.01	86.7	31.81	9	10.5	74.67	36.03
4	146	17.53	87.13	37.53	27	10.58	65.06	24.32
5	381	5.76	88.77	41.29	26	10.07	93.53	34.62
6	174	12.78	96.85	39.32	15	15.89	62.34	30.48

Table 3.3: Summary statistics for AFP data recorded.

(II) UE3

	UNAFFECTED PREGNANCIES				DOWN'S SYNDROME PREGNANCIES			
CENTRE	FREQ.	MIN.	MAX.	MEDIAN	FREQ.	MIN.	MAX.	MEDIAN
1	60	1.51	7.7	4.035	N/A	N/A	N/A	N/A
2	259	0.69	9.67	5.36	16	1.61	8.5	4.165
3	278	1.16	9.82	3.59	9	1.33	8.69	3.75
4	146	2.08	9.6	4.51	27	0.5	10.88	4.11
5	399	0.49	15.08	5.01	26	0.98	7.84	3.85
6	191	0.75	11.71	4.79	15	2.51	7.34	3.61

Table 3.4: Summary statistics for UE3 data recorded.

(III) HCG

	UNAFFECTED PREGNANCIES				DOWN'S SYNDROME PREGNANCIES			
CENTRE	FREQ.	MIN.	MAX.	MEDIAN	FREQ.	MIN.	MAX.	MEDIAN
1	60	3.77	125.5	29.1	N/A	N/A	N/A	N/A
2	259	1.12	123.4	25.45	16	27.05	83.06	52.11
3	285	7.22	148.4	31	9	22.84	356.1	69.18
4	146	4.77	103.9	26.04	27	12.32	165.9	58
5	399	4.78	88.57	23.74	26	12.73	111.2	46.2
6	207	3.67	89.49	26.43	15	11.52	142.4	51.57

Table 3.5: Summary statistics for HCG data recorded.

3.3 Modelling

Wald *et al* (1988) apply weighted least squares to regress unaffected weekly median concentrations on gestational age, the weights being taken as the number of pregnancies at each week. The motivation for using this approach is that median values are more robust to outliers and abnormalities in the distributional shape. As described in Chapter one, concentration values are expressed as MoM values and it is generally accepted that the distribution of MoM analyte concentrations are adequately represented by a lognormal distribution. More specifically, the log(MoMs) are assumed to be normally distributed with a mean of zero and a constant variance. Multivariate Gaussian densities are fitted to the distributions of log(MoMs), truncation limits are applied and excessively high or low values are replaced by the nearest end limit.

However, the log(MoM) values are equivalent to the residuals, after a log transformation, of the fitted regression models. Therefore, it is sensible to fit multivariate Gaussian densities to the residuals of the regression models rather than converting the data to MoM values. This approach avoids the need to group gestational dates into weeks so the fitted models are more representative of true gestational dates rather than those used for statistical analysis. This method also allows other explanatory variables such as weight and smoking status to be included in a multiple regression model.

This method of modelling is applied to the data described in section 3.2. Residuals associated with Down's syndrome pregnancies can be determined using the regression coefficients estimated from the unaffected data. Gaussian densities can be fitted to the distributions of residuals from unaffected and affected pregnancies and the ratio of these distributions provides a likelihood function that can be used in a discriminant analysis to calculate risks.

3.3.1 Modelling location

Initial scatter diagrams of the raw data are plotted. A uniformly distributed random number over the range (-0.5, 0.5) is added to the gestational age to make the points distinguishable on the scatter diagrams. Since the added noise is small in comparison to the range of the data the overall shape of the distribution is unchanged, but each point appears as a unique dot. The location for unaffected pregnancies is modelled by fitting appropriate regression models with gestational age and centre as explanatory variables. The adequacy of the models are assessed graphically. Diagrams of the fitted models are illustrated. Also tables of parameter estimates are provided.

3.3.2 Detection of outliers

An effective method for detecting model deficiencies in the regression analysis is to examine the residuals. The residuals, e_i, should be independently distributed normal deviates with mean zero and a constant variance. The graphical analysis of residuals of the fitted models consists of normal probability plots and plots of the standardized residuals on length of gestation. The graphs provide initial checks for any assumption violations such as non-normality, the presence of heteroscedastic errors, and the presence of outliers. Outliers are identified by removing cases for which the standardized residuals, $r_i = \frac{e_i}{sd}$, are in excess of 3 in magnitude and these are classified as outliers and removed.

3.3.3 Modelling variability

A standard assumption of least squares theory is the homogeneity of error variance. Although minor deviations from the assumptions will have little effect on the regression analysis, gross violations can lead to least squares estimates that are inefficient. Heteroscedastic errors can be dealt with by a suitable transformation, such as a logarithmic transformation or more generally Box Cox transformations (Box and Cox (1964)). Altman and Chitty (1994) describe a process to identify heteroscedastic errors by modelling the standard deviations of a response variable as a function of gestational age. The paper discusses the study design and analysis necessary to derive centiles for fetal size and references a simple method for modelling the change in error standard deviations with gestation (Altman (1993)). The absolute value of the residuals of a fitted model are regressed on length of gestation in days, using a linear or quadratic model. The fitted values of this model multiplied by $\sqrt{\frac{\pi}{2}}$ provide estimates of the gestational age specific residual standard deviations. The significance of the slope parameter is used to detect the presence of heteroscedasticity. Suitable transformations are applied to stabilize the error variance. A weighted regression using the reciprocal of the square of the estimated age specific standard deviations allows for the increase but Altman and Chitty (1994) point out the change is almost always rather small.

The method of modelling gestational age specific residuals standard deviations described by Altman (1993) is used in the regression analysis of the trials data set out in sections 3.4-3.6. Illustrations of the fitted models are provided. Lowess curves are also fitted to the data. These provide robust estimates of the trend of the data over the range of gestational age which are based on median values rather than mean values. It has to be

recognized, however, that the distribution of the absolute values of the residuals is not a Gaussian distribution even if the residuals themselves have a Gaussian distribution. Significance tests for the relationship between the absolute residuals and gestational age are, therefore, invalid and are only used here as a guide.

3.4 AFP

Scatter diagrams of AFP and log(AFP) against gestational age for the pooled data for unaffected pregnancies are shown in Figure 3.1 and Figure 3.2. A random uniform effect is added to make the points distinguishable on the graph.

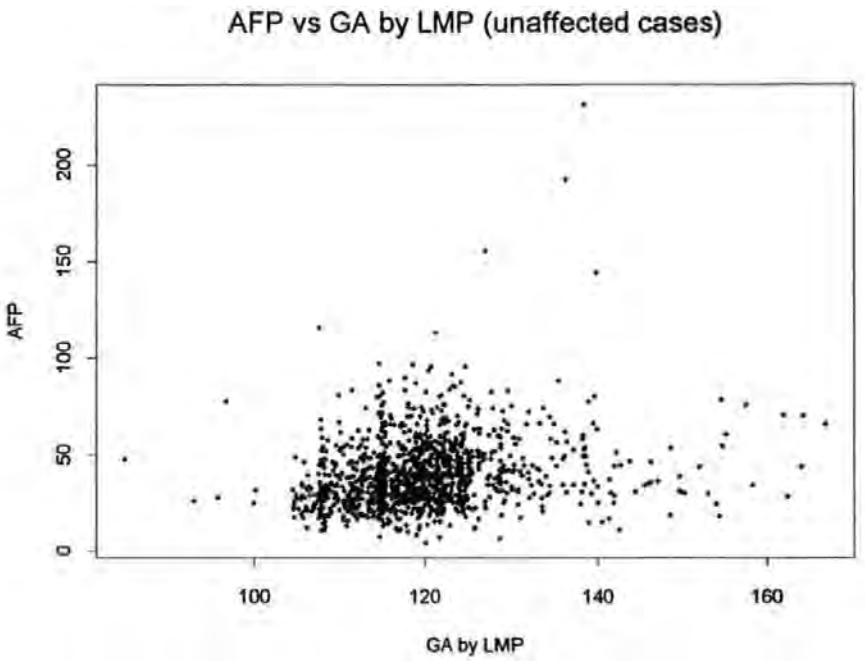

Figure 3.1: Plot of AFP concentrations against gestational age by LMP with a random uniform effect (unaffected pregnancies, n = 1299).

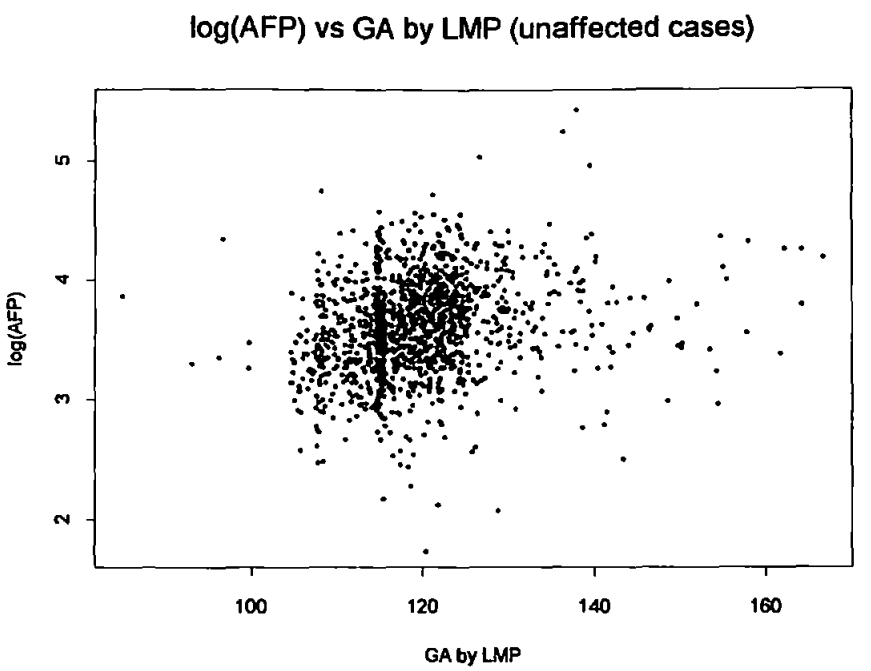

Figure 3.2: Plot of log(AFP) concentrations against gestational age by LMP with a random uniform effect (unaffected pregnancies, n = 1299).

Current approaches to model AFP concentrations use weighted median regression of log(AFP) on gestational age. It is generally accepted that the relationship between log(AFP) and gestational age is linear and that the variation about the regression is normal with constant variance. The models fitted in this section are of this form. The model for the full data set is defined as *ALIN1* and the model for the data set with outliers removed is defined as *ALIN2*. The form of the models is given below.

$$\log(\mathrm{AFP}) = \hat{\beta}_0 + \hat{\alpha}_j + \hat{\beta}_1(\mathrm{GA})$$

$\hat{\alpha}_j$ = centre effect for centres $j = 1, 2, 3, 4, 5$, $\hat{\beta}_0 + \hat{\alpha}_j$ = intercepts for centres $j = 1, 2, 3, 4, 5$, and $\hat{\beta}_0$ = intercept for centre 6.

The form of *ALIN1* is illustrated for centre 6 in Figure 3.3 which shows the fitted 10th, 50th and 90th centiles of AFP. The analysis of the fit of the models follows.

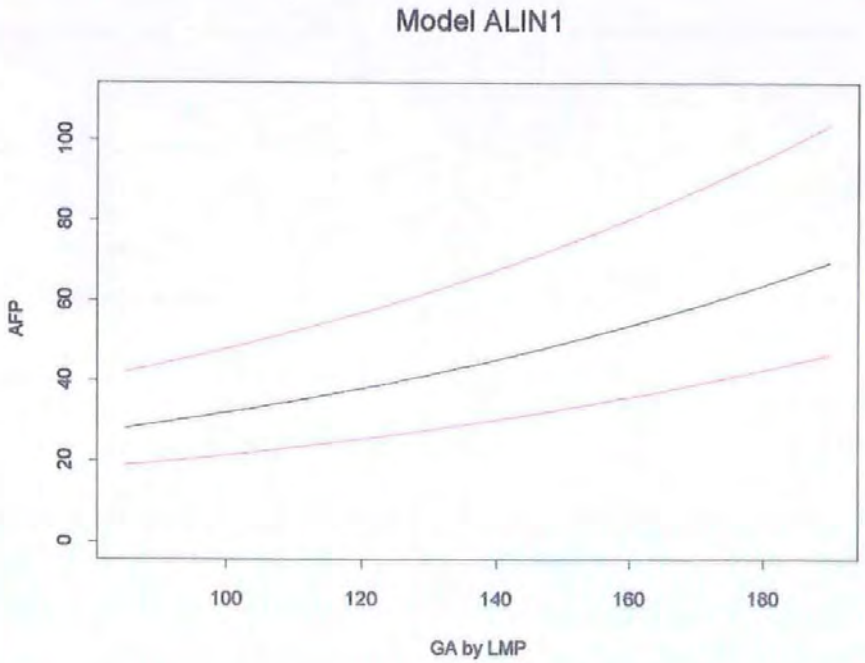

Figure 3.3: The fitted 10th, 50th and 90th centiles of AFP derived from model *ALIN1* for centre 6.

Model ALIN1

The parameter estimates for model *ALIN1* along with their standard errors are given in Table 3.6.

	VALUE	STANDARD ERROR
$\hat{\beta}_0$	2.6080	0.1611
$\hat{\alpha}_1$	0.0905	0.0278
$\hat{\alpha}_2$	-0.0462	0.0120
$\hat{\alpha}_3$	0.0182	0.0094
$\hat{\alpha}_4$	0.0015	0.0056
$\hat{\alpha}_5$	0.0048	0.0054
$\hat{\beta}_1$	0.0086	0.0014

Table 3.6: Parameter values of the fitted regression model *ALIN1*.

The R^2 value for the fitted model is 0.0859 indicating that 8.59% of the variation in the log(AFP) concentrations is explained by gestational age and centre effects. Although the R^2 value is small, the effects of gestational age and centre effect are significant ($p < 0.0001$). The normal probability plot of the residuals of the fitted model in Figure 3.4 shows evidence of deviations from a Gaussian form in the tails of the distribution but illustrates normality in the main body of the distribution. The plot of the standardized residuals against gestational age highlights the presence of outliers (Figure 3.5). The error variance remains constant across the range of gestational age. 11 outliers are identified and removed from the original data.

Model ALIN2

Table 3.7 lists the parameter estimates of the model fitted to the reduced data set.

	VALUE	STANDARD ERROR
$\hat{\beta}_0$	2.6699	0.1536
$\hat{\alpha}_1$	0.0859	0.0263
$\hat{\alpha}_2$	-0.0431	0.0114
$\hat{\alpha}_3$	0.0188	0.0089
$\hat{\alpha}_4$	0.0067	0.0054
$\hat{\alpha}_5$	0.0046	0.0052
$\hat{\beta}_1$	0.0081	0.0013

Table 3.7: Table of parameter estimates of the refitted regression *ALIN2*.

The R^2 value for the fitted model *ALIN2* is 0.0898 which is slightly greater than the R^2 value for model *ALIN1*. Figure 3.6 shows the normal probability plot of the residuals of

the fitted model. The outliers have little impact on the fitted model but they have a dramatic effect on the normal probability plot. Figures 3.7-3.12 illustrate the distribution of the centre specific residuals over the range of gestation. The centre specific residual standard deviations are shown in Table 3.8 along with estimates calculated from the product moment formula and robust estimates calculated between the $25^{th} - 75^{th}$ and $10^{th} - 90^{th}$ percentiles. In general the estimates are appreciably lower than those reported by Wald *et al* (1992). Figure 3.13 illustrates the distribution of the absolute residuals over the range of gestational age. The fitted values of this model provide estimates of the age specific standard deviations. The slope parameter of the fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ on length of gestation is not significant, $p > 0.05$. This confirms that the error variance can be assumed constant across the range of gestational age.

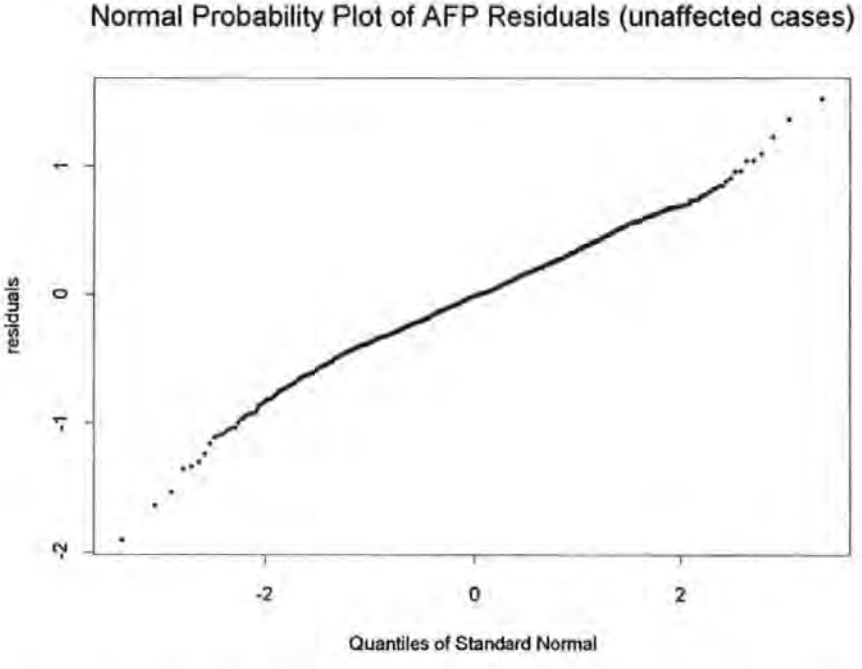

Figure 3.4: Normal probability plot of the residuals of the fitted model *ALIN1*.

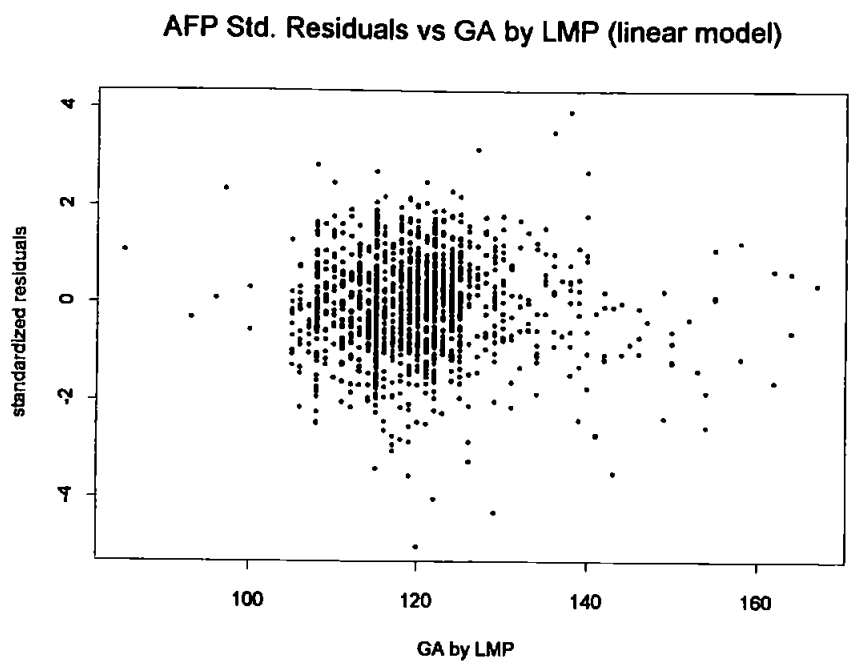

Figure 3.5: Plot of standardized residuals, r_i, of the fitted model *ALIN1* against gestational age.

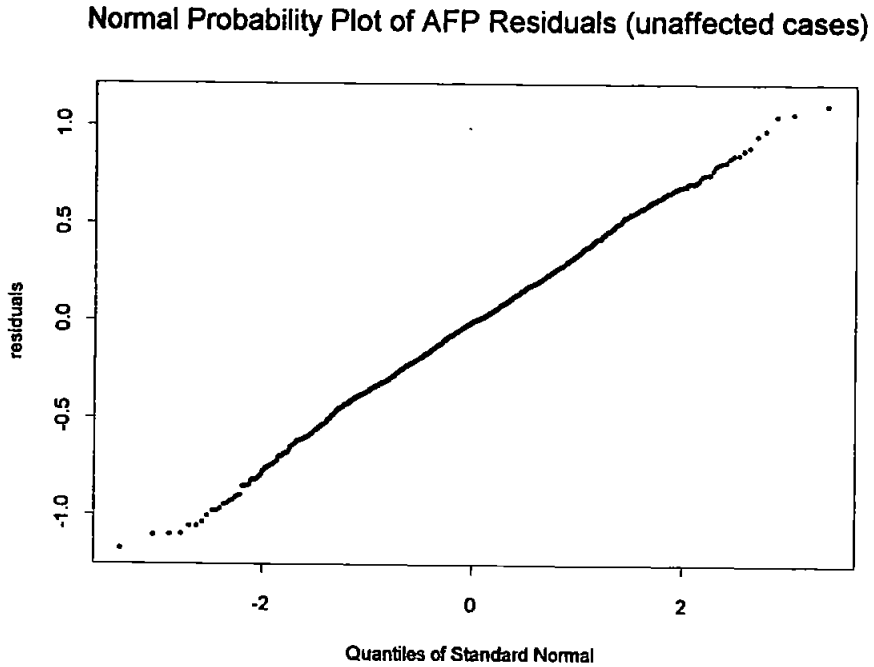

Figure 3.6: Normal probability plot of residuals of fitted model *ALIN2*.

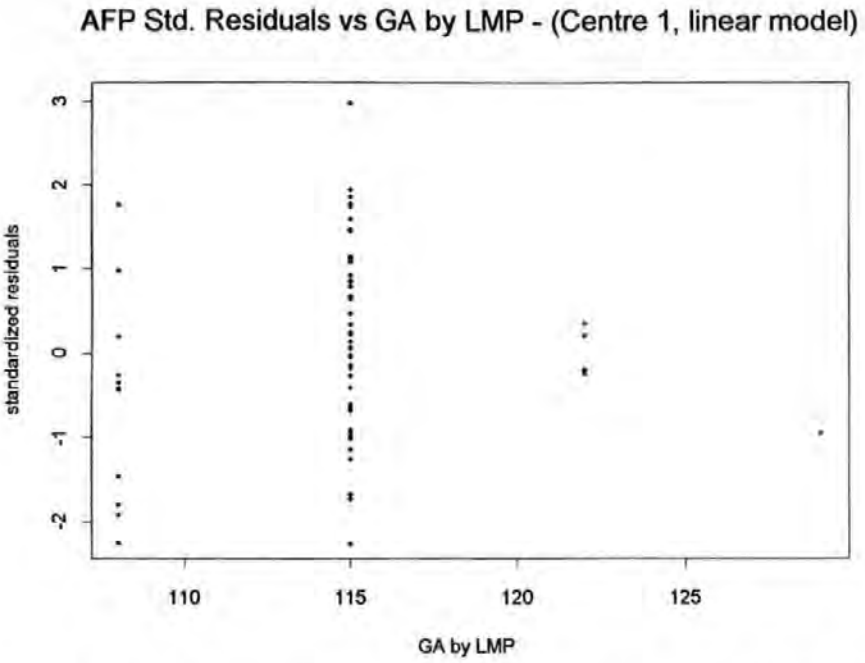

Figure 3.7: Plot of the standardized residuals , r_i , of the fitted model *ALIN2* for centre 1.

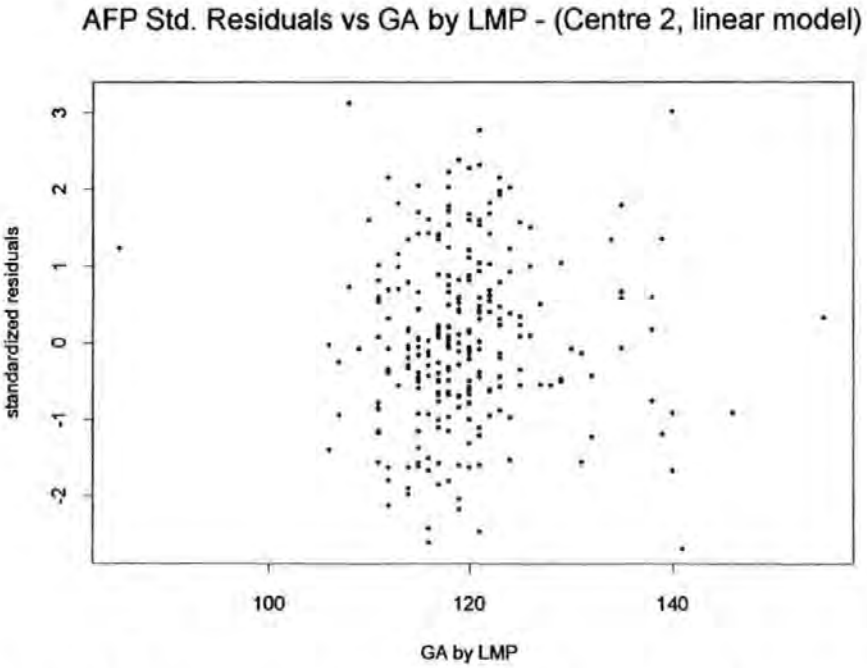

Figure 3.8: Plot of the standardized residuals, r_i , of the fitted model *ALIN2* for centre 2.

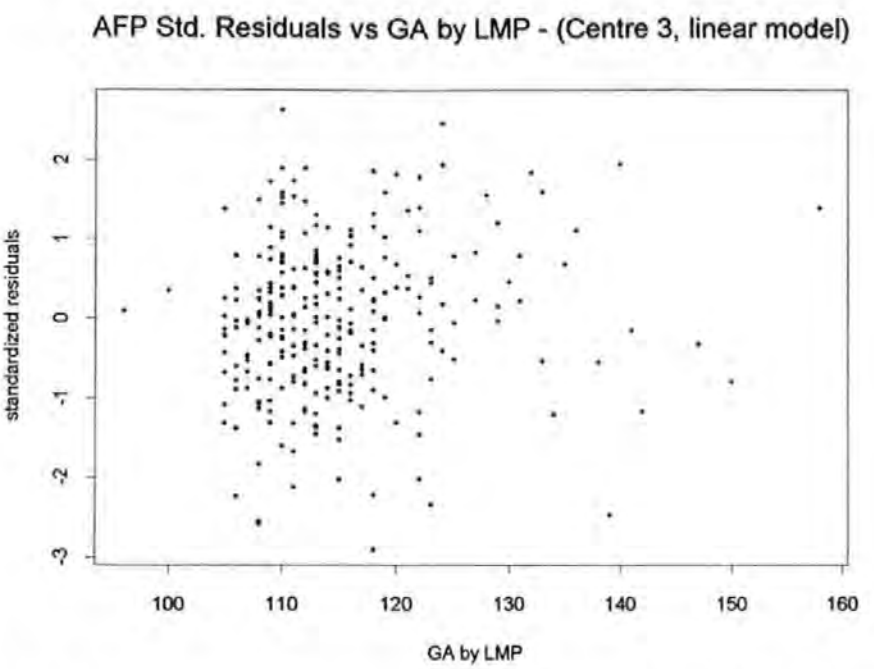

Figure 3.9: Plot of the standardized residuals, r_i, of the fitted model *ALIN2* for centre 3.

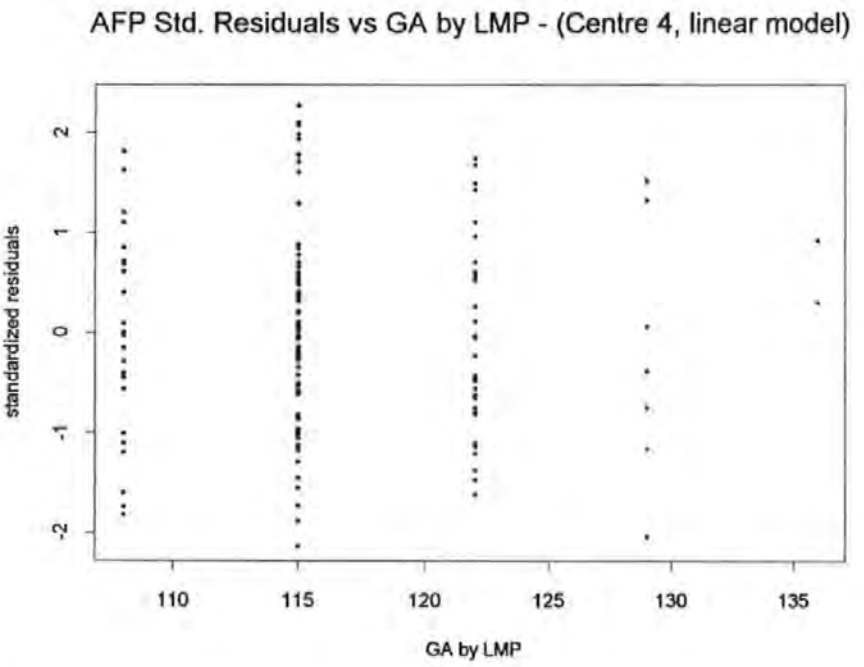

Figure 3.10: Plot of the standardized residuals, r_i, of the fitted model *ALIN2* for centre 4.

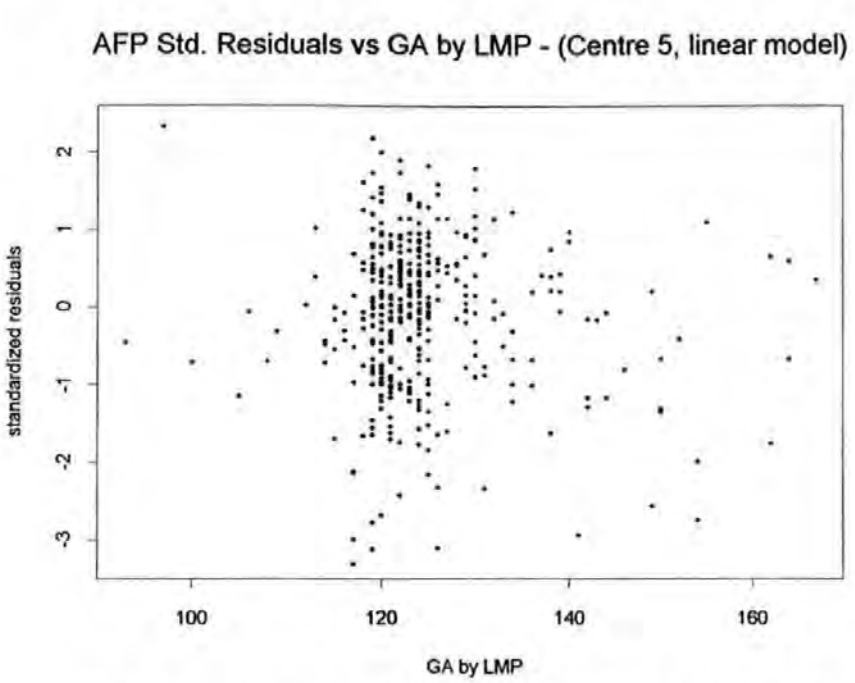

Figure 3.11: Plot of the standardized residuals, r_i, of the fitted model *ALIN2* for centre 5.

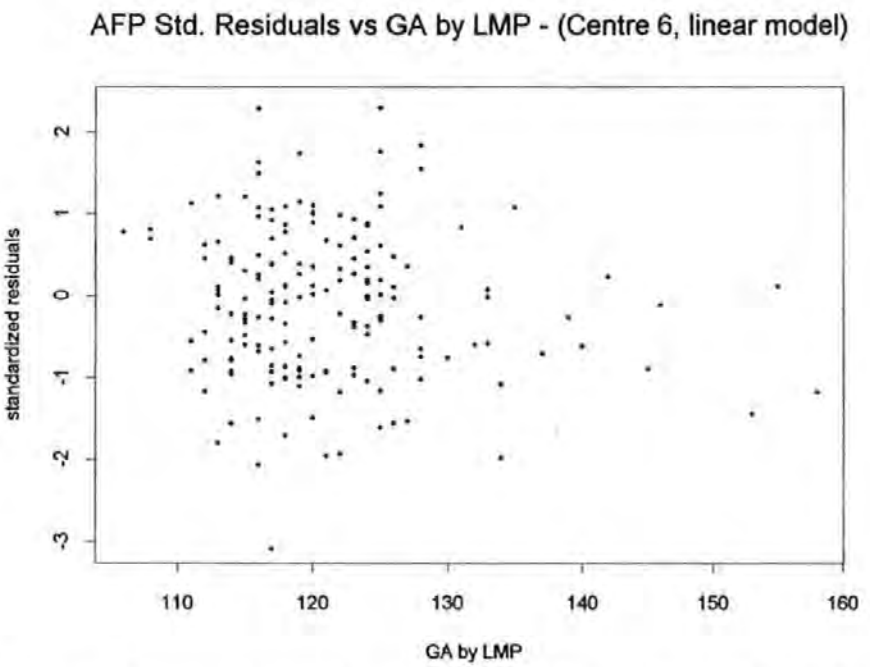

Figure 3.12: Plot of the standardized residuals, r_i, of the fitted model *ALIN2* for centre 6.

CENTRE No.	1	2	3	4	5	6
SD of residuals	0.4132	0.4010	0.3468	0.3619	0.3560	0.3318
SD estimate $Q_{0.25} - Q_{0.75}$	0.4056	0.3613	0.3263	0.3361	0.3492	0.3420
SD estimate $Q_{0.1} - Q_{0.9}$	0.4389	0.4323	0.3341	0.3814	0.3286	0.3140

Table 3.8: Summary of centre specific residual standard deviations of the fitted model *ALIN2*. Wald *et al* (1992) report the estimated standard deviation of log(MoM) AFP to be 0.4656, based on LMP dating methods and natural logarithms. The value is estimated between the $10^{th} - 90^{th}$ percentiles.

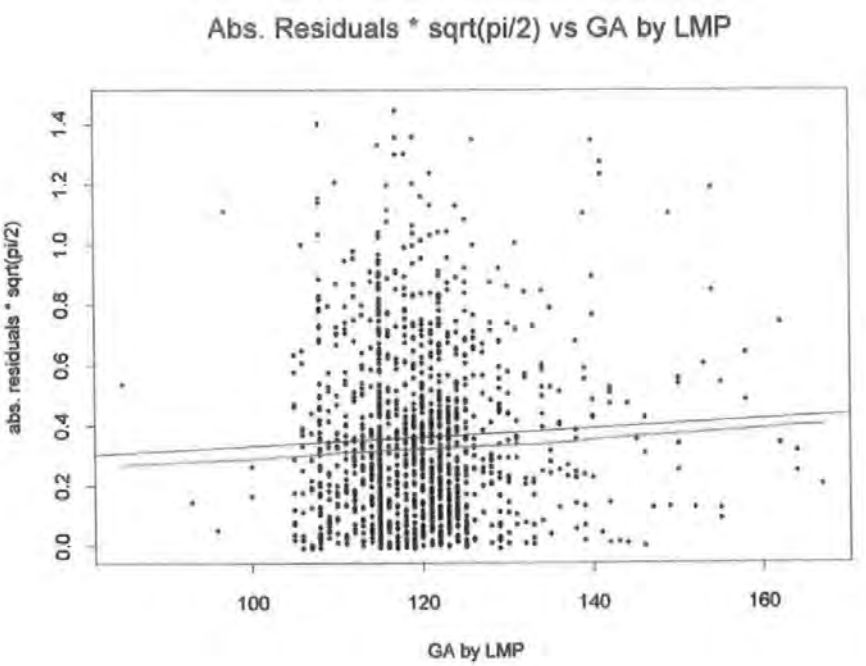

Figure 3.13: Plot of fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ of the model *ALIN2* against gestational age with lowess trend curve.

Fitted model :- $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i) = 0.1800+0.0015(GA)$

3.5 UE3

Scatter diagrams of UE3 and log(UE3) against gestational age for the pooled data for unaffected pregnancies are shown in Figure 3.14 and Figure 3.15. As with AFP, a random uniform effect is added to make the points distinguishable on the graph.

Wald *et al* (1988) assumed that UE3 depends linearly on gestational age and that the distribution of UE3 about the linear regression is normal with constant variance. Wald *et al* (1992) and (1993) conclude that the UE3 concentrations are more efficiently represented by a lognormal distribution. However, they assume that log(UE3) varies linearly with gestational age.

Wright *et al* (1995) point out that a better fit is obtained by using a model in which the location of UE3 depends linearly on gestational age, in accordance with Wald *et al* (1988), but that the variation about the line follows a lognormal distribution as in Wald *et al* (1992). With the above discussion in mind, three forms of models are fitted in this section. These are

(i) a simple linear regression of UE3 on gestational age (The models fitted of this form are defined as ULIN1 for the full data set, and ULIN2 for the data set with outliers removed);

(ii) a simple linear regression of log(UE3) on gestational age (The models fitted of this form are defined as ULOGLIN, for the full data set, and ULOGLIN2 for the data set with outliers removed.);

(iii) a non-linear regression of log(UE3) on gestational age in which the trend in UE3 depends linearly on gestational age but the distribution about the trend is lognormal. (The models fitted of this form are defined as UNLIN1 for the full data set, and UNLIN2 for the data set with outliers removed).

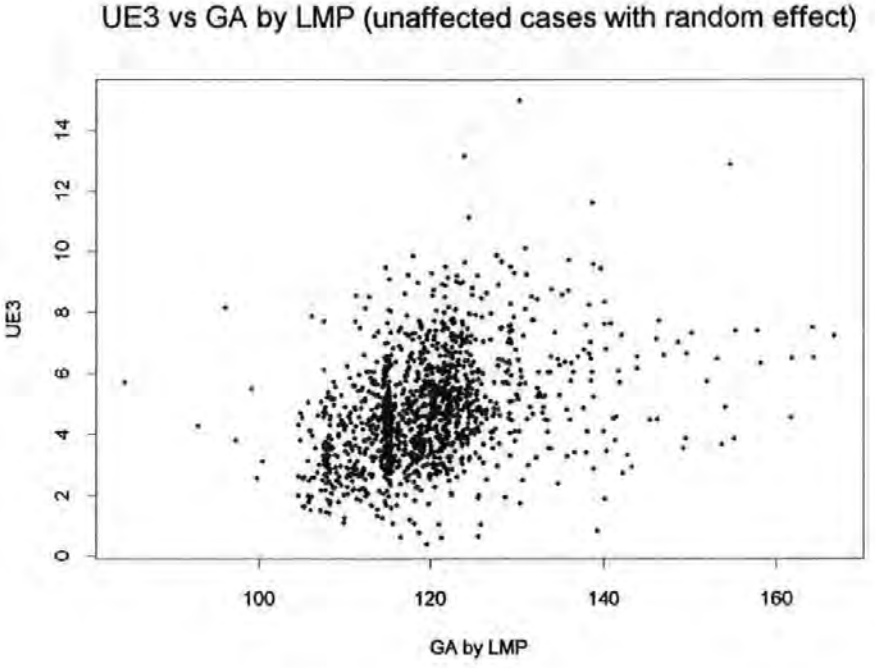

Figure 3.14: Plot of UE3 concentrations against gestational age by LMP with a random uniform effect (unaffected pregnancies, n = 1333).

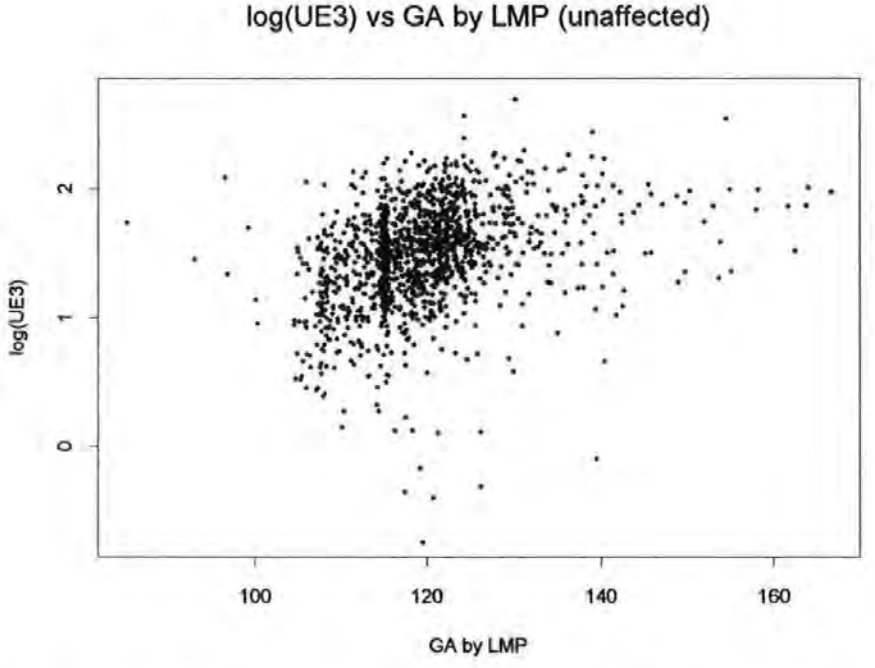

Figure 3.15: Plot of log(UE3) concentrations against gestational age by LMP with a random uniform effect (unaffected pregnancies, n = 1333).

Models described by (iii) can be considered as a hybrid between the model used in Wald *et al* (1988) and in Wald *et al* (1992). The analysis of the fit of the three models is given below. The forms of these models are illustrated for centre 6 in Figures 3.16-3.18 which show the fitted 10th, 50th and 90th centiles of UE3.

(i) A simple linear regression of UE3 on gestational age

$$UE3 = \hat{\beta}_0 + \hat{\alpha}_j + \hat{\beta}_1(GA)$$

$\hat{\alpha}_j$ = centre effect for centres $j = 1, 2, 3, 4, 5$, $\hat{\beta}_0 + \hat{\alpha}_j$ = intercepts for centres $j = 1, 2, 3, 4, 5$, and $\hat{\beta}_0$ = intercept for centre 6.

Model ULIN1

Table 3.9 lists the parameter estimates for the model fitted *ULIN1*.

	VALUE	STANDARD ERROR
$\hat{\beta}_0$	-2.7355	0.6630
$\hat{\alpha}_1$	0.4533	0.1144
$\hat{\alpha}_2$	-0.2976	0.0496
$\hat{\alpha}_3$	0.0483	0.0387
$\hat{\alpha}_4$	0.0363	0.0229
$\hat{\alpha}_5$	0.0173	0.0216
$\hat{\beta}_1$	0.0632	0.0056

Table 3.9: Parameter estimates of the fitted regression model *ULIN1*.

The R^2 value is 0.197, indicating that 19.7% of the variation in the UE3 concentrations is explained by gestational age and centre effects. Although the R^2 value is small, the effects of gestational age and centre are overwhelmingly significant ($p < 0.0001$ for both gestational age and centre effects). The normal probability plot of the residuals of the fitted model is shown in Figure 3.19. The plot illustrates deviations from a linear form in the tails. Figure 3.20 shows the distribution of the pooled standardized residuals of the fitted model. 7 outliers are identified and these are removed from the original data.

Model ULIN2

Table 3.10 lists the parameter estimates for the model fitted to the reduced data set.

	VALUE	STANDARD ERROR
$\hat{\beta}_0$	-2.4695	0.6390
$\hat{\alpha}_1$	0.4595	0.1091
$\hat{\alpha}_2$	-0.3063	0.0473
$\hat{\alpha}_3$	0.0502	0.0473
$\hat{\alpha}_4$	0.0237	0.0219
$\hat{\alpha}_5$	0.0173	0.0207
$\hat{\beta}_1$	0.0608	0.0054

Table 3.10 : Parameter estimates for the fitted regression model *ULIN2*.

The R^2 value for the fitted model *ULIN2* is 0.2024. This is an improvement on the R^2 value for model *ULIN1*. Figure 3.21 shows the normal probability plot of the residuals of the fitted model. Although the outliers have little impact on the regression coefficients, they have a dramatic effect on the normal probability plot. Figures 3.22-3.27 show the

centre specific standardized residuals across the range of gestation. There is evidence to suggest the error variance increases with gestational age. This is particularly noticeable in the plot for centre 3. The standard deviations estimated from the product moment formula and from values between the $25^{th} - 75^{th}$ and $10^{th} - 90^{th}$ percentiles are given in Table 3.15. These are somewhat larger than the estimate reported by Wald *et al* (1988). The fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ on gestational age is shown in Figure 3.28. The regression coefficient of this model is significantly different from zero ($p < 0.05$). This confirms that the error variance varies with gestational age. One way that this can be dealt with is to apply a log transformation to UE3 as in the models presented under (ii) and (iii) below.

(ii) A simple linear regression of log(UE3) on gestational age

$$\log(UE3) = \hat{\beta}_0 + \hat{\alpha}_j + \hat{\beta}_1(GA)$$

$\hat{\alpha}_j$ = centre effect for centres $j = 1, 2, 3, 4, 5$, $\hat{\beta}_0 + \hat{\alpha}_j$ = intercepts for centres $j = 1, 2, 3, 4, 5$, and $\hat{\beta}_0$ = intercept for centre 6.

Model ULOGLIN1

The parameter estimates for model *ULOGLIN1* along with their standard errors are given in Table 3.11.

	VALUE	STANDARD ERROR
$\hat{\beta}_0$	-0.0111	0.1489
$\hat{\alpha}_1$	0.0854	0.0257
$\hat{\alpha}_2$	-0.0731	0.0111
$\hat{\alpha}_3$	0.0188	0.0087
$\hat{\alpha}_4$	0.0092	0.0051
$\hat{\alpha}_5$	0.0063	0.0049
$\hat{\beta}_1$	0.0127	0.0013

Table 3.11: Parameter estimates of the fitted regression model *ULOGLIN1*.

The R^2 value is 0.1848 indicating that 18.48% of the variation in the log(UE3) concentrations is explained by gestational age and centre effects. Again, the R^2 value is small, but the effects of gestational age and centre effect are overwhelmingly significant ($p < 0.0001$ for both gestational age and centre effects). The normal probability plot of the residuals of the fitted model, show large deviations from a linear form in the tails of the plot in Figure 3.29. Figure 3.30 shows the distribution of the pooled standardized residuals of the fitted model. The error variance appears to be consistent across the range of gestational age. 13 outliers are identified and these are removed from the original data.

Model ULOGLIN2

Table 3.12 lists the parameter estimates of the model fitted to the reduced data set.

	VALUE	STANDARD ERROR
$\hat{\beta}_0$	-0.0877	0.1342
$\hat{\alpha}_1$	0.0996	0.0230
$\hat{\alpha}_2$	-0.0778	0.0100
$\hat{\alpha}_3$	0.0161	0.0078
$\hat{\alpha}_4$	0.0101	0.0046
$\hat{\alpha}_5$	0.0069	0.0044
$\hat{\beta}_1$	0.0134	0.0011

Table 3.12: Parameter estimates of the fitted regression model *ULOGLIN2.*

The R^2 value for the fitted model *ULOGLIN2* is 0.243 indicating that 24.3% of the variation in the log(UE3) concentrations is explained by gestational age and centre. This is an improvement on the R^2 value for model *ULOGLIN1*. Figure 3.31 shows the normal probability plot of the residuals of the fitted model. The plot is linear over a greater range than the plot associate with model *ULOGLIN1*. Figures 3.32-3.37 show the centre specific standardized residuals across the range of gestational age. There is no evidence to suggest the error variance varies with gestational age. The standard deviations estimated from the product moment formula and from the values between the $25^{th} - 75^{th}$ and $10^{th} - 90^{th}$ percentiles are given in Table 3.16. These are generally lower than the estimate reported by Wald *et al* (1993). The fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ on gestational age is shown in Figure 3.38. The regression coefficient of this model is not significantly different from

zero ($p > 0.05$) This confirms that the error variance can be assumed constant over the range of gestational age.

(iii) A non-linear regression of log(UE3) on gestational age in which the trend in UE3 depends linearly on gestational age but the distribution about the trend is lognormal

$$\log(\text{UE3}) = \hat{\alpha}_j + \log[\hat{\beta}_0 + \hat{\beta}_1(\text{GA})]$$

$\hat{\alpha}_j$ = centre effect for centres $j = 1, 2, 3, 4, 5$, $\log(\hat{\beta}_0) + \hat{\alpha}_j$ = intercepts for centres $j =$ 1, 2, 3, 4, 5, and $\log(\hat{\beta}_0)$ = intercept for centre 6.

Model UNLIN1

The parameter estimates for model *UNLIN1* along with their standard errors are given in Table 3.13.

	VALUE	STANDARD ERROR
$\hat{\alpha}_1$	-0.0725	0.0533
$\hat{\alpha}_2$	0.0935	0.0339
$\hat{\alpha}_3$	-0.2021	0.0347
$\hat{\alpha}_4$	0.0129	0.0395
$\hat{\alpha}_5$	-0.0025	0.0315
$\hat{\beta}_0$	-3.0285	0.7521
$\hat{\beta}_1$	0.0644	0.0064

Table 3.13: Parameter estimates of the fitted regression model *UNLIN1*.

It is hypothesized that a non-linear model that is a hybrid form of the models given in Wald *et al* (1988) and (1992) may provide a more appropriate description of the data in terms of fit. The R^2 value for the fitted model is 0.1875, indicating that 18.75% of the variation in the log(UE3) values is explained by gestational age and centre. This value is comparable to the R^2 values for the previous models. Both the effects of gestational age and centre effect are overwhelmingly significant ($p < 0.0001$ for both gestational age and centre effects). Figure 3.39 shows the normal probability plot of the residuals. The plot demonstrates large deviations from a linear form in the tails. The standardized residuals for the pooled data are plotted against gestational age in Figure 3.40. 15 outliers are identified and these are removed from the original data.

Model UNLIN2

The parameter estimates of the model fitted to the reduced data set are shown in Table 3.14.

	VALUE	STANDARD ERROR
$\hat{\alpha}_1$	-0.0807	0.0472
$\hat{\alpha}_2$	0.1114	0.0309
$\hat{\alpha}_3$	-0.2058	0.0309
$\hat{\alpha}_4$	0.0023	0.0351
$\hat{\alpha}_5$	-0.0010	0.0280
$\hat{\beta}_0$	-3.7616	0.6814
$\hat{\beta}_1$	0.0711	0.0059

Table 3.14: Parameter estimates for the fitted regression model *UNLIN2*.

The R^2 value of the fitted model *UNLIN2* is 0.2502 which is an improvement on the value for the fitted model *UNLIN1*. This value is marginally greater than the R^2 for model *ULOGLIN2*.

The normal probability plot of the residuals of the fitted model is shown in Figure 3.41. Again, the removal of outliers has little impact on the fitted model but greatly improves the linearity of the normal probability plot in Figure 3.39. The distribution of the centre specific residuals on length of gestation are given in Figures 3.42-3.47. There is no evidence in the plots to indicate that the error variance is correlated with gestational age. The estimated standard deviations, shown in Table 3.17, are similar to those calculated from the model *ULOGLIN2* and are generally lower than the estimates reported by Wald *et al* (1993). The slope parameter of the fitted model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ on gestational age is not significantly different from zero ($p > 0.05$) and therefore, there is no evidence of any violation of the assumption of homoscedasticity.

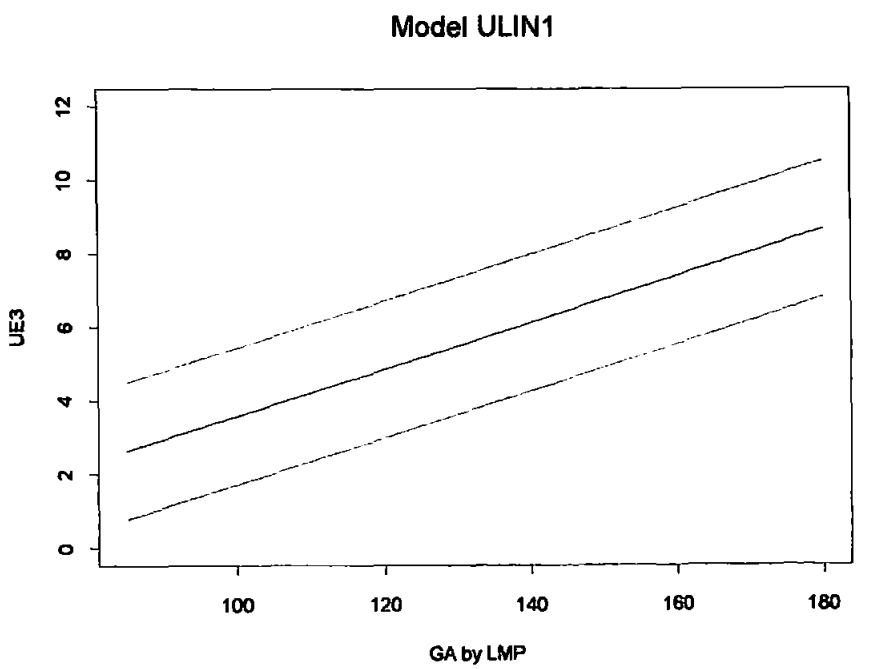

Figure 3.16: Plot of UE3 concentrations against gestational age by LMP with a random uniform effect (unaffected pregnancies, n = 1333).

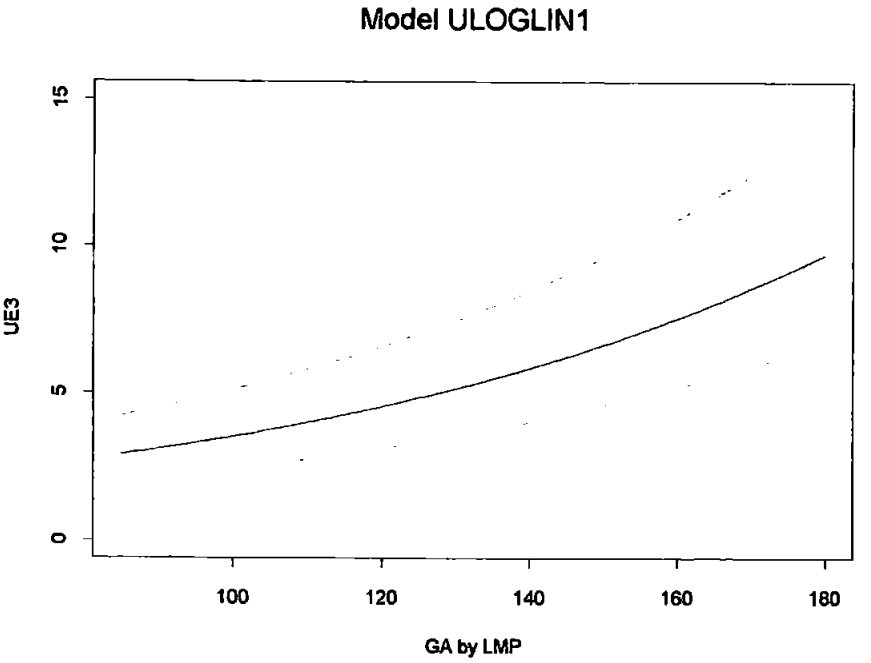

Figure 3.17: The fitted 10th, 50th and 90th centiles of UE3 derived from model *ULOGLIN1* for centre 6.

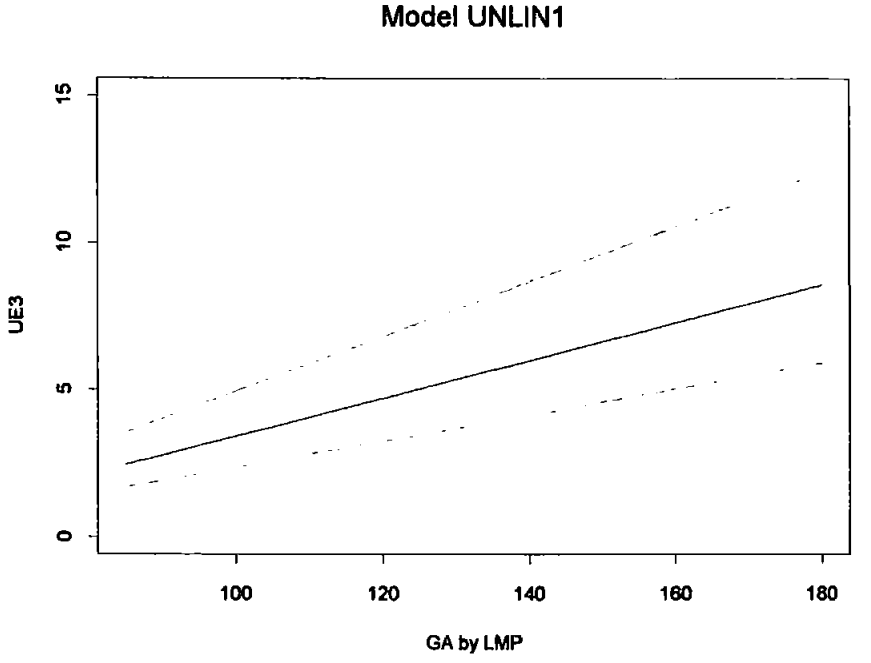

Figure 3.18: The fitted 10th, 50th and 90th centiles of UE3 derived from model *UNLIN1* for centre 6.

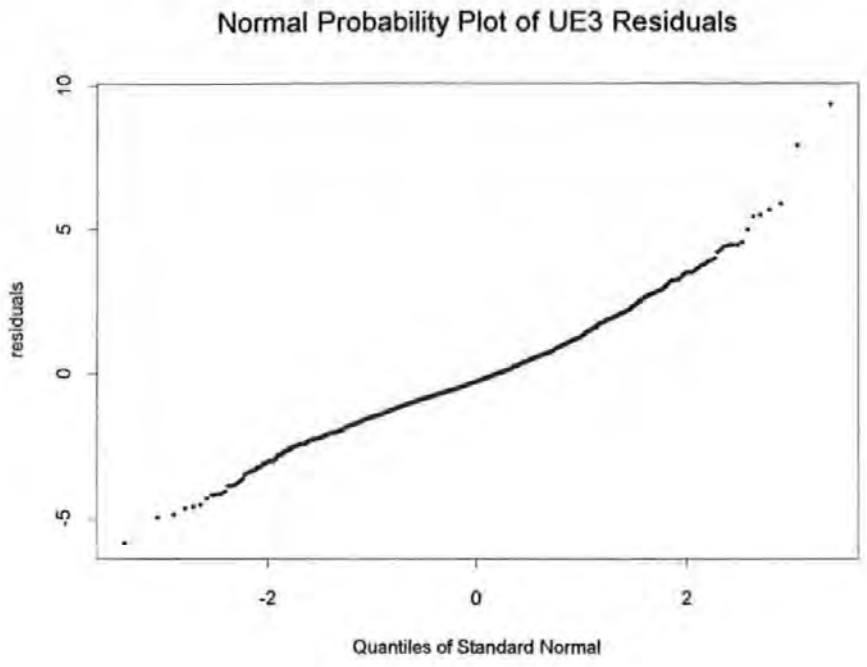

Figure 3.19: Normal probability plot of the residuals of the fitted model *ULIN1*.

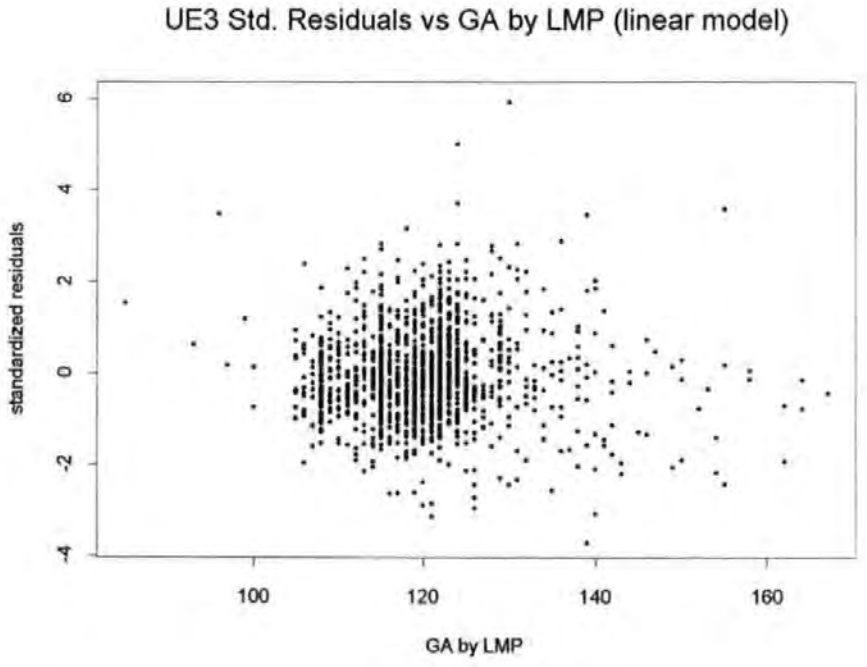

Figure 3.20: Plot of standardized residuals, r_i, of the fitted model *ULIN1* against gestational age.

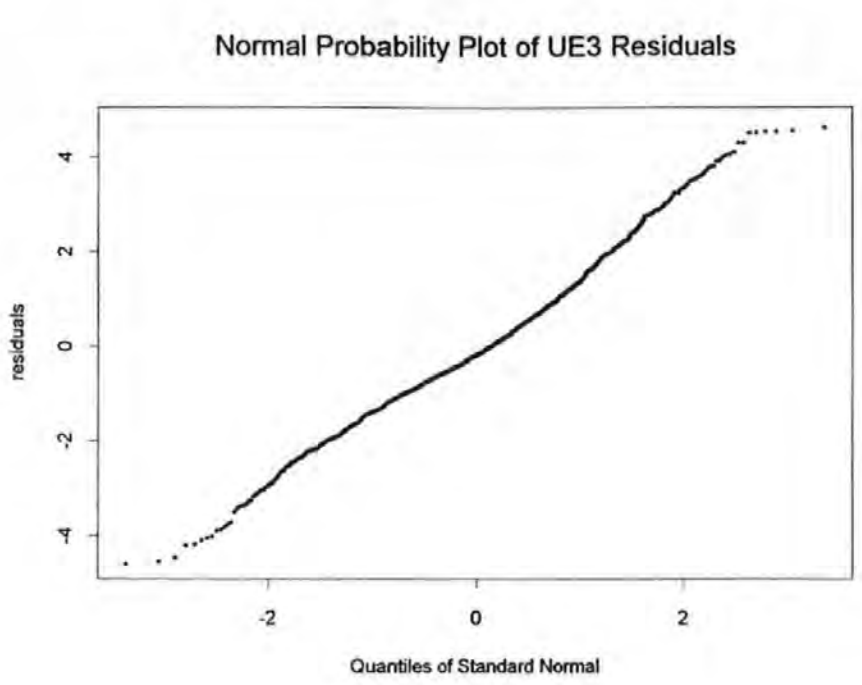

Figure 3.21: Normal probability plot of the residuals of the fitted model *ULIN2*.

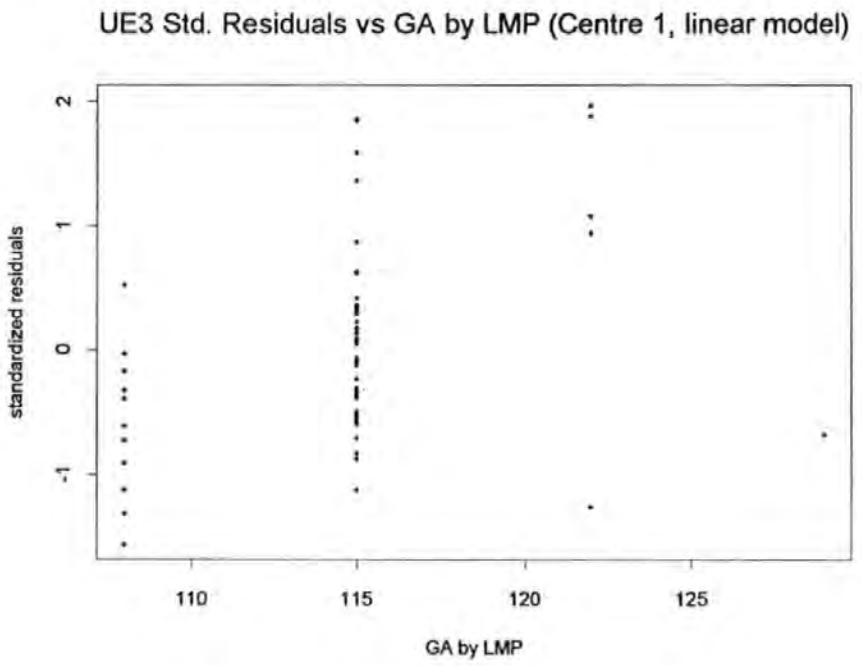

Figure 3.22: Plot of standardized residuals, r_i, of the fitted model *ULIN2* for centre 1.

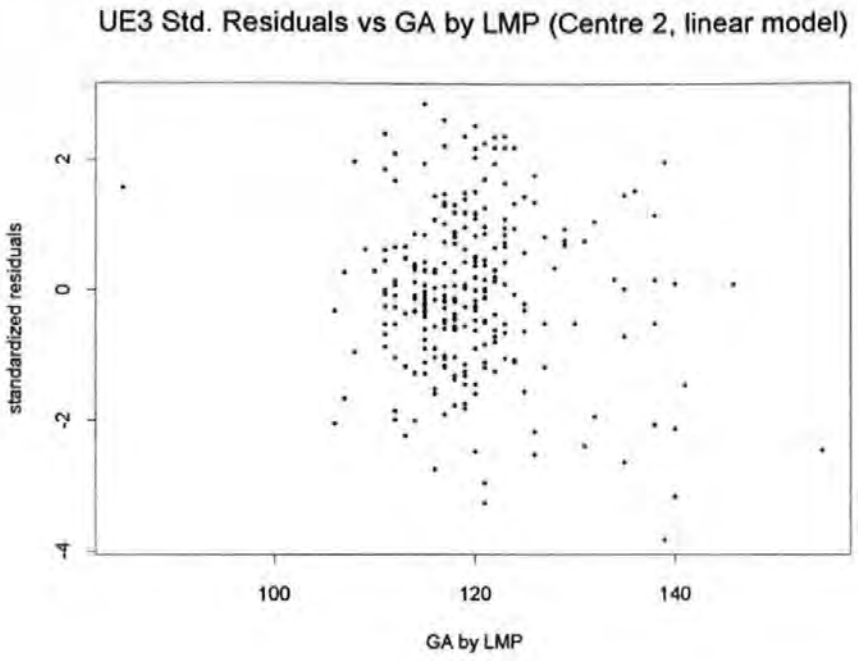

Figure 3.23: Plot of standardized residuals, r_i, of the fitted model *ULIN2* for centre 2.

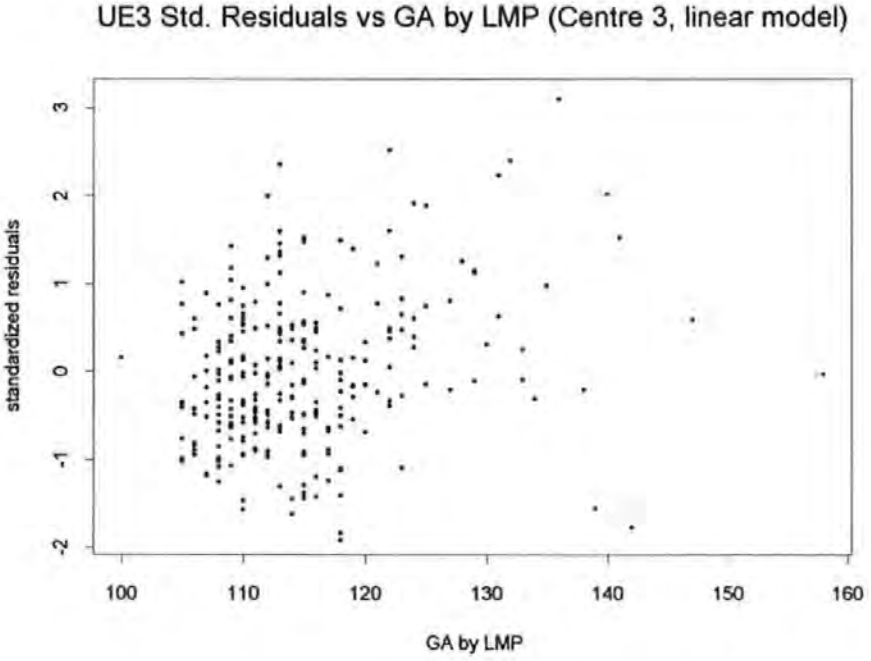

Figure 3.24: Plot of standardized residuals, r_i, of the fitted model *ULIN2* for centre 3.

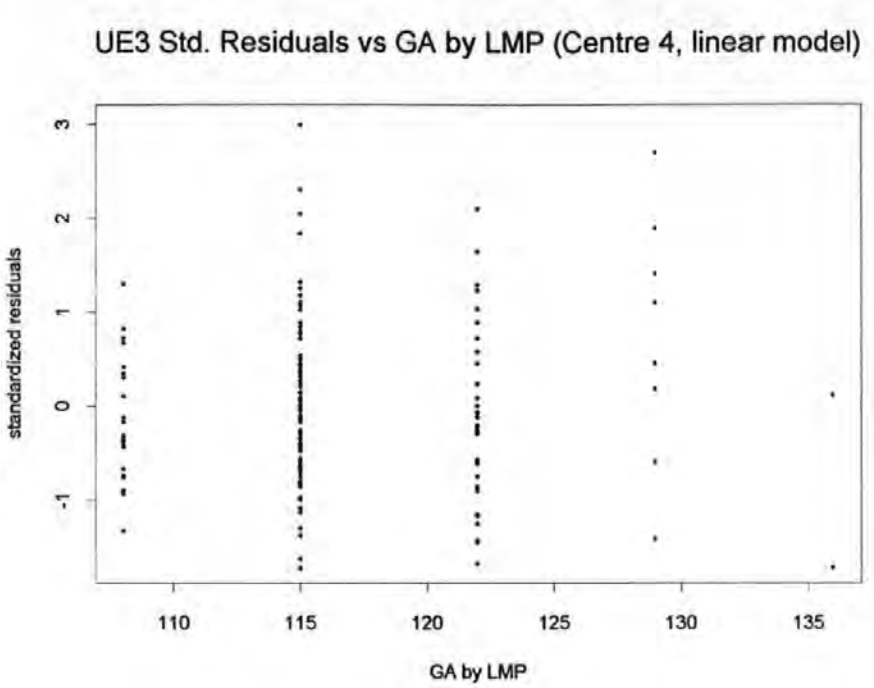

Figure 3.25: Plot of standardized residuals, r_i, of the fitted model *ULIN2* for centre 4.

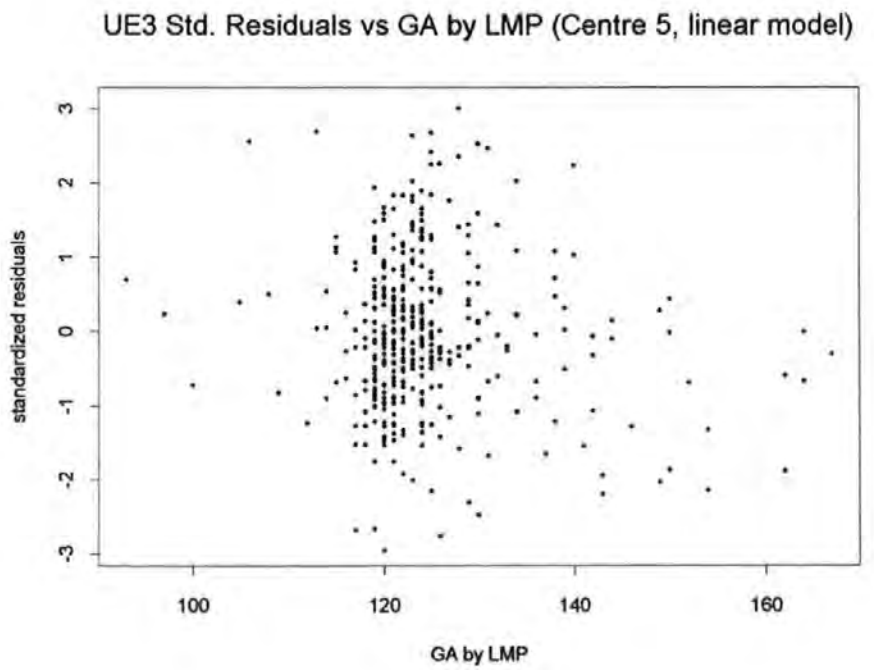

Figure 3.26: Plot of standardized residuals, r_i, of the fitted model *ULIN2* for centre 5.

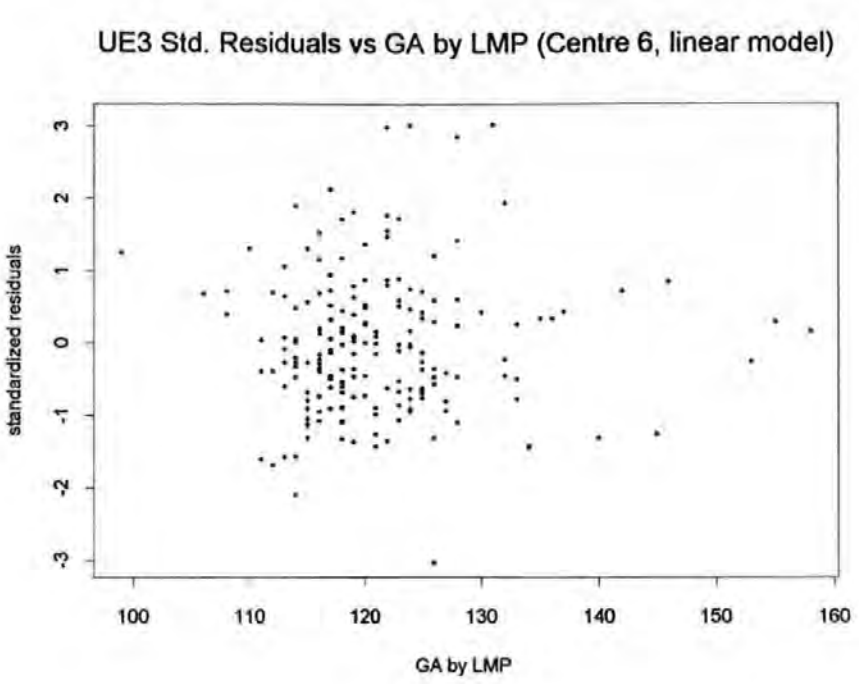

Figure 3.27: Plot of standardized residuals, r_i, of the fitted model *ULIN2* for centre 6.

CENTRE No.	1	2	3	4	5	6
SD of residuals	1.2476	1.8246	1.2770	1.3559	1.5528	1.4503
SD estimate $Q_{0.25} - Q_{0.75}$	1.0434	1.5991	1.1957	1.2295	1.3983	1.3258
SD estimate $Q_{0.1} - Q_{0.9}$	1.1715	1.8160	1.2455	1.3051	1.5153	1.3626

Table 3.15: Summary of centre specific residual standard deviations of the fitted model *ULIN2*. Wald *et al* (1993) report the estimated standard deviation of UE3 between the $10^{th} - 90^{th}$ percentiles to be 0.5663 based on LMP dating methods and natural logarithms.

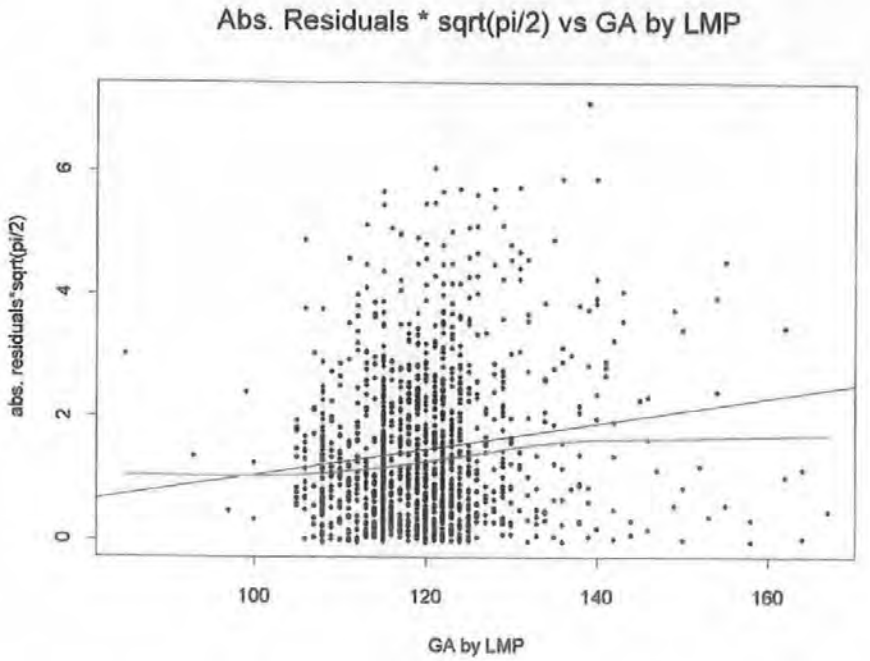

Figure 3.28: Plot of fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ of the model *ULIN2* against gestational age with lowess trend curve.

Fitted model :- $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i) = -1.0564+0.0210(\text{GA})$

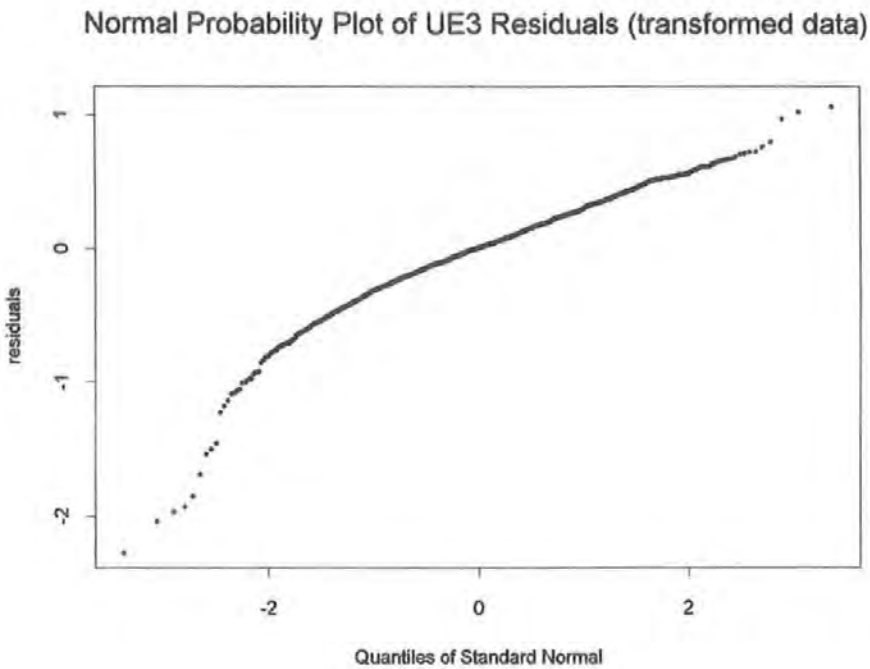

Figure 3.29: Normal probability plot of the residuals of the fitted model *ULOGLIN1*.

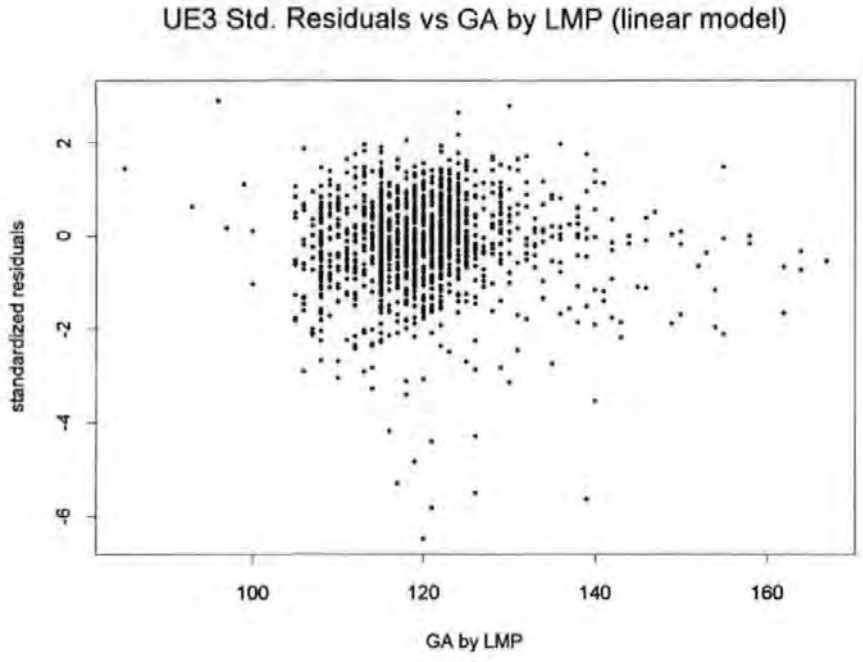

Figure 3.30: Plot of standardized residuals, r_i, of the fitted model *ULOGLIN1* against gestational age.

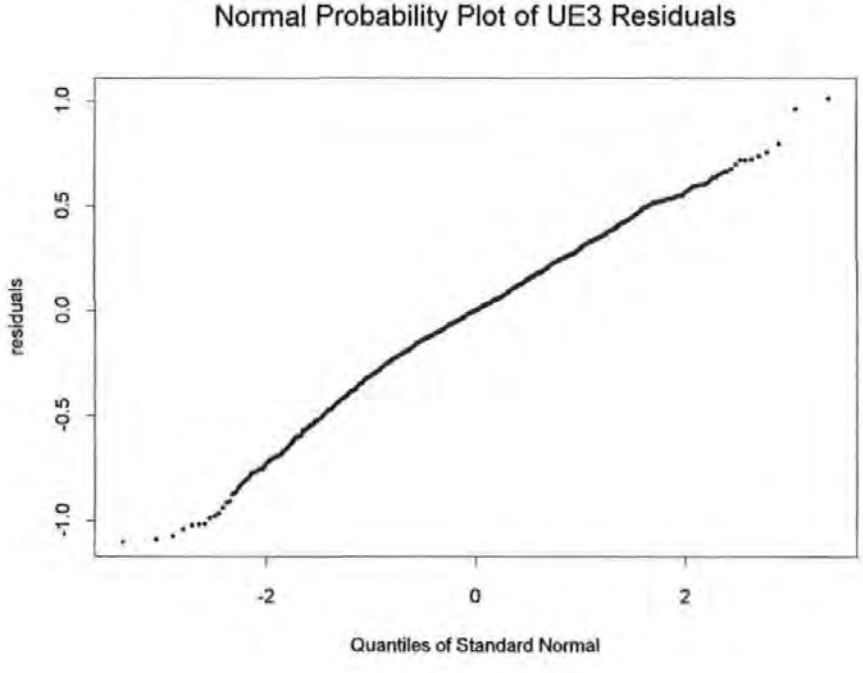

Figure 3.31: Normal probability plot of the residuals of the fitted model *ULOGLIN2*.

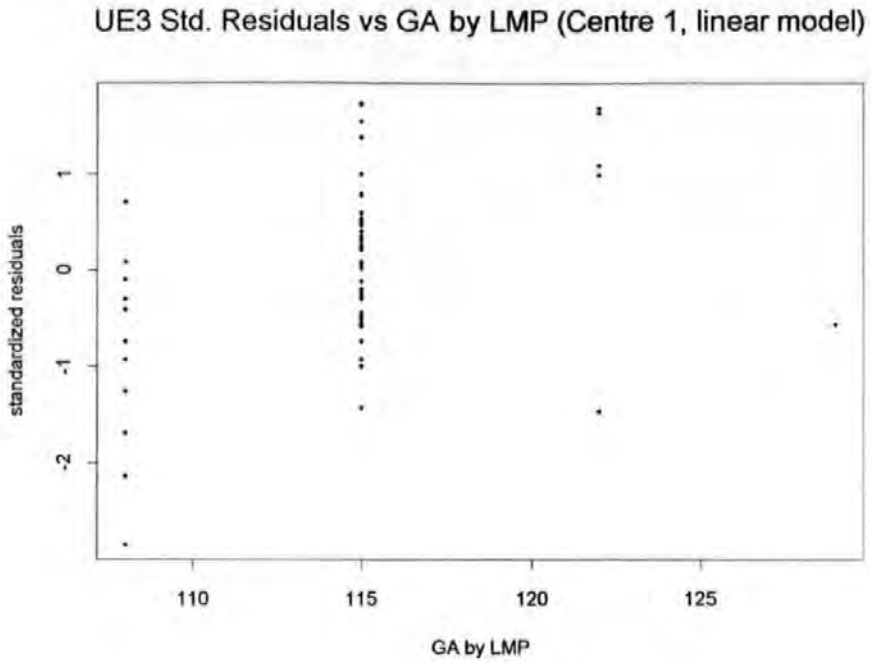

Figure 3.32: Plot of the standardized residuals, r_i, of the fitted model *ULOGLIN2* for centre 1.

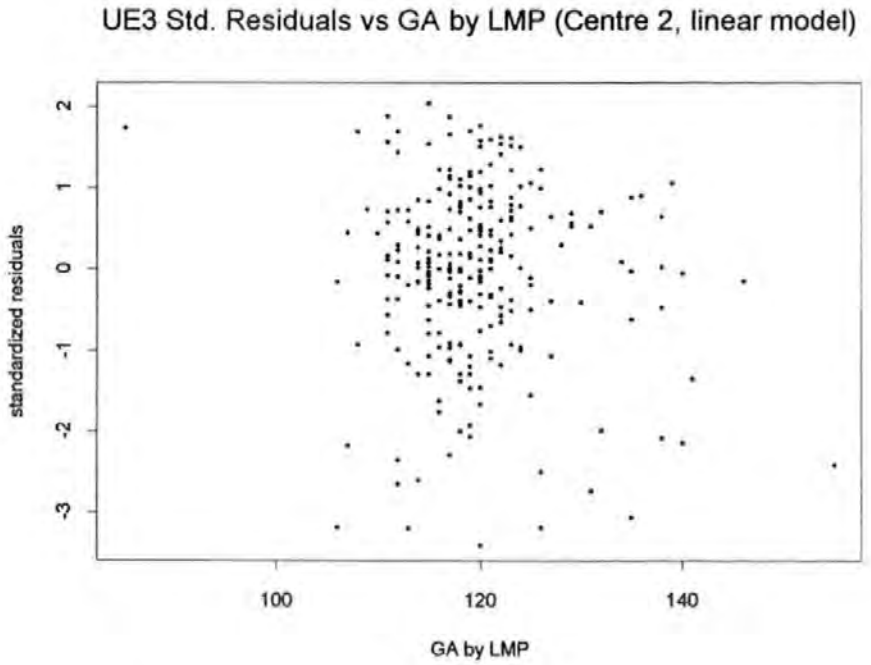

Figure 3.33: Plot of the standardized residuals, r_i, of the fitted model *ULOGLIN2* for centre 2.

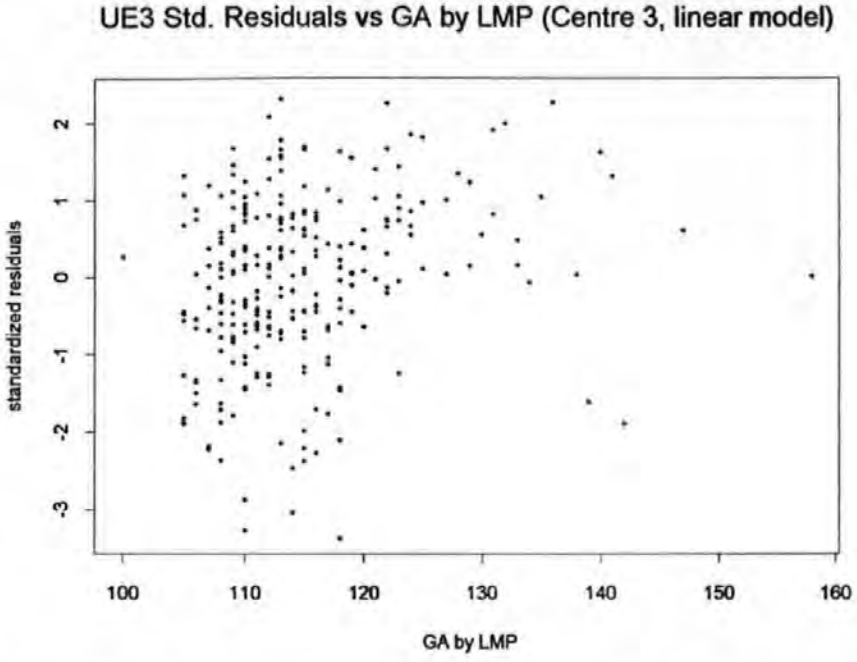

Figure 3.34: Plot of the standardized residuals, r_i, of the fitted model *ULOGLIN2* for centre 3.

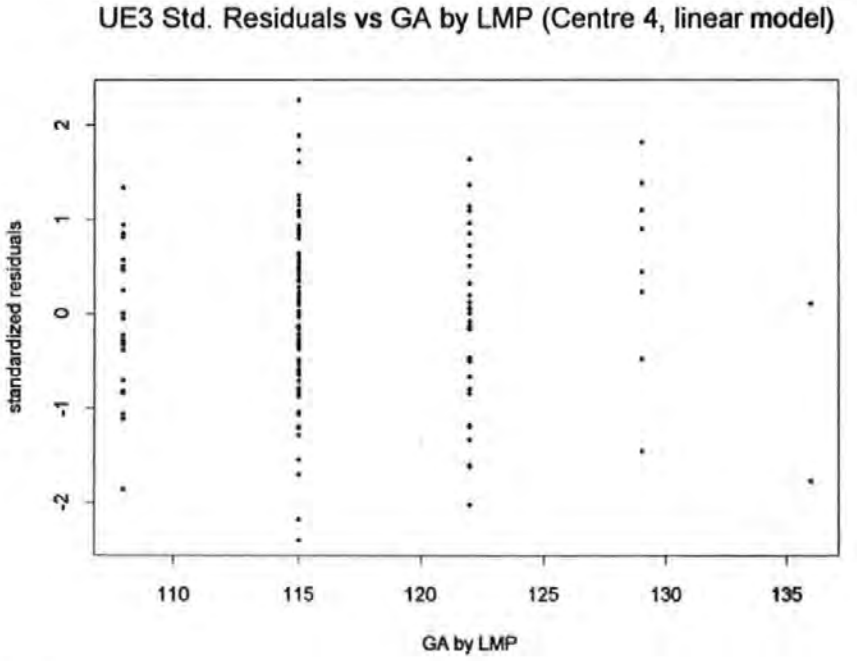

Figure 3.35: Plot of the standardized residuals, r_i, of the fitted model *ULOGLIN2* for centre 4.

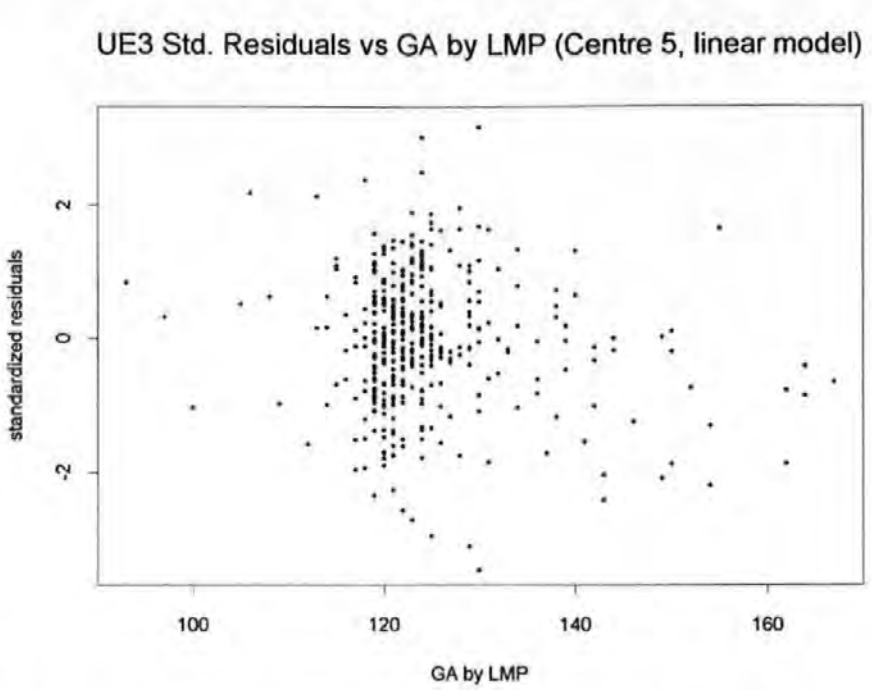

Figure 3.36: Plot of the standardized residuals, r_i, of the fitted model *ULOGLIN2* for centre 5.

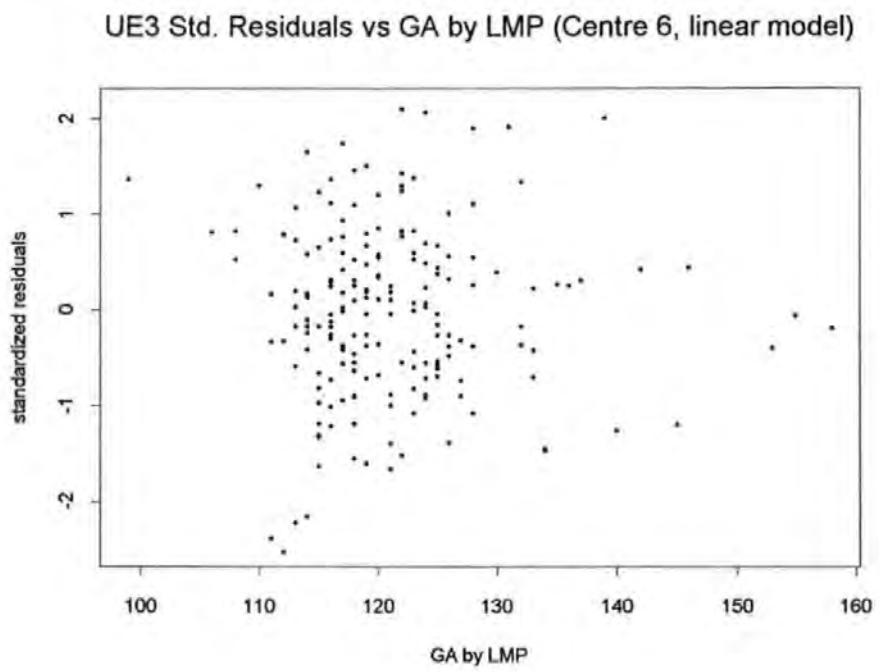

Figure 3.37: Plot of the standardized residuals, r_i, of the fitted model *ULOGLIN2* for centre 6.

CENTRE No.	1	2	3	4	5	6
SD of residuals	0.3000	0.3399	0.3451	0.2818	0.3152	0.2823
SD estimate $Q_{0.25} - Q_{0.75}$	0.2544	0.2763	0.3324	0.2733	0.2873	0.2676
SD estimate $Q_{0.1} - Q_{0.9}$	0.2660	0.3205	0.3445	0.2836	0.3118	0.2881

Table 3.16: Summary of centre specific residual standard deviations of the fitted model *ULOGLIN2*. Wald *et al* (1993) report the estimated standard deviation of log(MoM) UE3 between the $10^{th} - 90^{th}$ percentiles to be 0.3362 based on LMP dating methods and natural logarithms.

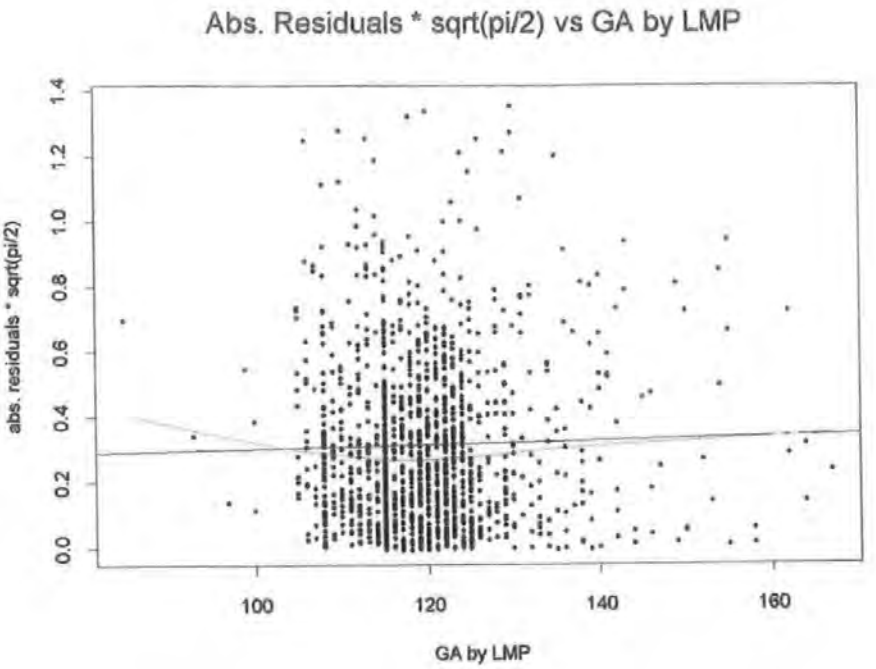

Figure 3.38: Plot of fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ of the model *ULOGLIN2* against gestational age with lowess trend curve.

Fitted model :- $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i) = 0.2419 + 0.0006(\text{GA})$

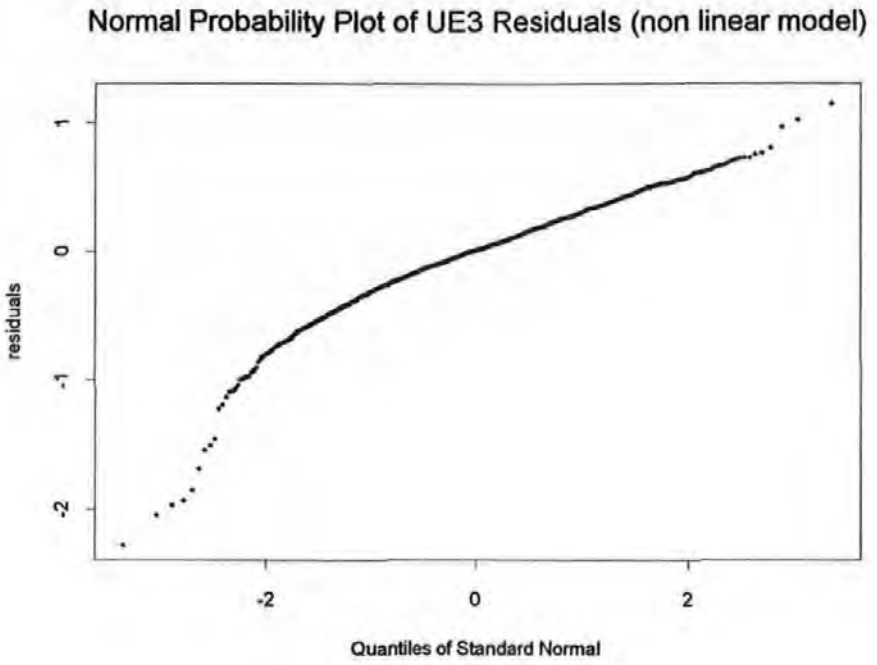

Figure 3.39: Normal probability plot of the residuals of the fitted model *UNLIN1*.

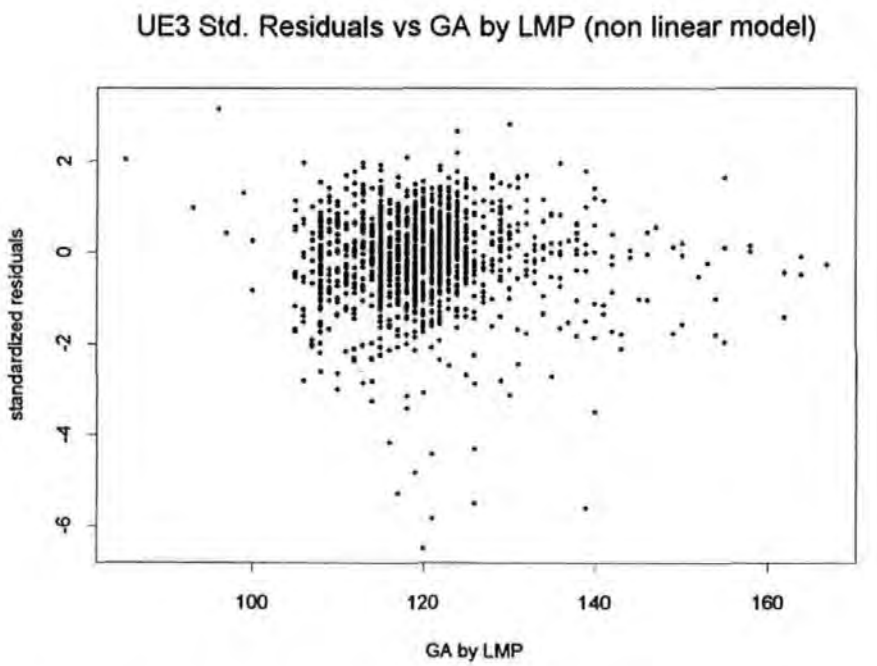

Figure 3.40: Plot of the standardized residuals, r_i, of the fitted model *UNLIN1* against gestational age.

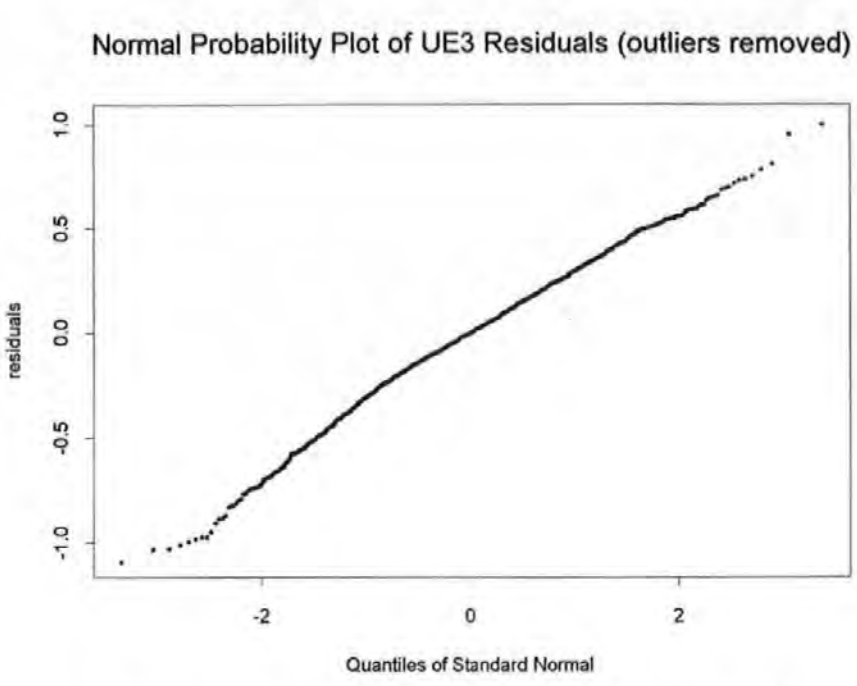

Figure 3.41: Normal probability plot of the residuals of the fitted model *UNLIN2*.

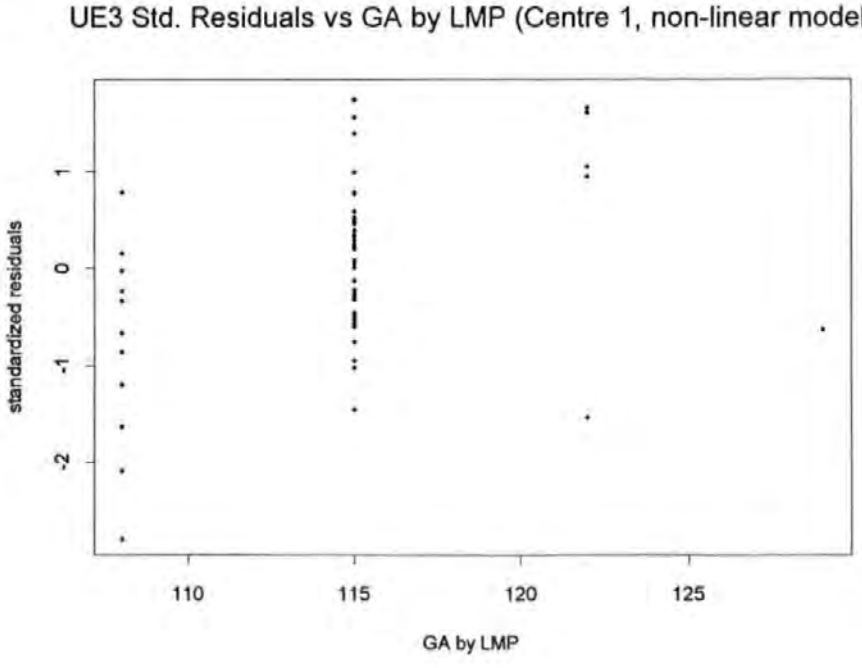

Figure 3.42: Plot of the standardized residuals, r_i, of the fitted model *UNLIN2* for centre 1.

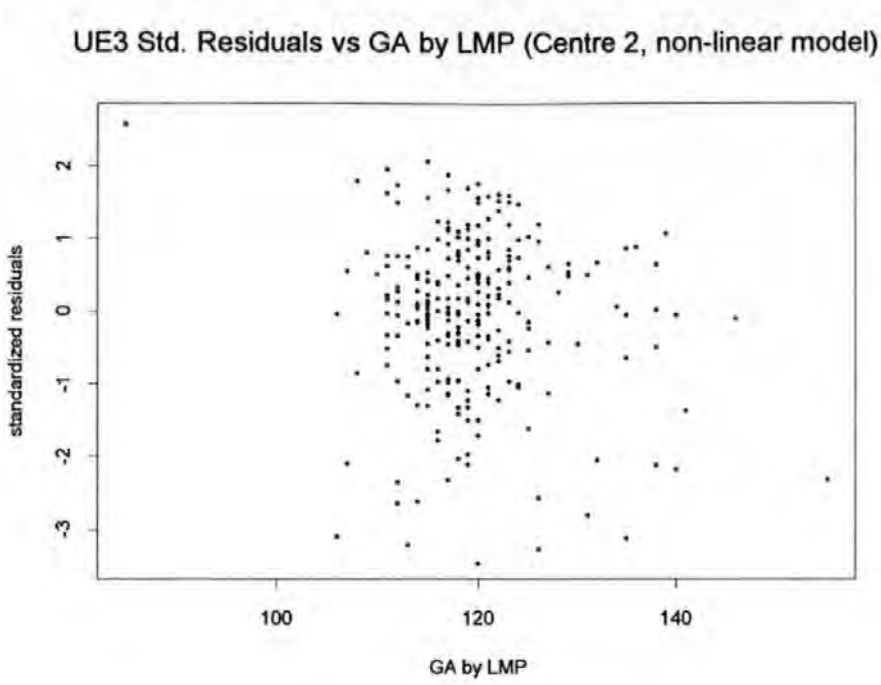

Figure 3.43: Plot of the standardized residuals, r_i, of the fitted model *UNLIN2* for centre 2.

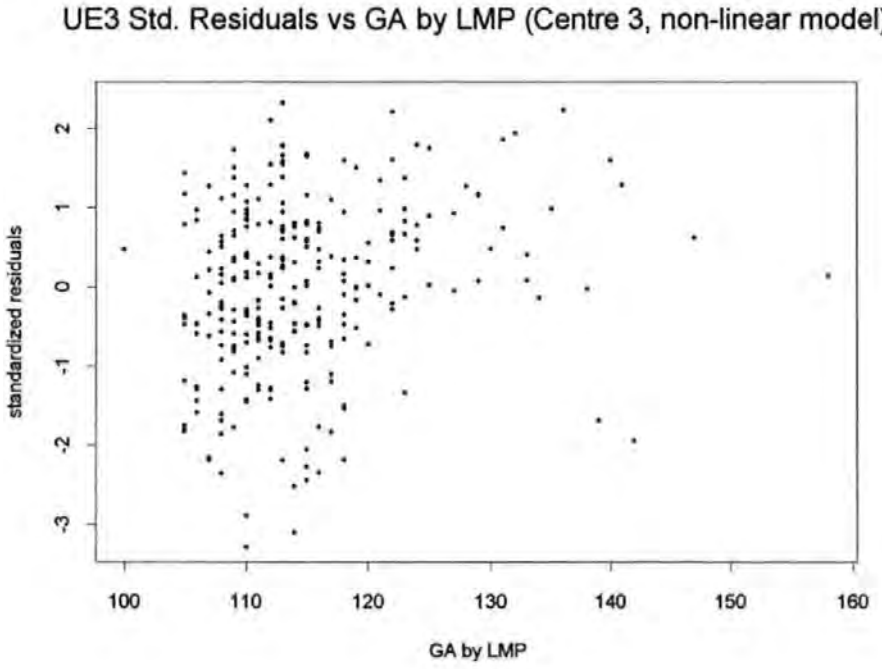

Figure 3.44: Plot of the standardized residuals, r_i, of the fitted model *UNLIN2* for centre 3.

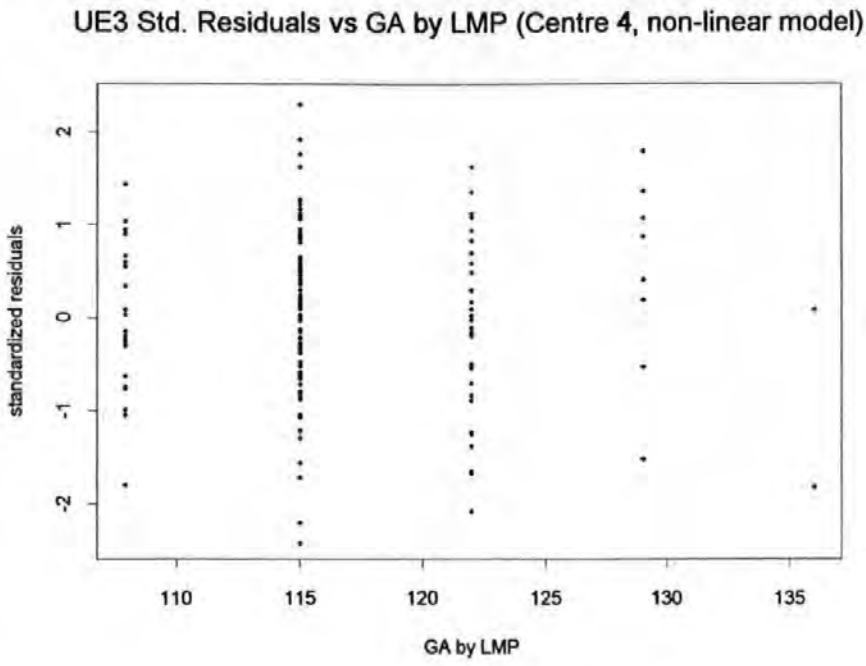

Figure 3.45: Plot of the standardized residuals, r_i, of the fitted model *UNLIN2* for centre 4.

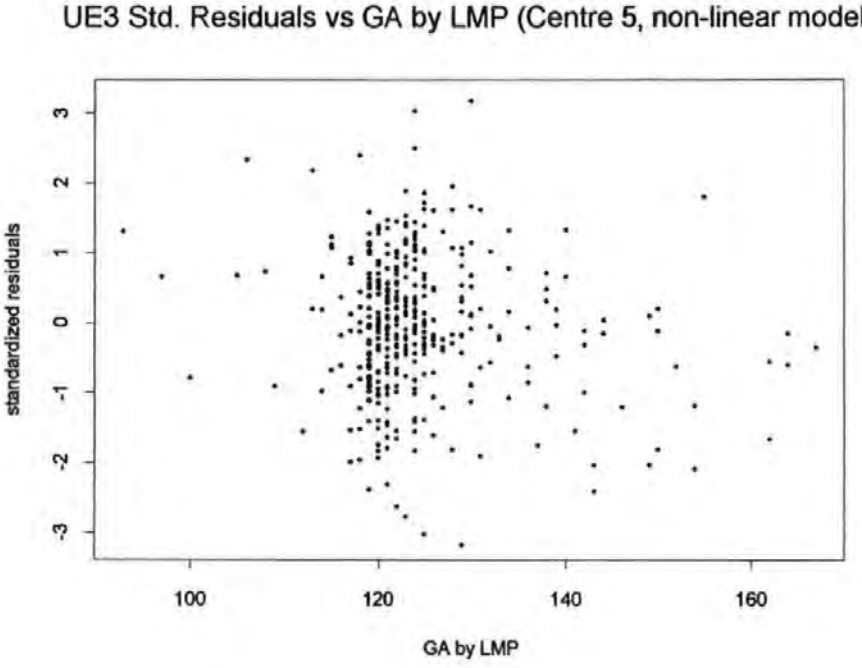

Figure 3.46: Plot of the standardized residuals, r_i, of the fitted model *UNLIN2* for centre 5.

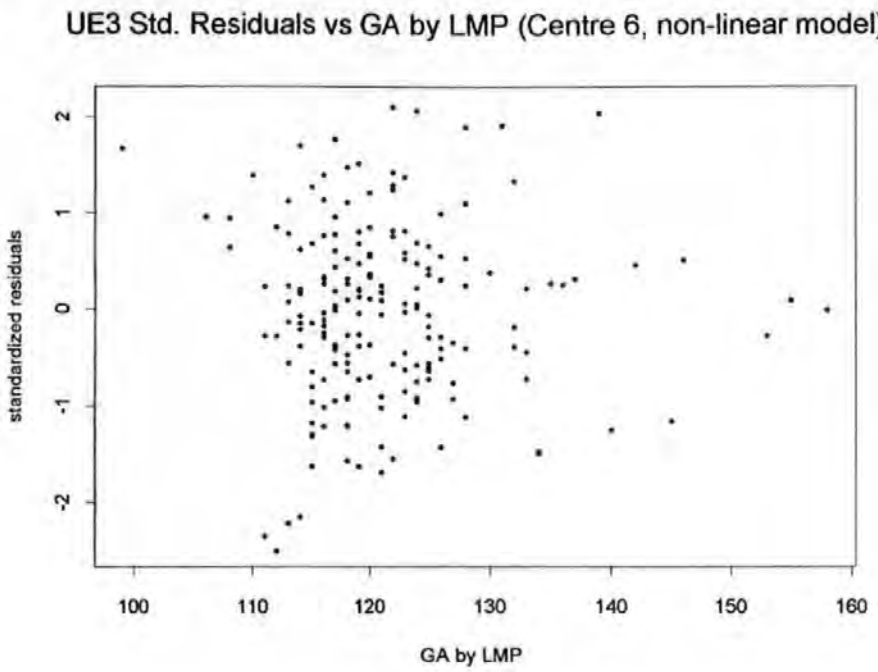

Figure 3.47: Plot of the standardized residuals, r_i, of the fitted model *UNLIN2* for centre 6.

CENTRE No.	1	2	3	4	5	6
SD of residuals	0.2950	0.3412	0.3352	0.2813	0.3096	0.2829
SD estimate $Q_{0.25} - Q_{0.75}$	0.2546	0.2769	0.3294	0.2754	0.2837	0.2686
SD estimate $Q_{0.1} - Q_{0.9}$	0.2601	0.3179	0.3353	0.2817	0.3118	0.2868

Table 3.17: Summary of centre specific residual standard deviations of the fitted model *UNLIN2*. Wald ***et al*** (1993) report the estimated standard deviation of log(MoM) UE3 between the $10^{th} - 90^{th}$ percentiles to be 0.3362 based on LMP dating methods and natural logarithms.

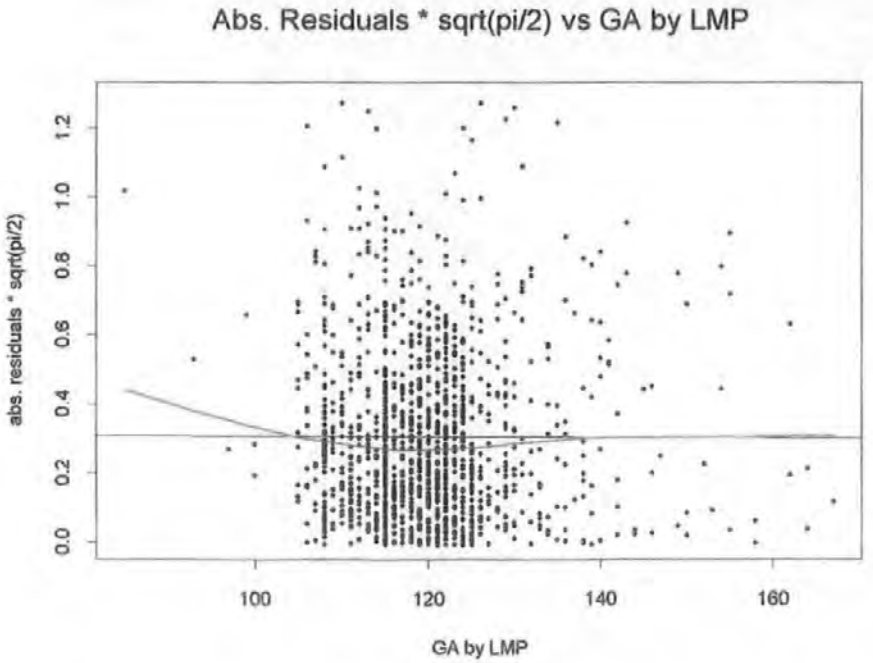

Figure 3.48: Plot of fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ of the model *UNLIN2* against gestational age with lowess trend curve.

Fitted model :- $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i) = 0.3182\text{-}0.0001\ (\text{GA})$

3.6 HCG

Scatter diagrams of the distributions of HCG and log(HCG) against gestational age for the pooled data for unaffected pregnancies are given in Figures 3.49 and 3.50. Again, a random uniform effect is added to the data to make the points distinguishable.

Wald *et al* (1988) model log(HCG) using a constant plus an exponential function of gestational age. This description was first offered by Bogart *et al* (1987). Two forms of models are fitted in this section. In the first model, log(HCG) depends linearly on gestational age and the distribution of log(HCG) about the regression is normal with constant variance. The second model is a non-linear model in which the location of HCG decays exponentially to a positive lower limit and the distribution about the regression is lognormal. This is the model used by Wald *et al* (1988) and (1992). The form of the models fitted in this section are given below

(i) a simple linear regression of log(HCG) on gestational age. (The models fitted of this form are defined as HLIN1 for the full data set, and HLIN2 for the data set with outliers removed).

(ii) a non-linear model in which the location of log(HCG) decays exponentially to a positive lower limit. The distribution about the location is normal. (The models fitted of this form are defined as HNLIN1 for the full data set, and HNLIN2 for the data set with outliers removed).

The analysis of the fit of the two models is given below. The form of these models are illustrated for centre 6 in Figures 3.51-3.52 which show the fitted 10th, 50th and 90th centiles of HCG.

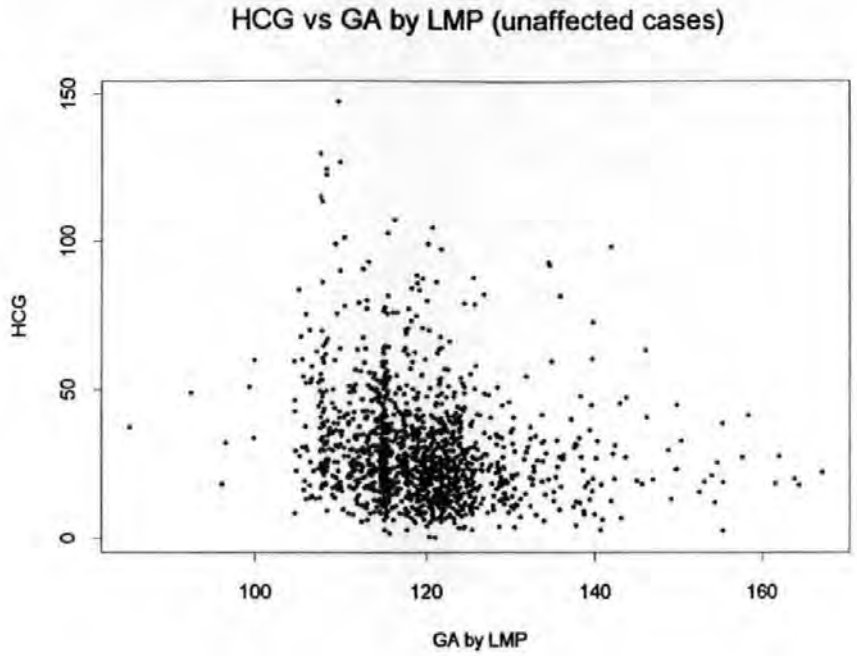

Figure 3.49: Plot of HCG concentrations against gestational age by LMP with a random uniform effect (unaffected pregnancies, n =1356).

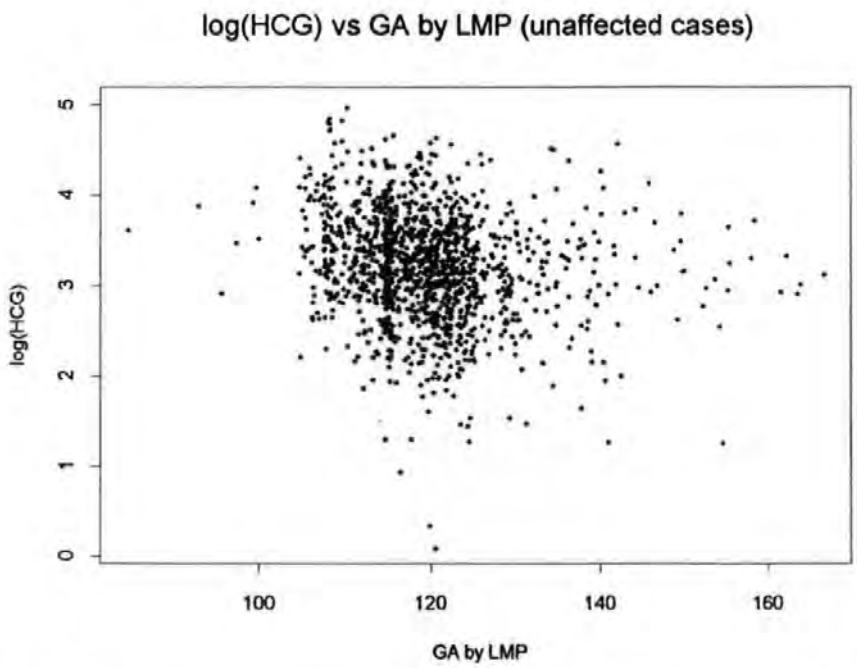

Figure 3.50: Plot of log(HCG) concentrations against gestational age by LMP with a random uniform effect (unaffected pregnancies, n =1356).

(i) A simple linear regression of log(HCG) on gestational age

$$\log(HCG) = \hat{\beta}_0 + \hat{\alpha}_j + \hat{\beta}_1(GA)$$

$\hat{\alpha}_j$ = centre effect for centres $j = 1, 2, 3, 4, 5$, $\hat{\beta}_0 + \hat{\alpha}_j$ = intercepts for centres $j = 1, 2, 3, 4, 5$, and $\hat{\beta}_0$ = intercept for centre 6.

Model HLIN1

Table 3.18 lists the parameter estimates for the model fitted *HLIN1*.

	VALUE	STANDARD ERROR
$\hat{\beta}_0$	4.3621	0.2360
$\hat{\alpha}_1$	-0.0630	0.0415
$\hat{\alpha}_2$	0.0316	0.0179
$\hat{\alpha}_3$	-0.0264	0.0140
$\hat{\alpha}_4$	-0.0258	0.0083
$\hat{\alpha}_5$	-0.0129	0.0076
$\hat{\beta}_1$	-0.0092	0.0020

Table 3.18: Parameter estimates for the fitted regression model *HLIN1*.

The R^2 value for the fitted model is 0.0536 indicating that 5.36% of the variation in the log(HCG) concentrations is explained by the gestational age and centre effects. The effects of gestational age and centre effect are significant ($p < 0.0001$ for both gestational age and centre effects). The normal probability plot of the residuals of the fitted model, in Figure 3.53, illustrates deviations from a Gaussian form in the tails of the distribution. The plot of the standarized residuals, Figure 3.54, against gestational age highlights the presence

of outliers. There is no evidence to suggest the error variance is correlated with gestational age. 8 outliers are identified and removed from the original data.

Model HLIN2

Table 3.19 lists the parameter estimates of the fitted model on the reduced data set.

	VALUE	STANDARD ERROR
$\hat{\beta}_0$	4.3143	0.2261
$\hat{\alpha}_1$	-0.0588	0.0400
$\hat{\alpha}_2$	0.0185	0.0172
$\hat{\alpha}_3$	-0.0332	0.0134
$\hat{\alpha}_4$	-0.0308	0.0080
$\hat{\alpha}_5$	-0.0128	0.0073
$\hat{\beta}_1$	-0.0086	0.0019

Table 3.19: Parameter estimates for the fitted regression model *HLIN2*.

The R^2 value of the fitted model *HLIN2* is 0.0575 which is only a slight improvement on the R^2 value for model *HLIN1*. The outliers have little effect on the fitted model. Figure 3.55 shows the normal probability plot of the residuals of the fitted model. The plot is more linear in the tails than the plot in Figure 3.53. Figures 3.56-3.61 plot the centre specific residuals on length of gestation. The standard deviations of the residuals estimated from the product moment formula and robust estimates calculated between the $25^{th} - 75^{th}$ and $10^{th} - 90^{th}$ percentiles are tabulated, Table 3.22. The estimates are generally lower than the estimate reported by Wald *et al* (1992). Figure 3.62 illustrates the distribution of the absolute residuals over the range of gestational age. The slope parameter

of the fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ against gestational age is not significant, confirming that the error variance can be assumed constant over the range of gestation ($p > 0.05$).

(ii) A non-linear regression of log(HCG) on gestational age in which the trend in HCG depends linearly on gestational age but the distribution about the trend is lognormal

$$\log(HCG) = \hat{\alpha}_j + \log[\hat{\beta}_0 + \beta_1 \exp(-\hat{\beta}_2(GA - 115)/10)]$$

$\hat{\alpha}_j$ = centre effect for centres $j = 1, 2, 3, 4, 5$.

Model HNLIN1

Table 3.20 lists the parameter estimates for the model fitted *HNLIN1*.

	VALUE	STANDARD ERROR
$\hat{\alpha}_1$	0.1415	0.0855
$\hat{\alpha}_2$	0.0314	0.0535
$\hat{\alpha}_3$	0.1548	0.0550
$\hat{\alpha}_4$	0.0116	0.0627
$\hat{\alpha}_5$	-0.0215	0.0496
$\hat{\beta}_0$	20.2058	1.7736
$\hat{\beta}_1$	5.6556	1.9233
$\hat{\beta}_2$	0.7541	0.2283

Table 3.20: Parameter estimates of the fitted regression model *HNLIN1*.

The R^2 value for the fitted model is 0.0633 indicating that 6.33% of the variation in the log(HCG) concentrations is explained by the gestational age and centre effects. The R^2 value is slightly greater than those given for the models previously fitted. The effects of gestational age and centre effect are significant ($p < 0.0001$ for both gestational age and centre effects). The normal probability plot of the residuals of the fitted model, in Figure 3.63, illustrates deviations from a Gaussian form in the tails of the distribution. The plot resembles the plot in Figure 3.53. The plot of the standarized residuals against gestational age, Figure 3.64, highlights the presence of outliers. There is no evidence to suggest the error variance is correlated with gestational age. 9 outliers are identified and removed from the original data.

Model HNLIN2

Table 3.21 lists the parameter estimates of the model for the reduced data set.

$\hat{\alpha}_1$	0.1657	0.0821
$\hat{\alpha}_2$	0.0612	0.0514
$\hat{\alpha}_3$	0.1442	0.0527
$\hat{\alpha}_4$	-0.0015	0.0601
$\hat{\alpha}_5$	-0.0432	0.0476
$\hat{\beta}_0$	21.1149	1.6664
$\hat{\beta}_1$	5.0452	1.8021
$\hat{\beta}_2$	0.7884	0.2361

Table 3.21: Parameter estimates of fitted regression model *HNLIN2*.

The R^2 value for the fitted model *HNLIN2* is 0.0648 which is a slight improvement on the R^2 value for the fitted model *HNLIN1*. The outliers have little impact on the fitted model. The normal probability plot shown in Figure 3.65 is more linear in the tails. The centre specific standardized residuals are illustrated in Figures 3.66-3.71. There is no evidence in the residual plots to suggest the error variance is correlated with gestational age. The standard deviations of the residuals estimated from the product moment formula and robust estimates calculated between the $25^{th} - 75^{th}$ and $10^{th} - 90^{th}$ percentiles are tabulated, Table 3.23, and are generally lower than the estimated reported by Wald *et al* (1992). Figure 3.72 graphs the fitted model of the absolute residuals on length of gestation. The slope parameter of the fitted model is not significant to the regression.

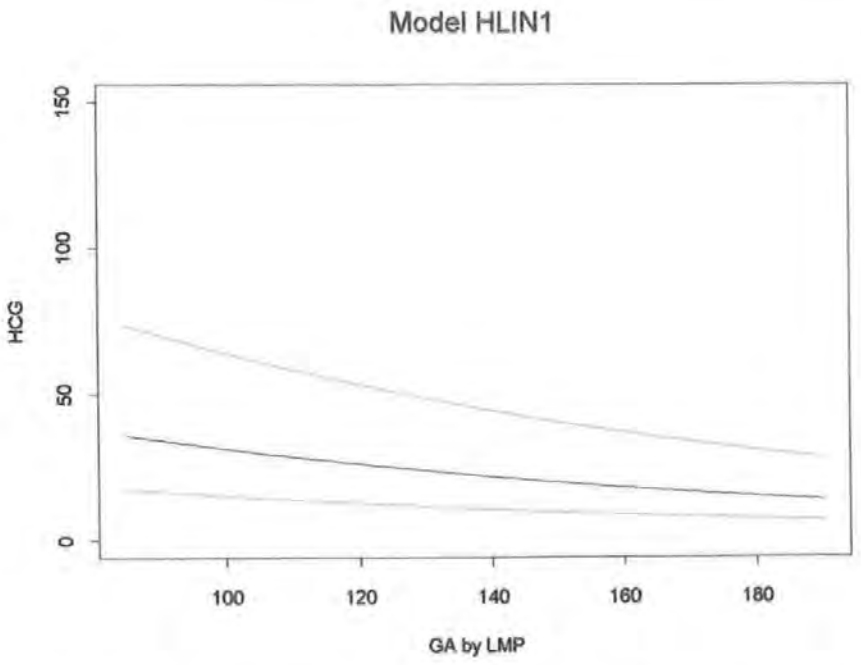

Figure 3.51: The fitted 10th, 50th and 90th centiles of HCG derived from model *HLIN1* for centre 6.

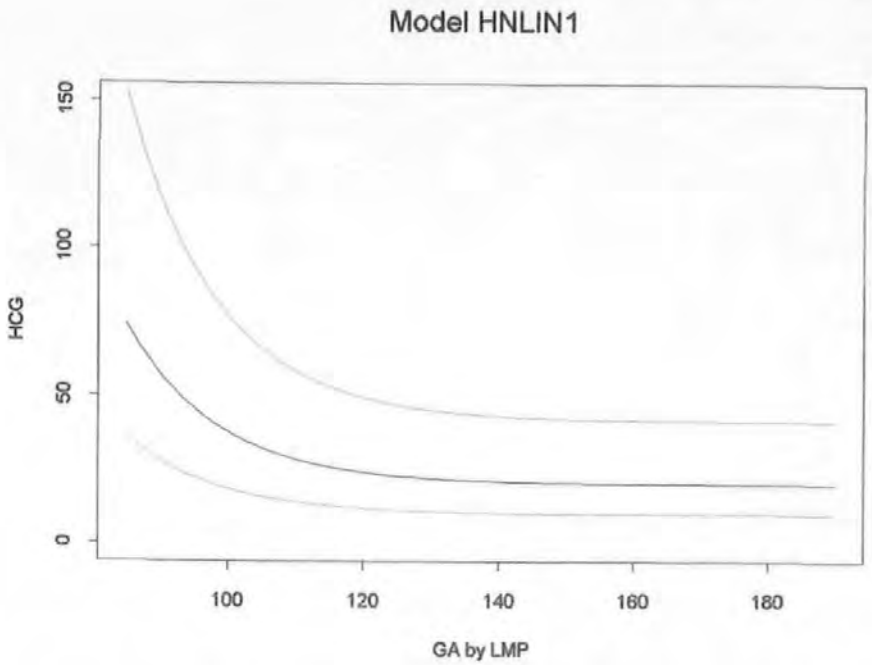

Figure 3.52: The fitted 10th, 50th and 90th centiles of HCG derived from model *HNLIN1* for centre 6.

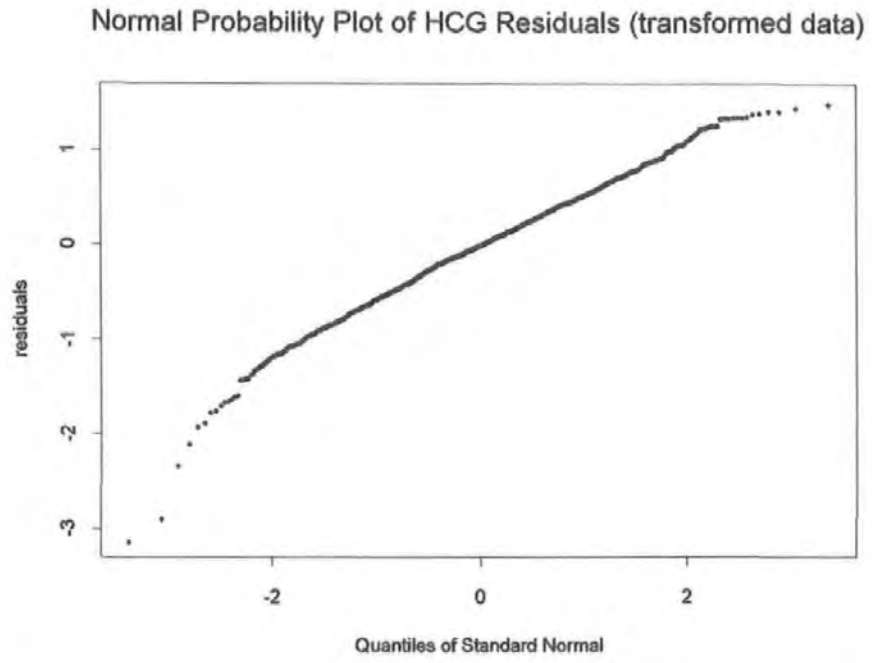

Figure 3.53: Normal probability plot of the residuals of the fitted model *HLIN1*.

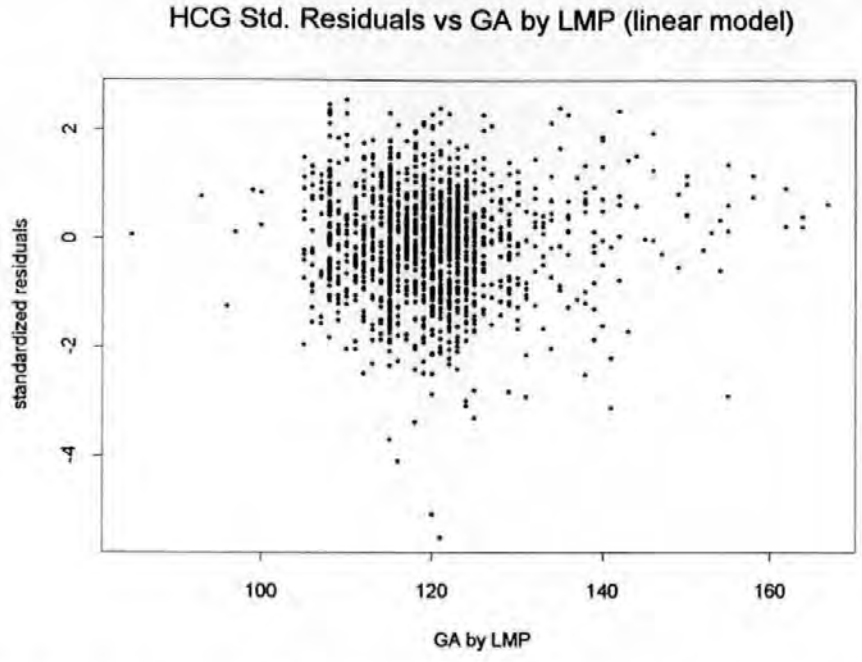

Figure 3.54: Plot of residuals of the fitted model *HLIN1* against gestational age.

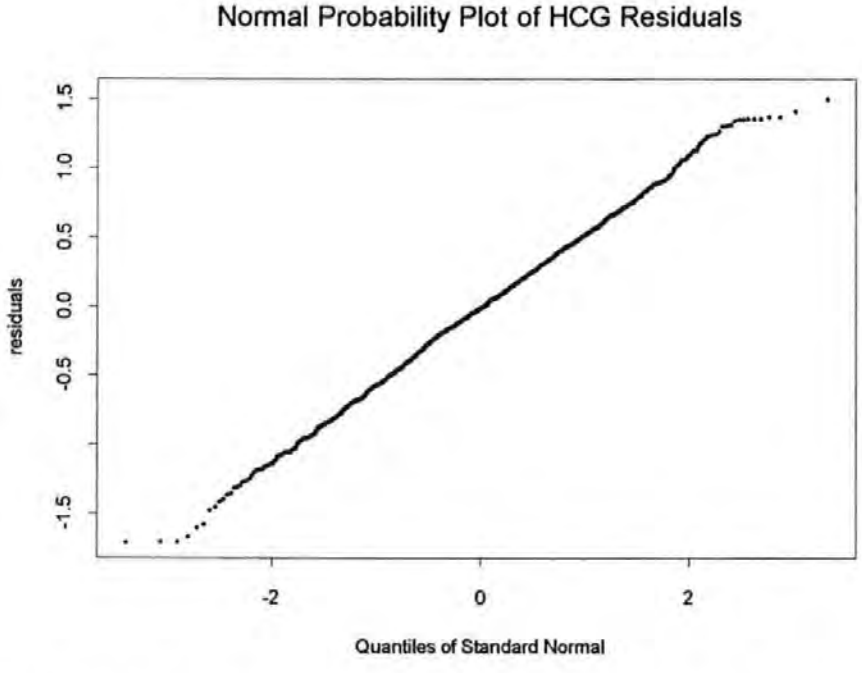

Figure 3.55: Normal probability plot of the residuals of the fitted model *HLIN2*.

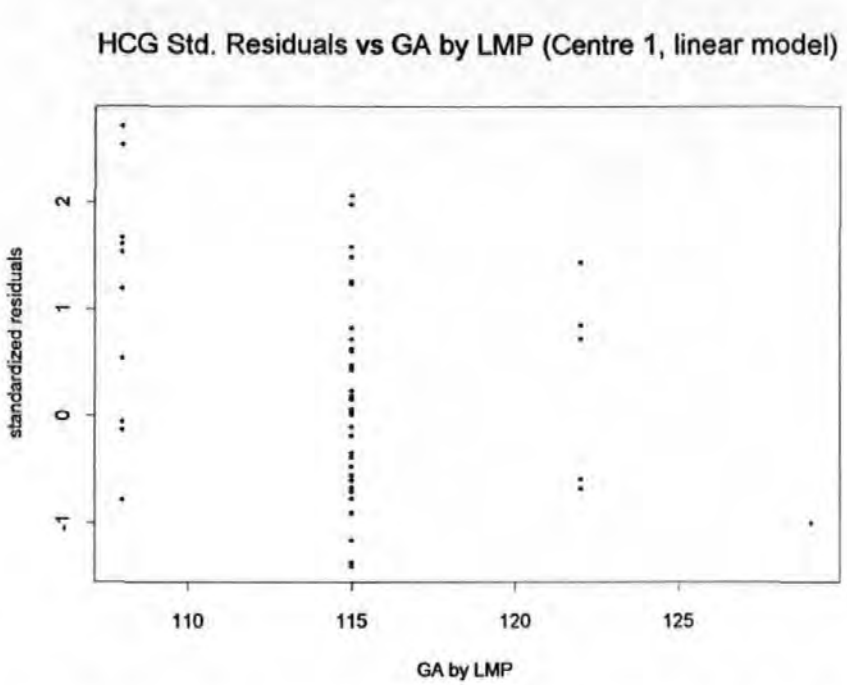

Figure 3.56: Plot of the standardized residuals, r_i, of the fitted model *HLIN2* for centre 1.

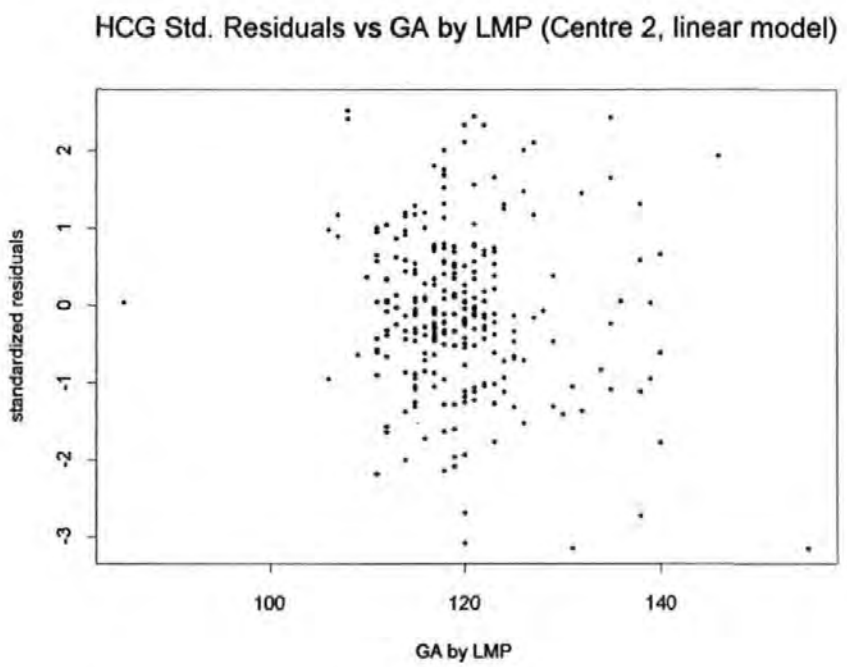

Figure 3.57: Plot of the standardized residuals, r_i, of the fitted model *HLIN2* for centre 2.

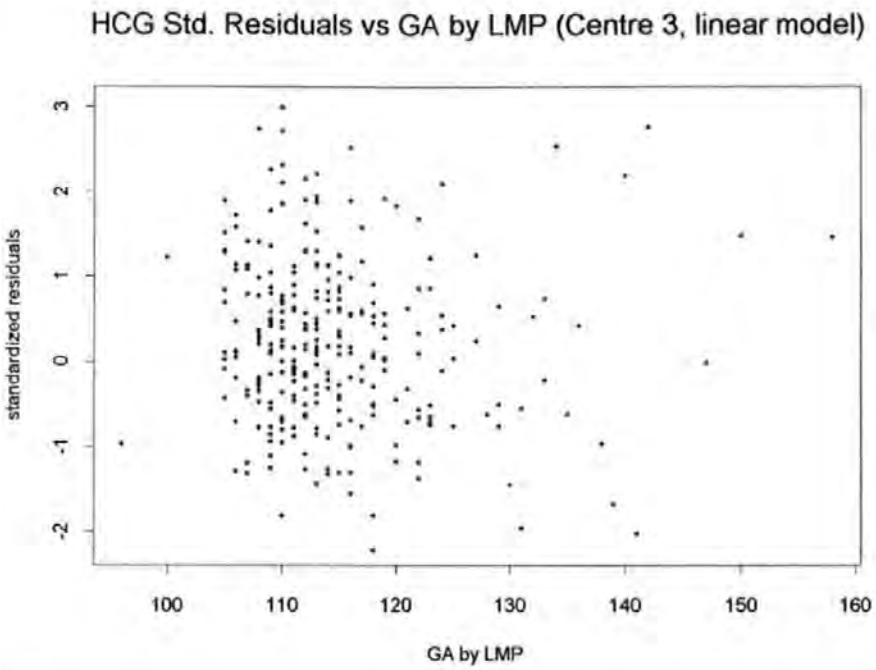

Figure 3.58: Plot of the standardized residuals, r_i, of the fitted model *HLIN2* for centre 3.

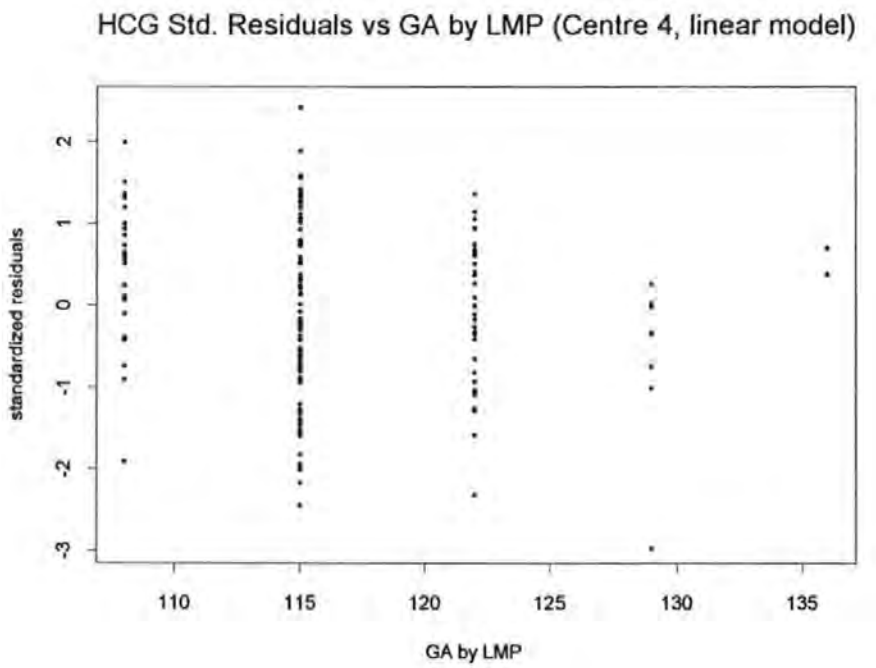

Figure 3.59: Plot of the standardized residuals, r_i, of the fitted model *HLIN2* for centre 4.

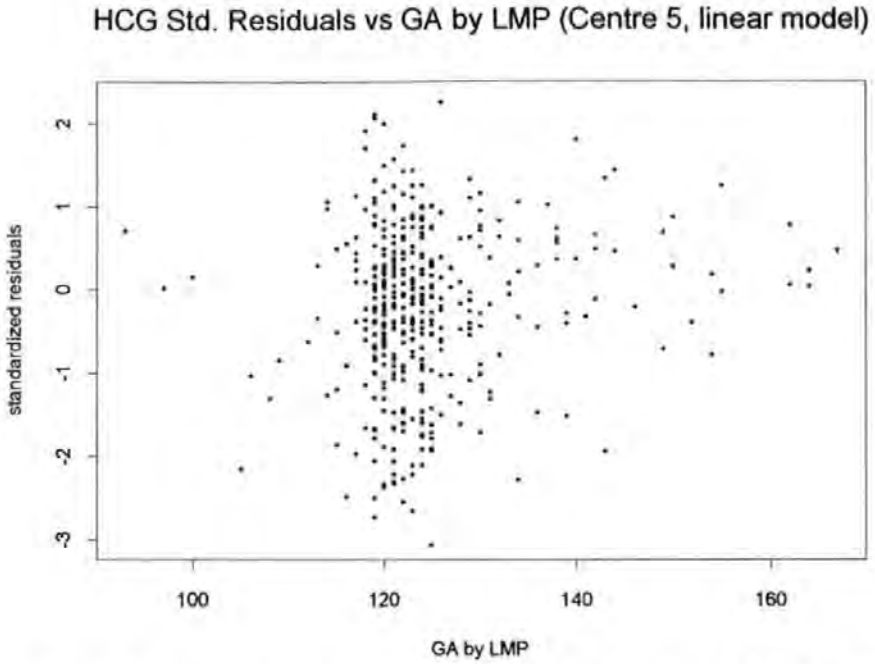

Figure 3.60: Plot of the standardized residuals, r_i, of the fitted model *HLIN2* for centre 5.

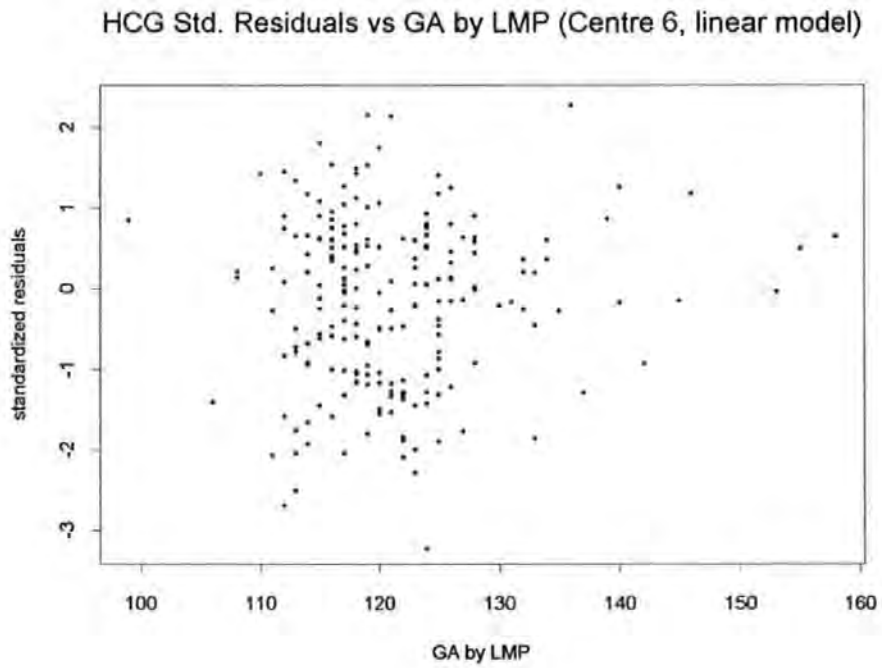

Figure 3.61: Plot of the standardized residuals, r_i, of the fitted model *HLIN2* for centre 6.

CENTRE No.	1	2	3	4	5	6
SD of residuals	0.5426	0.5726	0.5298	0.5620	0.5335	0.5684
SD estimate $Q_{0.25}-Q_{0.75}$	0.5489	0.4916	0.5055	0.6088	0.5144	0.6371
SD estimate $Q_{0.1}-Q_{0.9}$	0.5339	0.5408	0.5250	0.5659	0.5562	0.5609

Table 3.22: Summary of centre specific residual standard deviations of the fitted model *HLIN2*. Wald *et al* (1992) report the estimated standard deviation of log(MoM) HCG between the $10^{th}-90^{th}$ percentiles to be 0.5720 based on LMP dating methods and natural logarithms.

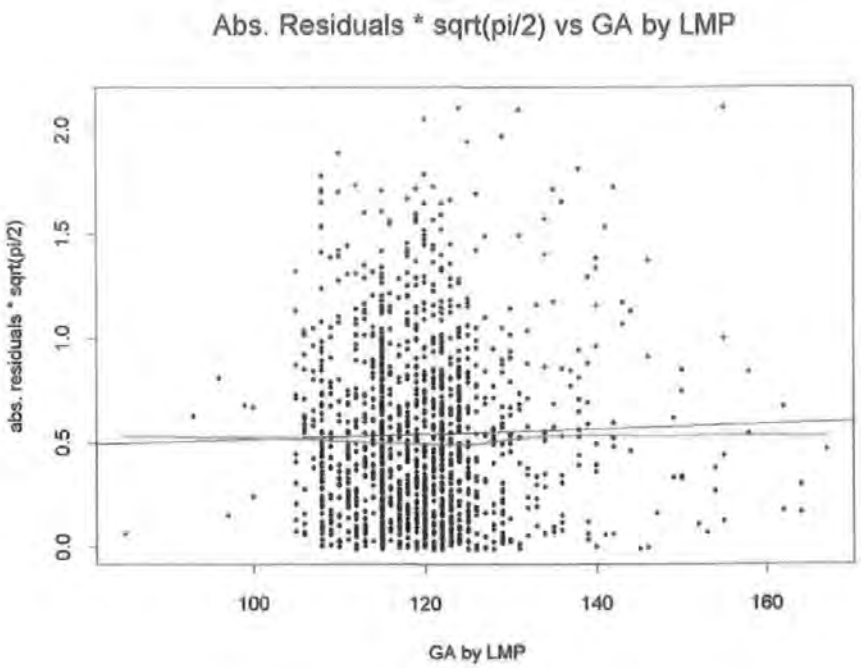

Figure 3.62: Plot of fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ of the model *HLIN2* against gestational age with lowess trend curve.

Fitted model :- $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i) = 0.3965 + 0.0012(\text{GA})$

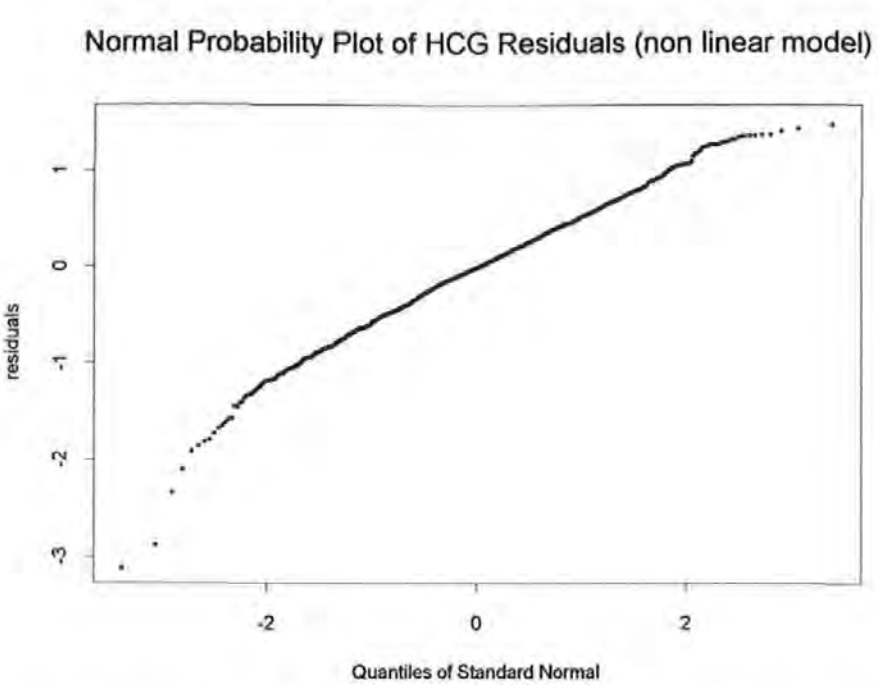

Figure 3.63: Normal probability plot of the residuals of the fitted model *HNLIN1*.

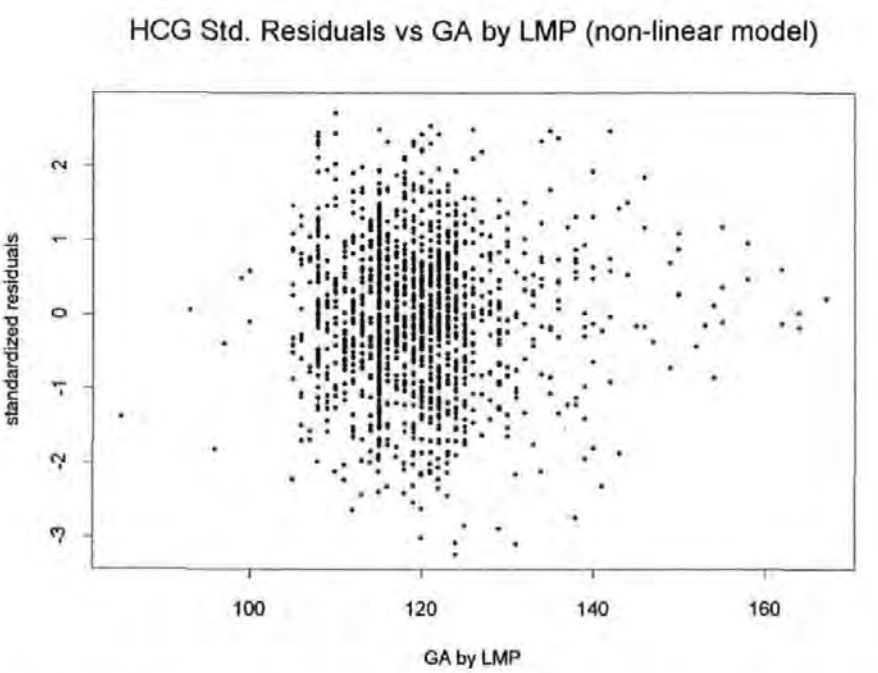

Figure 3.64: Plot of the standardized residuals, r_i, of the fitted model *HNLIN1* against gestational age.

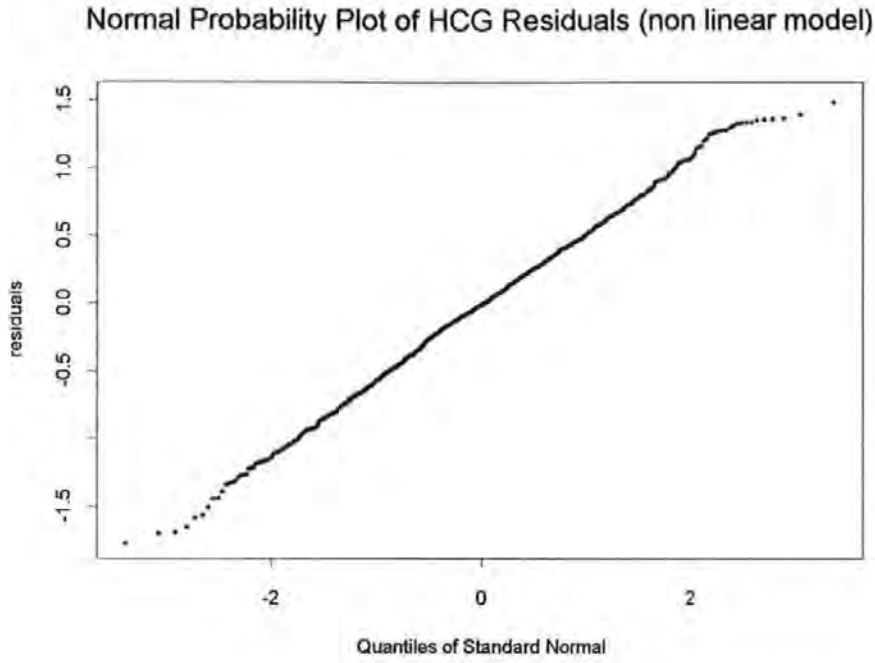

Figure 3.65: Normal probability plot of residuals of fitted model *HNLIN2*.

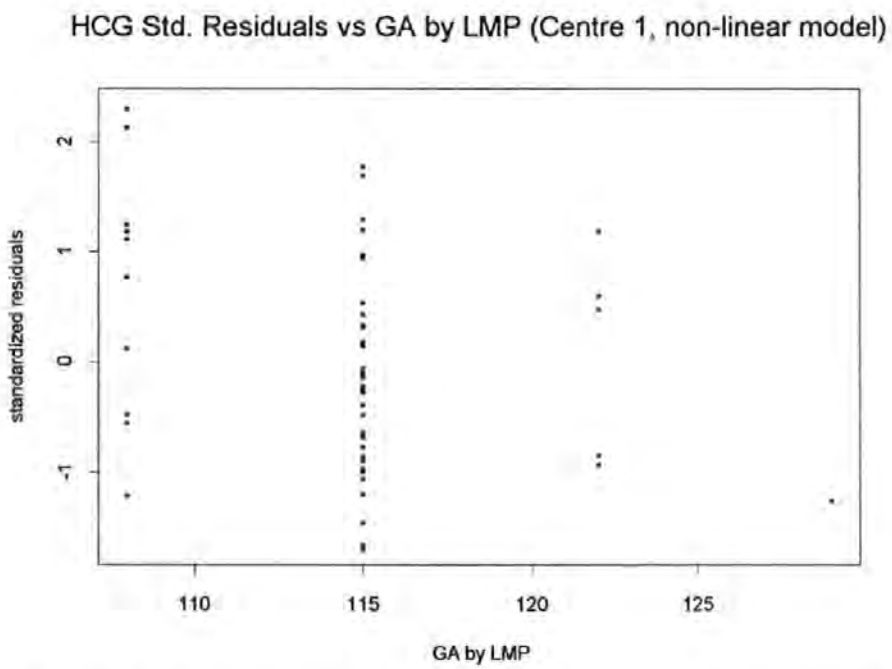

Figure 3.66: Plot of the standardized residuals, r_i, of the fitted model *HNLIN2* for centre 1.

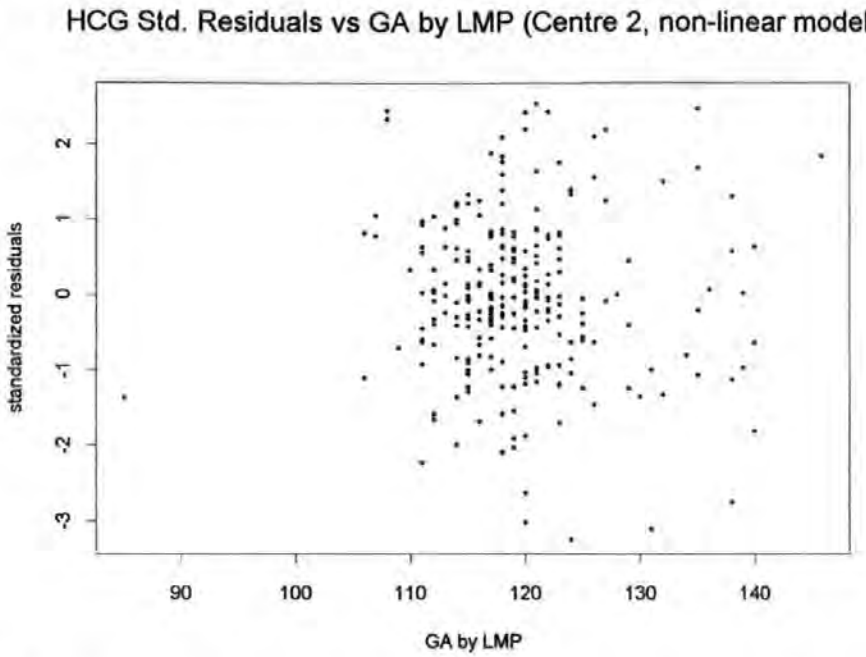

Figure 3.67: Plot of the standardized residuals, r_i, of the fitted model *HNLIN2* for centre 2.

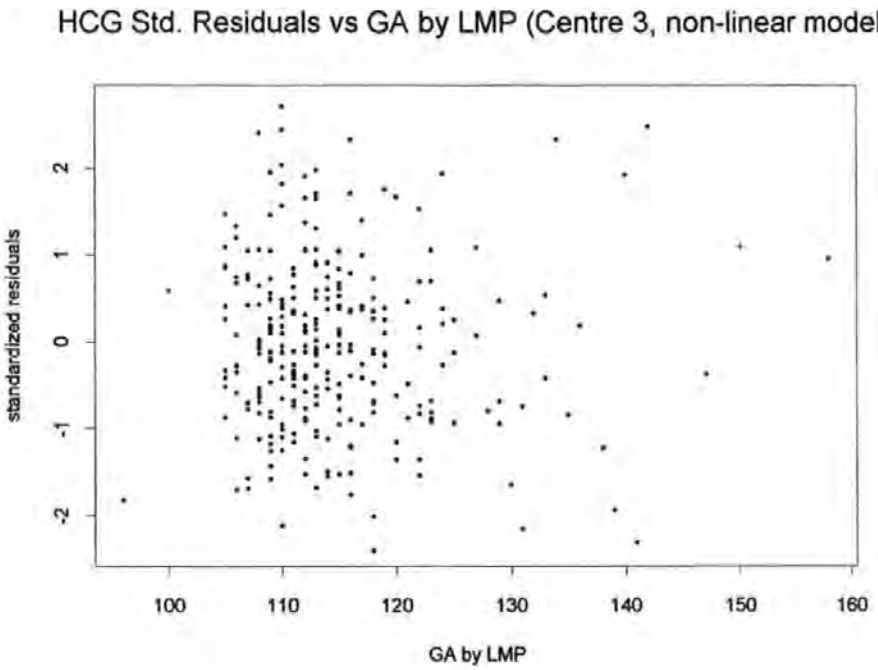

Figure 3.68: Plot of the standardized residuals, r_i, of the fitted model *HNLIN2* for centre 3.

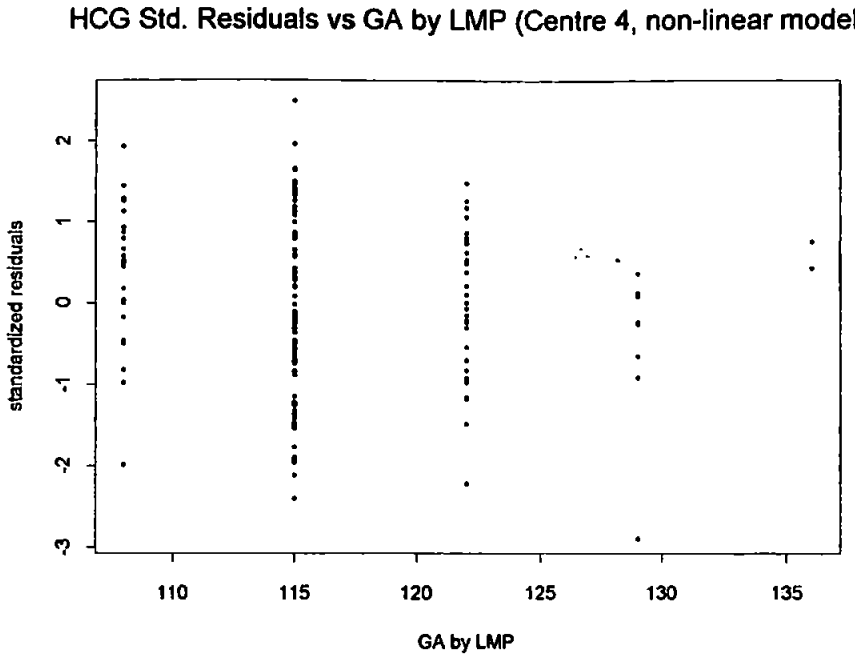

Figure 3.69: Plot of the standardized residuals, r_i, of the fitted model *HNLIN2* for centre 4.

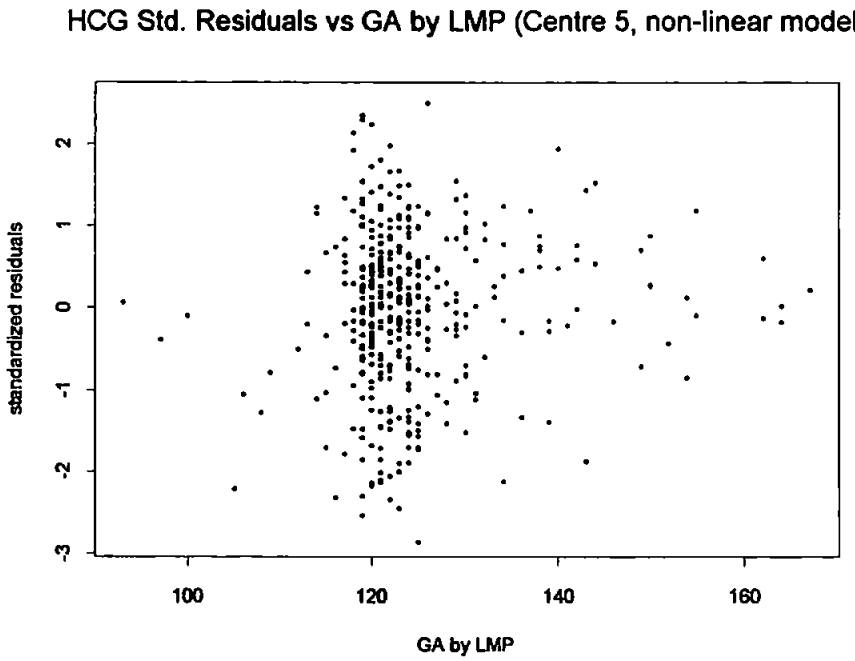

Figure 3.70: Plot of the standardized residuals, r_i, of the fitted model *HNLIN2* for centre 5.

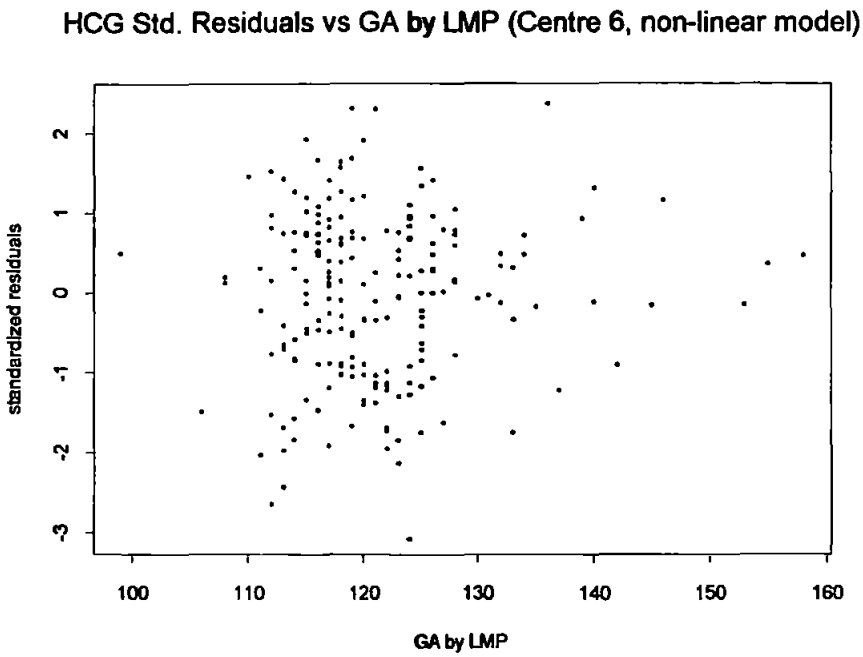

Figure 3.71: Plot of the standardized residuals, r_i, of the fitted model *HNLIN2* for centre 6.

CENTRE No.	1	2	3	4	5	6
SD of residuals	0.5318	0.5734	0.5271	0.5310	0.5296	0.5659
SD estimate $Q_{0.25} - Q_{0.75}$	0.5573	0.5076	0.4920	0.5016	0.4945	0.6423
SD estimate $Q_{0.1} - Q_{0.9}$	0.5164	0.5404	0.4857	0.5521	0.5528	0.5627

Table 3.23: Summary of centre specific residual standard deviations of the fitted model *HNLIN2*. Wald *et al* (1992) report the estimated standard deviation of log(MoM) HCG between the $10^{th} - 90^{th}$ percentiles to be 0.5720 based on LMP dating methods and natural logarithms.

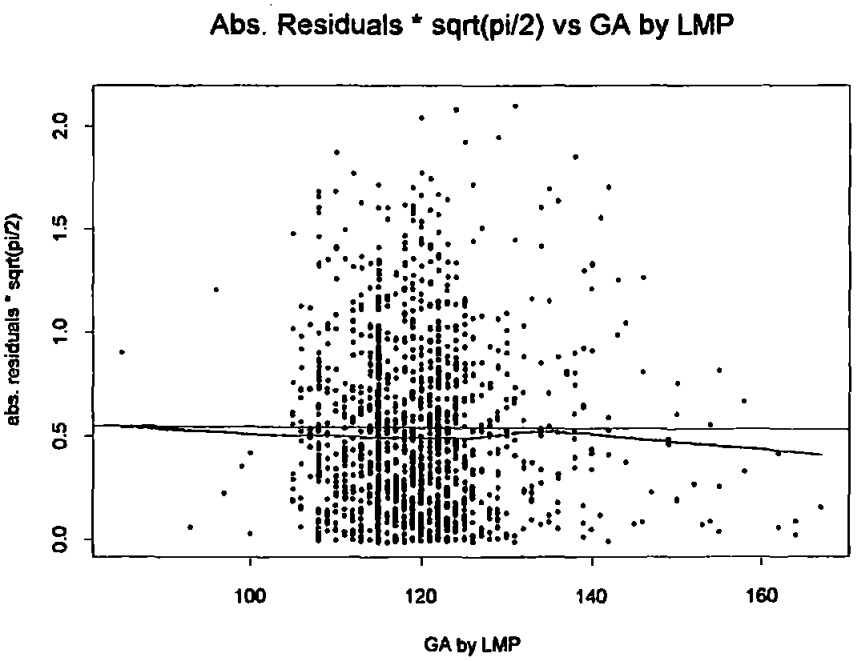

Figure 3.72: Plot of fitted regression model of $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i)$ of the model *HNLIN2* against gestational age with lowess trend curve.

Fitted model :- $\sqrt{\frac{\pi}{2}} \times \text{abs}(e_i) = 0.5579\text{-}0.0001(\text{GA})$

3.7 Conclusion

The effect of centre is significant for all analytes. This confirms the dangers of pooling data from different centres without adjusting for the centre effects. The effect of outliers on the fitted models is negligible as there is no real impact on the regression coefficients. However, they have a large influence on the shape of the probability plots and the standard deviations and correlations.

The analysis has demonstrated that linear and non-linear least squares are efficient methods of modelling analyte concentrations. The residuals of the fitted regression models are equivalent to the log(MoM) values that are currently used in clinical trials analysis. This approach avoids the need to standardize the trials data as Gaussian densities can directly be fitted to these residuals and the ratio of the fitted densities provides a likelihood function

which can be used in the discriminant analysis. Grouping the data into gestational weeks is unnecessary, therefore the fitted models more precisely describe the true relationship between concentration values and gestational dates.

The loglinear and non-linear models that are fitted in this chapter confirm the assumption of homogeneity. The R^2 value for the linear models of log(UE3) on gestational age and log(HCG) on gestational age are similar to the values given with the non-linear models. There is little difference between the fit of the loglinear and the non-linear models. However, the non-linear model may provide a better description of the trend of the distribution of log(HCG) over greater gestations. It would be informative to compare the use of these models with data having gestational ages recorded by sonography since this method of dating has less error.

Residuals of the linear model *ALIN2* and of the non-linear models *UNLIN2* and *HNLIN2* are used for the analyses conducted in chapters four and six of this thesis, however, the choice between the use of the linear and the non-linear models is somewhat arbitrary.

Chapter 4

Detection rates and false positive rates

4.1 Introduction

The performance of the screening algorithms used by Wald *et al* (1988) and others, see for example Crossley *et al* (1993), is assessed in terms of the detection rate and false positive rate which are the respective proportions of pregnancies that are correctly and incorrectly screened positive. The detection rates are used to compare the accuracy of different assay kits and to compare the effective use of analytes in screening. The estimated detection rates and false positive rates reported in the literature are generally provided as point estimates. Little or no attention is paid to sampling error.

Failure to consider the sampling error of such estimates has led to debates over the differences between published detection rates and the relative utility of the various combinations of analytes - in particular the benefit of adding UE3 to a combination of maternal age, AFP and HCG. More recent studies have indicated that free β HCG, (total HCG minus the β subunit), is a more useful marker than intact HCG (Macri *et al* (1990)).

The estimated performance measures are also prone to bias from different sources. It is well known that the disciminant rule is optimized on a design set. Therefore, the performance statistics that are estimated from this design set do not reflect the values that would be attained if the rule was applied to a randomly selected data set from the same distribution. Moreover, it is well established that Down's syndrome pregnancies are

associated with older women and since data for affected pregnancies are relatively rare it is often difficult to draw representative samples.

This chapter addresses the issue of sampling error and investigates the possible bias in the estimated error rates. Section 4.2 gives a brief review of the debate over the differences between the published detection rates. Section 4.3 sets out the algorithm used to calculate the false positive rate and the detection rate. Section 4.4 reviews the application of these methods of estimation to screening. In section 4.5, Monte Carlo methods of simulation are used to simulate a number of repeats of the clinical trials data given in Chapter three and of the clinical trial conducted by Wald *et al* (1992). The detection rates are calculated for each simulated study and the sampling errors of the detection rates are estimated. For reference, this method is also used in Wright *et al* (1993a). Section 4.6 discusses parametric and nonparametric methods of estimating discriminant rules and defines error rates of allocation. Some methods of computing parametric and nonparametric bias corrected error rates are reviewed. Section 4.7 applies the nonparametric method of bootstrapping to assess the extent of the bias in the detection rates and false positive rates. The results of the studies are summarized in section 4.8.

4.2 Background

The seminal article published by Wald *et al* (1988) reports a detection rate of 60%, with a false positive rate of 5% using the three analytes AFP, HCG and UE3 and the maternal age distribution of pregnancies for England and Wales in 1981-1985. The data consists of 77 affected pregnancies and 385 controls. The paper concludes that the addition of UE3 has the advantage of increasing the detection rate and decreasing the false positive rate. MacDonald *et al* (1991) use the same combination of analytes and report a detection

rate of 60% at a false positive rate of 7.7% with data based on a 'representative screened population' and a sample of 18 cases. Macri *et al* (1990) find no basis to support the hypothesis that low levels of UE3, in 41 affected pregnancies, can be used to detect fetal Down's syndrome.

Almost as much controversy surrounds the advantage of screening with free β HCG rather than Intact HCG. Spencer (1991) compares the efficiency of intact HCG and free β HCG when combined with AFP and UE3 at a false positive rate of 5.9%. The study, which includes 29 affected pregnancies, gives rise to a detection rate of 52% with the intact molecule and 66% with the free β subunit. Macri *et al* (1990) also support the use of free β HCG. The relatively recent introduction of new biochemical predictors of Down's syndrome, such as PAPP-A and inhibin, has added to the argument.

The debate over the relative use of analytes as predictors of Down's syndrome disregards sampling error. It is plausible that the reported performance statistics show little difference when considered along with their standard errors. The possibility of bias in the estimated detection rates and false positive rates must also be considered before comparisons can be made between the reported estimates.

The problem of selecting screening variables is essentially a problem of variable selection. Variable selection and error rate estimation are central to discriminant analysis and a large amount of relevant work is available in the statistical literature (Habbema *et al* (1978), Lachenbruch (1975), McLachlan (1992)).

4.3 A simple algorithm for establishing the detection rate and false positive rate of the screen

The estimated performance of the discriminant rule is assessed in terms of the detection rates and false positive rates. A discriminant rule of the form given in equation (1.3) uses estimated risks to classify pregnancies as either unaffected or Down's syndrome. The proportions of women with maternal age m that are classified as unaffected and Down's syndrome are calculated. The detection rate is given by summing those proportions of affected pregnancies that are correctly classified as having the abnormality over the range of m. Conversely, the false positive rate is given by summing those proportions of unaffected pregnancies that are incorrectly classified. Essentially, the summations are formed by numerically integrating these proportions of pregnancies over the appropriate maternal age distribution.

4.3.1 Detection rate

The detection rate of the screen, $p(screen+ve/D)$, gives the proportions of pregnancies correctly classified as Down's syndrome and is defined as

$$p(screen+ve/D)=\sum_{m} p(screen+ve,m/D) \qquad (4.1)$$

$$=\sum_{m} p(screen+ve/m,D)p(m/D)$$

$$=\frac{\sum_{m} p(screen+ve/m,D)p(D/m)p(m)}{p(D)} \qquad (4.2)$$

The probability of a Down's syndrome pregnancy can be written as

$$p(D) = \sum_m p(D \cap m) = \sum_m p(D/m)p(m)$$

equation (4.2) therefore becomes

$$p(screen+ve/D) = \frac{\sum_m p(screen+ve/m,\ D)\ P(D/m)p(m)}{\sum_m p(D/m)p(m)} \qquad (4.3)$$

4.3.2 False Positive Rate

The false positive rate, $p(screen+ve/N)$, gives the proportion of unaffected pregnancies incorrectly screened positive.

$$p(screen+ve/N) = \sum_m p(screen+ve, m/N)$$

Simply, by replacing D by N in equation (4.2) and equation (4.3) gives

$$p(screen+ve/N) = \frac{\sum_m p(screen+ve/m,\ N)\ p(N/m)p(m)}{\sum_m p(N/m)p(m)} \qquad (4.4)$$

4.3.3 Age specific performance levels

Maternal age specific performance statistics can be estimated by numerically integrating the class conditional probability density functions of transformed MoM analyte values over the regions where pregnancies are screened postive.

The detection rate that is specific to women of age m is given by

$$p(screen\ +ve/D,m) = \int_{W_m} p(\underline{x}/D,m)d\underline{x} \qquad (4.5)$$

where W_m is the region where pregnancies of women aged m are screened positive

The false positive rate that is specific to women of age m is, therefore, given by

$$p(screen\ +ve/N,m) = \int_{W_m} p(\underline{x}/N,m)d\underline{x} \qquad (4.6)$$

4.4 Current methodology for estimating the detection rates and false positive rates

Several methods have been used to estimate detection rates and false positive rates (Wald *et al* (1988), Davies *et al* (1991)). These methods are discussed below.

Raw proportions

The simplest approach calculates the number of Down's syndrome pregnancies and unaffected pregnancies, from a test set, whose risks are higher than a selected cut-off value,

and expresses each as a proportion of the total number of pregnancies with the same outcome. Bishop (1994) shows that this method of estimating the detection rate is not the most efficient. This approach disregards the bias caused by screening with unrepresentative maternal age distributions. Samples that over represent more mature women will optimistically bias the performance measures.

Weighted age specific proportions

A second method calculates estimates of the false positive rate and detection rate by directly modelling the clinical trials data. The approach corrects for unrepresentative maternal age distributions and has the advantage of utilizing each independent analyte value. Likelihood ratios are derived from the fitted Gaussian densities of appropriately transformed MoM values that are recorded from the trials data. The age specific risk of Down's syndrome given by Cuckle *et al* (1987) is applied to a standardized age distribution to estimate the expected proportion of unaffected and Down's syndrome births. The value of the likelihood ratio, $\lambda(\underline{x})$, required to modify the age specific risk to below a selected cut off value, c, for each single year of maternal age is determined. A pregnancy is screened positive if

$$\lambda(\underline{x}) \quad < \quad \frac{c\, p(D/m)}{p(N/m)} \qquad \forall m \in M, \underline{x} \in \underline{X} \tag{4.7}$$

The sample proportions of pregnancies from the trials data screened positive for each maternal age and outcome are computed. The standardized age distribution is then used to estimate the age specific proportions of unaffected and Down's syndrome pregnancies that are screened positive which are given by

$$\hat{p}(screen+ve/N,m) = proportion\ of\ unaffected\ cases\ for\ which\ \lambda(\underline{x}) < c\,\frac{p(D/m)}{p(N/m)} \tag{4.8}$$

and

$$\hat{p}(screen+ve/D,m) = proportion\ of\ affected\ cases\ for\ which\ \lambda(\underline{x}) < c\,\frac{p(D/m)}{p(N/m)} \tag{4.9}$$

The estimated detection rate and false positive rate are given by summing these proportions over the standardized age distribution.

$$\hat{p}(screen+ve/N) = \sum_m \hat{p}(screen+ve/N,m)\hat{p}(m/N) \tag{4.10}$$

$$\hat{p}(screen+ve/D) = \sum_m \hat{p}(screen+ve/D,m)\hat{p}(m/D) \tag{4.11}$$

The overall screen positive rate of the algorithm is given by

$$\hat{p}(screen+ve) = \hat{p}(screen+ve/D)\,\hat{p}(D) + \hat{p}(screen+ve/N)\,\hat{p}(N)$$

Modelling

A third method, adopted by Wald *et al* (1988), does not use test data but assumes multivariate Gaussian for the distributions of log(MoM)'s. The likelihood ratios are applied to a standardized age distribution as described previously and the detection rate and false

positive rate are given by numerically integrating the multivariate Gaussian densities over the regions where the likelihood ratio modifies the age specific risk of Down's syndrome to below a selected cut-off c, for each maternal age. This approach has the advantage of supplying more precise estimates of performance, but is necessarily based on more modelling assumptions and may lead to over estimated performance.

An alternative approach to evaluating the integrals is to use Monte Carlo methods to simulate samples of $(\underline{x}/N)$ and $(\underline{x}/D)$ from the original multivariate Gaussian model of the trials data and compute likelihood ratios. These are integrated by simulation methods over a standardized age distribution in the regions where pregnancies are screened positive. Such an approach provides flexibility since the distributional assumptions and the existing risk algorithm can be changed with little programming effort. Repeated simulation of samples drawn from the multivariate Gaussian model of the original trial, corresponding in number to the trial, provides a tool for calculating standard errors to assess the statistical accuracy.

Bishop (1994) shows that the methods of estimating the detection rates give rise to a variance that is proportional to the reciprocal of the sample size used. Bishop (1994) conducts a simulation study to produce the standard errors of the detection rates and false positive rates and discusses the confidence intervals attained from different sample sizes. Since the samples of affected data are small, the standard errors of the detection rates are large which leads to extremely wide confidence intervals for the detection rates.

The following section describes the simulation algorithm used in Wright *et al* (1993a) to calculate the standard errors of the detection rates. The algorithm is used to simulate samples from the multivariate distribution that is specified by the parameter estimates under the fitted models *ALIN2*, *UNLIN2*, and *HNLIN2*. These are shown in Table 4.1. Detection rates are calculated for the simulated samples under these parameter

estimates and under the parameter estimates specified by Wald *et al* (1992) and (1993). A copy of the simulation software that is used to conduct the analysis is given in Appendix A of this thesis.

PARAMETER	ANALYTE	UNAFFECTED	DOWN'S SYNDROME
MEANS	AFP	0	-0.2399
	UE3	0	-0.2824
	HCG	0	0.7133
SD	AFP	0.3747	0.4617
	UE3	0.3153	0.3840
	HCG	0.5416	0.5492
R	AFP-UE3	0.2971	0.4778
	AFP-HCG	0.1438	0.0663
	UE3-HCG	-0.0819	-0.1718

Table 4.1: Means, standard deviations, (SD), and correlation coefficients, (R), of residuals data recorded from 93 affected and 1284 unaffected pregnancies.

4.5 Application of Monte Carlo Simulation Methods

The effect of sampling error on the reported detection rates can be established by repeatedly applying the following algorithm to build up a sample of detection rates.

The whole study is simulated by simply sampling from the assumed distributions of MoM's, or appropriately transformed values, with the sample sizes equivalent to those in the study. The sample means, standard deviations and correlations for the simulated study are specified. A risk cut-off, c, is selected, which when applied to a given population, will produce a screen positive rate of p. This is done by repeatedly calculating the screen positive rate, using the method set out in section 4.3, with different cut-off levels until a rate of p is attained. The detection rates and false positives rates are calculated using the methods described under *Modelling*.

The Monte Carlo methods of simulation are used to simulate 100 samples from the distributions specified by the parameters from the models *ALIN2, UNLIN2* and *HNLIN2,* that were fitted in Chapter three. These are given in Table 4.1. The samples are used to calculate confidence intervals for estimates of detection rates which are derived from the algorithm based on the parameters in Table 4.1 and from the algorithm described in Wald *et al* (1988). All studies are modelled against the same maternal age distributions for England and Wales for the period 1986-1988 (Birth statistics 1986 - 1988). This distribution is tabulated in Appendix B. The detection rates are estimated at an overall 5% screen positive rate. Confidence intervals for the detection rates based on the analytes AFP, UE3 and HCG, and also for the rates based on AFP and HCG are derived, thus, the benefit of screening with UE3 can be assessed. The results of the study are given in Tables 4.2-4.4.

Markers	Dating method	No. of controls	No. of cases	Detection rate and 95% confidence interval at 5% screen positive rate	
AFP, UE3, HCG	LMP	1284	93	59.9	(52.77 , 67.01)
AFP, HCG	LMP	1284	93	56.9	(50.14 , 63.72)

Table 4.2: Detection rates and confidence intervals based on the parameter estimates given in Table 4.1.

Markers	Dating method	No. of controls	No. of cases	Detection rate and 95% confidence interval at 5% screen positive rate	
AFP, UE3, HCG	LMP	2113	77	58.56	(51.56 , 65.55)
AFP, HCG	LMP	2113	77	56.31	(49.88 , 62.74)

Table 4.3: Detection rates and confidence intervals based on the parameter estimates given in Wald *et al* (1992) and in Table 3.1.

Parameter estimates	Difference in detection rate and 95% confidence interval at 5% screen positive rate	
Table 4.1	2.98	(-0.22 , 6.17)
Wald *et al* (1992)	2.25	(-1.07 , 5.58)

Table 4.4: Differences in detection rates based on AFP and HCG and the rates based on AFP, UE3 and HCG and confidence intervals using both sets of parameter estimates.

Table 4.2 and Table 4.3 show the mean detection rates plus confidence intervals for the mean detection rates based on the parameter estimates given in Table 4.1 and the estimates reported in Wald *et al* (1992), respectively. The standard errors are large which probably explains the argument over the differences in reported detection rates. Table 4.4 shows the differences in the mean detection rates when UE3 is added to AFP and HCG and confidence intervals for these differences, for both studies. Both studies show an increase in detection when UE3 is added to the screen. However, when the detection rates are considered along with their standard errors there is little benefit in screening with UE3. Moreover, it is statistically very likely that some studies will show an apparent decrease in detection rate on addition of UE3. The study also indicates that any benefit in screening with UE3 is greater using the screening algorithm presented in this thesis.

4.6 Some results from discriminant analysis and error rate estimation

This section reviews some of the important results from the literature on discriminant analysis and screening. These are then adapted to examine the degree of bias in the estimates of detection rates and false positive rates.

Discriminant rules are designed to minimize the error rates. The error rates associated with the screening algorithms are the false positive rates and false negative rates. Definitions of these are given in Chapter one. An age related risk is combined with the likelihood ratio of the class conditional densities of transformed MoM analyte values to produce a risk which is used to classify pregnancies as either unaffected or Down's syndrome. If the group prior probabilities and class conditional densities are known, the realized feature vectors can simply be 'plugged in' to the probability model and the optimal rates of allocation can be computed.

The problem of unknown, or partially known group conditional densities is greater than the problem of absent prior probabilities (McLachlan (1992)). If the group conditional densities are unknown reliable estimates of the probability density functions can be constructed from feature vectors whose classifications are already known. This gives rise to a sample based discriminant rule that is designed from training data. The sample based rule provides estimates of the true rates of allocation.

Parametric and nonparametric techniques can be used to estimate the class conditional densities. The screening algorithm adopted by Wald *et al* (1988) uses a parametric approach to estimate the class conditional densities and thus to formulate a sample based allocation rule. The parameters are often estimated from a design set by robust methods. The sample based allocation rule is used to reclassify the design set and the error rates of allocation are computed. However, the estimated parameters are optimized for this design set so the apparent error rates of the allocation rule do not reflect the true error rates of the rule when it is applied to an independently selected data set from the same distribution. McLachlan (1992) discusses the difficulties in obtaining unbiased estimates of the error rates of a sample based allocation rule. The following section defines the types of error rates associated with discriminant rules and it also fixes the notation for the subsequent sections.

4.6.1 Notation

The notation provided is in accordance with that given in McLachlan (1992). Let $r(\underline{x};\underline{t})$ define a decision rule that is formed from realized training data $\underline{t}$, consisting of n, p-dimensional feature vectors $\underline{x}$. The feature vectors are drawn from a feature space for the purpose of allocating the entity to one of g mutually exclusive and exhaustive groups,

$G_1, G_2, \ldots, G_g$. Define $n = n_1 + n_2 + \ldots + n_g$ and n_i to be the number of observations sampled from group G_i. The allocation rates associated with the optimal Bayes rule are defined

$$eo_{i,j}(F_i) = \Pr\{r_o(\underline{X}; F_i) = j / G_i\} \qquad i, j = 1,2,\ldots,g \tag{4.12}$$

which denote the probability that a randomly selected feature vector from G_i is allocated to G_j, via $r_o(\underline{X}; F_i)$ where F_i is the distribution function of $\underline{X}$ in group i.

The group specific optimal error rates are given as

$$eo_i(F_i) = \sum_{j \neq i}^{g} eo_{i,j}(F_i) \quad i = 1,2,\ldots,g \tag{4.13}$$

and the overall optimal error rate is defined

$$eo(F) = \sum_{i=1}^{g} p(G_i) eo_i(F_i) \tag{4.14}$$

where $p(G_i)$ denote the arrival rates of each group and F is the distribution function of $\underline{X}$.

The allocation rates of the sample based discriminant rule, $r(\underline{x}; \underline{t})$, which is formed from training data, $\underline{t}$, are

$$ec_{i,j}(F_i; \underline{t}) = pr\{r(\underline{X}; \underline{t}) = j / G_i, \underline{t}\} \tag{4.15}$$

For groups i and j, equation (4.15) represents the probability that any random vector in the training set, $\underline{t}$, belonging to G_i is allocated to G_j, $(i, j = 1,2,\ldots,g)$. The conditional error rates are often referred to as actual error rates. The group specific conditional error rates are defined

$$ec_i(F_i; \underline{t}) = \sum_{j \neq i}^{g} ec_{i,j}(F_i; \underline{t}) \qquad (i = 1,2,\ldots,g) \tag{4.16}$$

and the overall conditional error rate is then

$$ec(F; \underline{t}) = \sum_{i=1}^{g} p(G_i) ec_i(F_i; \underline{t}) \tag{4.17}$$

The unconditional, or expected rates of allocation are formed by averaging the conditional allocation rates over the distribution of training sets. The unconditional allocation rates are given as

$$eu_{i,j}(F_i) = E[ec_{i,j}(F_i;\underline{T})] \tag{4.18}$$

$$= \Pr\{r(\underline{X};\underline{T}) = j / G_i\} \qquad (i,j = 1,2,\ldots,g)$$

The unconditional group specific error rates are denoted as

$$eu_i(F_i) = \sum_{j \neq i}^{g} eu_{i,j}(F_i) \tag{4.19}$$

which gives the expected misclassification rate of each group. The overall unconditional error rate is defined

$$eu(F) = \sum_{i=1}^{g} p(G_i) eu_i(F_i) \tag{4.20}$$

The true error rate is the expected error rate of the rule on future samples drawn from the same distribution as the design set. The unconditional allocation rates are of less importance than the conditional allocation rates in the context of diagnostic testing. The conditional allocation rates are used to monitor the performance of the screening algorithm.

This section has discussed the problem of bias when estimating error rates from the same data that is used to formulated a sample based allocation rule. Some nonparametric and parametric methods of computing bias corrected error rates are discussed in the following section. Special attention is paid to the conditional error rates for the reasons noted above. Definitions are supplied in terms of a fixed group, G_1, since extensions to other groups is straightforward.

4.6.2 Some parametric and nonparametric error rate estimators

The simplest estimator of the conditional error rate of $r(\underline{x};\underline{t})$, when allocating entities from G_1, $(ec_1(F_1,\underline{t}))$, is the apparent error rate, $A_1(\underline{t})$, which is calculated by reclassifying the design set. The apparent error rate is the proportion of observations from G_1 in $\underline{t}$, that are misclassified by $r(\underline{x};\underline{t})$.

If $\underline{z}$ is a (gxn) matrix of group indicators such that

$$z_{i,j} = \begin{cases} 1 \text{ if entity j belongs to group i} \\ 0 \text{ elsewhere} \end{cases}$$

$$\text{then } A_1(\underline{t}) = \frac{1}{n_1}\sum_{j=1}^{n} z_{1,j} Q[1, r(\underline{x}_j;\underline{t})] \tag{4.21}$$

$$\text{where } Q[1, r(\underline{x}_j;\underline{t})] = \begin{cases} 0 & if\ r(\underline{x}_j;\underline{t}) = 1 \\ 1 & if\ r(\underline{x}_j;\underline{t}) \neq 1 \end{cases} \quad \text{and } n_1 = \sum_{j=1}^{n} z_{1,j}$$

The apparent error rate of the rule is computed from the same design set that is used to formulate the rule. This provides an optimistic view of the overall performance. The extent of the bias relates to the complexity of the discriminant rule. Parametric and nonparametric methods of bias correction can be applied to remove the bias and provide reliable assessments of the true error rates of the rule. A review of these methods is provided in McLachlan (1992). This section reviews some of these nonparametric approaches.

Data resampling techniques are central to unbiased error rate estimation. Interest in computer intensive methods has surged since Efron's series of publications on the bootstrap, jackknife and cross validation approaches to estimation. (Efron (1979), Efron (1982), Efron (1983)). The improvement in computer technology has also increased the relative utility of these methods. One approach that uses nonparametric resampling methods

to eliminate the bias in the apparent error rate is the leave-one-out, or cross validation method, which was first introduced by Lachenbruch and Mickey (1974). An entity is omitted from the training data and the discriminant rule is recalculated from the remaining observations. The omitted entity is allocated on the basis of the new decision rule and checks for misclassification are made. The process is repeated so that each observation is removed from the training data, and a new allocation rule is formed from the remaining observations which is then used to classify the entity. Records of misclassification are made at each stage of the procedure until the training data is reduced to a single entity.

The cross validation estimator of the apparent error rate, provides estimates of the conditional error rate and is defined as

$$A_1^{(cv)} = \frac{1}{n_1}\sum_{j=1}^{n} z_{1,j} Q[1, r(\underline{x}_j; \underline{t}_{(j)})] \tag{4.22}$$

where $\underline{t}_{(j)}$ denotes the training data $\underline{t}$ with the j^{th} observation omitted.

The cross validation estimator of the overall apparent error rate, $A^{(cv)}$, is nearly unbiased but often is highly variable if n is small.

Another nonparametric method of estimating bias uses the jackknife resampling technique. The jackknife estimate of bias was proposed by Quenouille in the mid 1950's. (Efron and Tibshirani (1993)). It was the first computer based resampling method for estimating bias and standard errors. The jackknife estimate of the apparent error rate omits one observation from the training data, and formulates the discriminant rule on the remaining observations. This is then used to classify each observation that is remaining in the training set. The proportion of misclassified observations is calculated at each stage. The apparent error rate is estimated from each set of allocated training data with one distinct observation omitted. The jackknife estimator of the apparent error rate is given as

$$A_1^{(J_o)} = A_1 + (n_1 - 1)(A_1 - A_{1(\cdot)}) \tag{4.23}$$

where $A_{1(j)} = \sum_{k \neq j}^{n} z_{1,k} Q[1, r(\underline{x}_k; \underline{t}_{(j)})] / (n_1 - 1)$ and $A_{1(.)} = \sum_{j=1}^{n} A_{1(j)} / n_1$

The jackknife form of the apparent error rate is appropriate for estimating the unconditional error rate as n approaches infinity.

The nonparametric bootstrap method of resampling was introduced by Efron (1979), and a series of related publications concerning its applications has followed. A full review of the techniques involving the bootstrap is given in Efron and Tibshirani (1993). The nonparametric method of bootstrapping forms an estimate, $\hat{F}$, of the underlying distribution function, F, from the realized training data. $\hat{F}$ is referred to as the bootstrap distribution. The nonparametric version of the bootstrap calculates $\hat{F}$ as the empirical distribution, which approximates the true distribution of the observations by placing a mass of $\frac{1}{n}$ on each of them. Monte Carlo methods of simulation are used to draw bootstrap samples, with replacement, from $\hat{F}$, which are subsequently used to calculated bootstrap estimates of apparent error.

An algorithm for the nonparametric bootstrap bias correction of the apparent error rate, based on G_1, follows.

(1) Form an estimate, $\hat{F}_i$, of the underlying distribution, F_i from the realized training data, such that $\hat{F}_i$ is the empirical distribution with mass $\frac{1}{n_i}$ at each $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{n_i}$ in G_i, $i = 1,2,\ldots,g$.

(2) Use Monte Carlo methods of simulation to simulate a new set of data, $\underline{t}^*$, from $\underline{t}$. Samples are drawn independently and with replacement such that $\underline{t}^*$ consists of the realized values of $\underline{X}^*_{i,1}, \underline{X}^*_{i,2}, \ldots, \underline{X}^*_{i,n_i} \overset{iid}{\approx} \hat{F}_i$ i = 1,2,..,g.

(3) Form the rule $r(\underline{x};\underline{t}^*)$ from $\underline{t}^*$ in the same manner as $r(\underline{x};\underline{t})$ is formed from the original training set.

(4) The apparent error rate, $A_1(\underline{t}^*)$, of $r(\underline{x};\underline{t}^*)$ for group G_1 is given as

$$A_1(\underline{t}^*) = \frac{1}{n_1^*}\sum_{j=1}^{n} z_{1,j}{}^* Q[1, r(\underline{x}_j;\underline{t}^*)] \tag{4.24}$$

where $n_1^* = \sum_{j=1}^{n} z_{1,j}{}^*$

Under separate sampling $ec_1(\hat{F}_1;\underline{t}^*) = \frac{1}{n_1}\sum_{j=1}^{n} z_{1,j} Q[1, r(\underline{x}_j;\underline{t}^*)]$ and the difference $\hat{d}_1^* = A_1(\underline{t}^*) - ec_1(\hat{F}_1;\underline{t}^*)$ is computed.

(5) The bootstrap bias of the apparent error rate for the first group is defined as

$$\hat{b}_1^{(B)} = E^*(\hat{d}_1^*) = E^*\left\{A_1(T^*) - ec_1(\hat{F}_1;T^*)\right\}$$

where E^* is the expectation over the distribution of T^*. The bias $\hat{b}_1{}^{(B)}$ can be estimated by $\overline{\hat{d}_1^*}$ which is obtained by averaging over K independent samples of training data drawn from T^* such that

$$\overline{\hat{d}_1^*} = \sum_{k=1}^{K} \hat{d}_{1,k}{}^* / K \tag{4.25}$$

and $\hat{d}_{1,k}{}^* = A_1(\underline{t}_k{}^*) - ec_1(\hat{F}_1;\underline{t}_k{}^*)$ for the k^{th} bootstrap replication.

The bootstrap bias corrected version of the apparent error rate for G_1 is given by

$$A_1{}^{(B)} = A_1 - \hat{b}_1{}^{(B)} \tag{4.26}$$

Efron and Tibshirani (1993) point out that 50 -100 replicated bootstrap samples is a sufficient number for standard error and bias estimation. The standard error of the Monte Carlo approximation $\overline{\hat{d}_1^*}$ to the bootstrap bias is the positive square root of

$$\sum_{k=1}^{K}\left(\hat{d}_{1,k}^* - \overline{\hat{d}_1^*}\right)^2 / \{K(K-1)\} \qquad (4.27)$$

The error rate estimators discussed so far use nonparametric approaches to estimation. If adequate information concerning the class conditional densities is available, the parametric bootstrap may be appropriate.

The parametric bootstrap postulates a form of the class conditional distributions, F, or F_i in the case of separate sampling. The unknown parameters of the distributions $\hat{F}$ or $\hat{F}_i$ are estimated from the training data. Maximum likelihood estimates are commonly used for these estimates. Instead of sampling from the data, Monte Carlo simulation methods are used to generate pseudo bootstrap samples from $\hat{F}$ or $\hat{F}_i$, corresponding in size to the original training data. An algorithm for calculating the parametric bootstrap under a separate sampling scheme follows.

(1) Postulate forms for the population distribution functions F_i. Define the vector of unknown parameters as Ψ.

(2) Calculate estimates of the distribution functions, $\hat{F}_i$ that have the same form as F_i and estimate the unknown parameters, $\hat{\Psi}$, from the training data, $\underline{t}$.

(3) Apply Monte Carlo methods of simulation to generate a parametric bootstrap sample, $\underline{t}^*$, of size n_i from each $\hat{F}_i$.

(4) Continue from step (4) of the algorithm for calculating nonparametric bootstrap estimates of bias.

4.7 Application of the nonparametric bootstrap to screening to calculate estimates of bias

The nonparametric method of bootstrapping is employed to calculate the bias in the detection rates and false positive rates of the screening algorithm given by Wald *et al* (1988). The original detection rates and false positive rates, and the bias corrected rates, are calculated over a standardized maternal age distribution. The extent of the bias is assessed over 100 bootstrap replications. Maternal age specific bias corrected estimates of detection rates and false positive rates are also obtained. This enables an assessment of the changes in the bias caused by reclassiying the design set when screening with unrepresentative maternal age distributions. The values of the weighted bias in the detection rate and false positive rate are shown in Table 4.5.

	DR (%)	FPR (%)
Weighted Bias	+0.17	-0.51

Table 4.5: Weighted bias in detection rates and false positive rates.

The study indicates that the estimated detection rates and false positive rates are only marginally affected by bias. The bias in the rates for maternal ages in the ranges of 11-15 and 45-55 was zero and the bias in the rates associated with the ages in the range of 16-44 showed little variation from the values quoted in Table 4.5. Therefore, screening with maternal age distributions that are unrepresentative of the target population will have little effect on the extent of the bias caused by reclassifying the design set.

4.8 Conclusion

This chapter has investigated the extent of the sampling error in the estimated detection rates and has shown how the differences in the reported estimates can be explained by sampling error. The results of the simulation studies suggest there is little benefit in screening with UE3 in addition to AFP and HCG, however, the algorithm used by Wald *et al* (1992) reduces the potential use of UE3 as a screening variable for Down's syndrome. This chapter has also discussed the problem of bias in error rates that are estimated by reclassifying the design set with the allocation rule that was formulated from this set. It has been shown that the bias in the estimated detection rates and false positive rates is small. Also, screening with unrepresentative maternal age distributions does not lead to an increase in this bias.

Chapter 5

The inclusion of a non-specific screen: a question of atypicality

5.1 Introduction

The screening algorithm used by Wald *et al* (1988) classifies pregnancies as either unaffected or Down's syndrome. Frequently other abnormalities, notably trisomy 18 and trisomy 13, may have MoM values, or some appropriate transformation of them, that translate into low risks.

A study was conducted by Heyl *et al* (1990) to assess the performance of the screening algorithm when dealing with various abnormalities as well as trisomy 21. Serum samples were collected from 16 trisomy 21 pregnancies and 18 with other autosomal aneuploides, including trisomy 18 and trisomy 13, whose mothers were known to have had an amniocentesis on the basis of advanced age.

Heyl *et al* (1990) reported that the algorithm detected 63% of trisomy 21 pregnancies with a false positive rate of 5%, using a risk cut off of 1:365. Only 3 out of the 18 pregnancies with other abnormalities were screened positive, none of which were either trisomies 18 or 13. Most abnormal pregnancies received extremely low risks, emphasising why such risks cannot legitimately be used to reassure a mother that her pregnancy is normal.

Low detection in non-Down's syndrome abnormalities is reflected through the MoM analyte values. Often non-Down's syndrome abnormalities have MoM values that are dissimilar to those from trisomy 21 pregnancies and also have a low probability of being

associated with an unaffected pregnancy. Frequently these pregnancies are classified as unaffected even though they may be atypical of this outcome.

A strategy to improve detection of non-Down's syndrome abnormalities would be to incorporate other specific screens into the screening algorithm. Such screens are not generally practiced because of the lower incidence rates of other abnormalities. Moreover, in cases such as trisomy 18, it may be considered unnecessary to incorporate a specific screen due to their lethality. However, interest surrounding specific screening of this kind has more recently developed. Staples *et al* (1991) examines the feasibility of extending second trimester screening for Down's syndrome to incorporate a specific screen for trisomy 18. Staples *et al* (1991) reports the most useful analytes for identifying trisomy 18 are UE3, free α-subunit HCG, free β-subunit HCG, estradiol and Human placental lactogen. The study focuses on 12 pregnancies with trisomy 18 outcomes and 390 matched controls. At a risk cut off of 1:400, 83.3% of affected pregnancies were detected at a false positive rate of 2.6%.

It would be impractical to specifically screen for many fetal abnormalities simultaneously. An alternative approach to the problem is suggested by Wright *et al* (1993). The approach incorporates a non-specific classification into the existing screen for those abnormalities that are unlike trisomy 21 but are also highly atypical of unaffected pregnancies. An index of atypicality can be constructed using the well established statistic, the Mahalanobis distance. By assigning an atypicality index to all pregnancies that are classified as unaffected, those that have sufficiently large indices but are associated with low risks of Down's syndrome can be classified as non-specific with no risk being reported until further investigations have been undertaken.

Section 5.2 of this report describes the methodology employed by Wright *et al* (1993) to monitor atypicality in data from a multivariate Gaussian distribution using

Mahalanobis distances and illustrates how the calculation can easily be introduced into the existing screen with just a simple modification to the screening algorithm used by Wald *et al* (1988). This section also discusses the advantages should the enhancement be used as part of a screen for Down's syndrome. A summary of the materials used to illustrate the value of the modified screen by Wright *et al* (1993) is given in Table 5.1.

Section 5.3 provides full details of the results of the analysis. The effects of the modified screen on the classification of abnormal pregnancies is considered. The risks and Mahalanobis distances for these pregnancies are provided in Table 5.2, and a summary of these results is given in Table 5.3. A discussion of the effects of the modified screen on the classification of unaffected pregnancies is also included. Section 5.4 reviews the consequences of using the enhanced screen in conjunction with the existing algorithm. The results are shown in Table 5.4. Figure 5.1 provides a plot of the 99 % contours of the atypicality indices for the unaffected and Down's syndrome pregnancies.

5.2 Methodology

5.2.1 Monitoring for atypicality in multivariate Gaussian data

Given an observation $\underline{y}$ from a p-dimensional multivariate Gaussian distribution with mean vector $\underline{\mu}$ and covariance matrix Σ, the Mahalanobis distance, d, is defined

$$d = \left(\underline{y} - \underline{\mu}\right)^T \Sigma^{-1} \left(\underline{y} - \underline{\mu}\right) \qquad (5.1)$$

By calculating d for each p-dimensional feature vector, $\underline{y}$, an assessment of atypicality is given by declaring as atypical values those observations whose Mahalanobis distance exceeds the upper (1 - α) 100% quantile of the chi-squared distribution with p

degrees of freedom, for suitably chosen values of α. Atypical observations are considered to be outliers of the distribution owing to the extremeness of their Mahalanobis distances. Larger indices of atypicality present more evidence to suggest the observations have been misclassified.

5.2.2 Incorporating the atypicality index into the existing screen

The computation of the Mahalanobis distance can easily be incorporated into the existing screen with just a simple modification to the current algorithm given by Wald *et al* (1988). Since the Mahalanobis distance is the exponent part of the Gaussian density function, which is already computed when the likelihood ratios are evaluated, the information is readily available for extraction.

A copy of the computer software, written in Turbo Pascal Version 4, which is designed to imitate the risk algorithm given by Wald (1988) but also has the enhancement of reporting atypicality indices for both Down's syndrome and unaffected classifications, is contained in Appendix (D) of this thesis, along with detailed documentation.

5.2.3 An assessment of atypicality as part of a screen for Down's syndrome

It has been suggested that one method of ensuring the classification of a pregnancy is reasonable is to assign to each feature vector of log(MoM) values an index of atypicality as an assessment of how typical the observation is of its particular classification. Also, it is considered that the inclusion of a non-specific category for those observations deemed atypical of unaffected pregnancies, but unlike Down's syndrome pregnancies may partly overcome the problem of low risks being assigned to pregnancies with disorders other than trisomy 21.

If these hypotheses are true the enhanced algorithm that is adjusted to incorporate atypicality indices would be expected, not only to adequately detect pregnancies with Down's syndrome but also to identify a proportion of those pregnancies with other disorders. Moreover, this would have the effect of reducing the false positive rate.

5.2.4 Method

The 1993 paper by Wright *et al* assesses the modified algorithm by examining its detection rate based on transformed AFP and HCG MoM values that are recorded from pregnancies associated with various fetal chromosomal disorders. In order to directly compare the existing screen with the modified screen, the parameter estimates used in the modified screen are those reported by Wald *et al* (1988) and (1992) and are shown in Table 5.1.

OUTCOME	UNAFFECTED		DOWN'S SYNDROME	
	AFP	HCG	AFP	HCG
Sample Size	385	385	77	77
Means	0	0	-0.3286	0.6961
Std. dev.	0.4656	0.5720	0.4720	0.6309
Correlations	0.0723		0.1703	

Table 5.1: Means, standard deviations and correlations assumed in calculating Down's syndrome risks and Mahalanobis distances. (From Wald *et al* (1988) and (1992)). Values are in logs.

To demonstrate the value of the inclusion of a non-specific screen, figures for a total of 37 pregnancies with various abnormalities were taken from published literature (Bogart

et al (1987), Staples *et al* (1991), Johnson *et al* (1991)) and analysed using the modified screen given in Appendix (D) of this report. In order to provide an unbiased assessment of the algorithm, all mothers with unknown ages were assumed to be 35 years old, thus ensuring the best possible chance of a screen positive result.

A further 2000 unaffected pregnancies from the Royal Gwent Down's syndrome Screening program were analysed to determine the possibility of a reduction in false positive rates with the modified screen. UE3 was not used because the analyte is not assayed by the Gwent screening program.

5.3 Illustration

5.3.1 Abnormal pregnancies

Risks and values of the Mahalanobis distances for the 37 abnormal pregnancies are given in Table 5.3.

The Mahalanobis distances should be compared with quantiles of the chi-squared distribution with $\upsilon = 2$ df. Selected contours for comparison were 95%, 98% and 99%, the corresponding chi-squared statistics being 5.991, 7.824 and 9.210 respectively. The data are plotted together with the 99% contours of both unaffected and Down's syndrome distributions in Figure 5.1. Table 5.4 gives a summary of the performance of the screening algorithm.

A most concerning result of this analysis is that such low risks can frequently be assigned to abnormal pregnancies when these pregnancies are clearly highly atypical of the this outcome.

As indicated by Table 5.5 it is notable that overall an additional 15 out of the 37 abnormal outcomes are classified as being abnormal using a 1% cut off on the Mahalanobis

distance. These 15 outcomes correspond to the 15 points which fall outside the 'Normal outcomes' contour in Figure 5.1. Of these 15, 73.33% are known to be trisomy 18 .

Table 5.3 also gives Mahalanobis distances for the distribution of Down's syndrome. It is notable that the low risk, abnormal pregnancies are even more atypical of Down's syndrome than they are of unaffected outcomes. Using a risk cut-off of 1:300, only 2 pregnancies were screened positive.

5.3.2 Unaffected pregnancies

A summary of the results of the analysis of the 2000 unaffected pregnancies is given in Table 5.4. Using the nominal cut-off, 1.5% of normal pregnancies were screened negative and classified as atypical. This is rather large compared with the nominal 1% reflecting the fact that the distribution of log (MoMs) is only approximately Gaussian in form.

5.4 Conclusions

This chapter has illustrated how the current screening algorithm frequently fails to recognise other congenital abnormalities that occur during pregnancy and often assigns to them extremely low risks, misclassifying them as being unaffected by any disorder.

The modification needed to the current algorithm to monitor for atypicality is a simple one, and would enable pregnancies that were highly atypical of one outcome but unlike the other to be classified in their own right as non-specific to either outcome.

In the sample of 37 abnormal pregnancies extracted from the literature (Bogart *et al*, (1987), Staples *et al* (1991), Johnson *et al* (1991)) an additional 15 pregnancies were identified as atypical of unaffected pregnancies but were assigned risks that eliminated any suspicions of Down's syndrome. It is most concerning that such low risks can lead to false

reassurances. A set of atypical results would provide an indication for a considered review of the case before any notes are reported.

Case	MoM		Abnormality	Maternal Age	Risk	Source	Mahalanobis Distance	
	AFP	HCG					Unaffected	Down's
1	3.59	5.31	+13	35*	53	B	16.12	12.86
2	1.03	1.53	+13	35*	509	B	0.57	0.93
3	0.25	0.37	+13	35*	1100	J	11.84	11.65
4	1.67	0.45	+13	35*	17400	J	3.67	11.09
5	0.89	0.65	+13	35*	2430	J	0.73	4.21
6	0.50	0.75	+13	35*	593	J	2.45	3.11
7	0.50	0.78	+13	35*	553	J	2.39	2.91
8	1.76	4.96	XO	35*	41	B	9.77	5.11
9	1.46	1.77	XXY	35*	639	B	1.64	2.44
10	0.33	0.10	69,XXX	35*	3720	B	23.29	26.51
11	0.75	0.16	46,XX/47,XX,+9	35*	12600	J	11.96	18.73
12	1.67	0.20	46,XY/47,XY,+Mar	35*	64800	J	10.66	20.71
13	1.28	0.53	46,XY,7q+	35*	7400	J	1.81	7.52
14	0.20	0.58	47,XXY	35*	580	J	12.68	10.40
15	0.58	1.15	45,X+	35*	346	J	1.45	1.03
16	0.80	1.53	46,XY,-18,+der18	35*	329	J	0.84	27.38
17	0.96	0.10	+18	35*	35900	B	18.51	20.99
18	1.01	0.15	+18	35*	27800	B	12.63	17.47
19	1.00	0.19	+18	35*	20700	J	9.7	23.79
20	2.50	0.21	+18	35*	153000	J	13.03	20.05
21	2.50	0.30	+18	35*	85300	J	9.46	12.28
22	1.25	0.32	+18	35*	16800	J	4.94	4.01
23	0.67	0.62	+18	35*	1490	J	1.51	4.47
24	1.33	0.85	+18	35*	3090	J	0.51	2.45
25	1.00	0.95	+18	35*	1420	J	0.02	3.12
26	0.72	0.72	+18	33	1970	S	0.86	15.40
27	0.69	0.20	+18	21	31800	S	9.51	13.43
28	0.53	0.23	+18	32	6620	S	9.10	32.04
29	0.47	0.07	+18	27	21300	S	26.51	1.95
30	0.82	0.93	+18	30	2420	S	0.20	20.41

Table 5.2 Cont...

Cont...

31	0.66	0.14	+18	23	39300	S	14.02	20.41
32	1.06	0.56	+18	29	12100	S	1.26	5.99
33	0.56	0.20	+18	18	20000	S	10.29	15.19
34	0.51	0.45	+18	32	2530	S	4.17	6.58
35	1.41	0.97	+18	30	6100	S	0.56	4.16
36	0.78	0.54	+18	27	8150	S	1.600	5.21
37	0.49	0.08	+18	25	25100	S	23.91	29.54

Table 5.2: Risks and Maholanobis distances for a total of 37 abnormal pregnancies, continued (Risks given to three significant digits.) Sources B = Bogart *et al* (1987).; S = Stables *et al* (1991).; J = Johnson *et al* (1991). Value for risk = n such that the risk is 1 : n. (Risks given to 3 significant digits.)
* No maternal age given, 35 assumed for purpose of risk calculation.

Outcome	Outcome Total	No. screened +ve	No. screened -ve but atypical of Unaffected at		
			5%	2%	1%
TR18	21	0	11	11	10
TR13	7	1	1	1	1
Others	9	1	4	4	4
Total	37	2	16	16	15

Table 5.3: Summary of results of modified screening algorithm for abnormal pregnancies.

	Screened Positive	Screened -ve but atypical of Unaffected at		
		5%	2%	1%
Frequency	58	85	43	30
%	2.9%	4.25%	2.2%	1.5%

Table 5.4: Summary of results of modified screening algorithm on 2000 unaffected pregnancies.

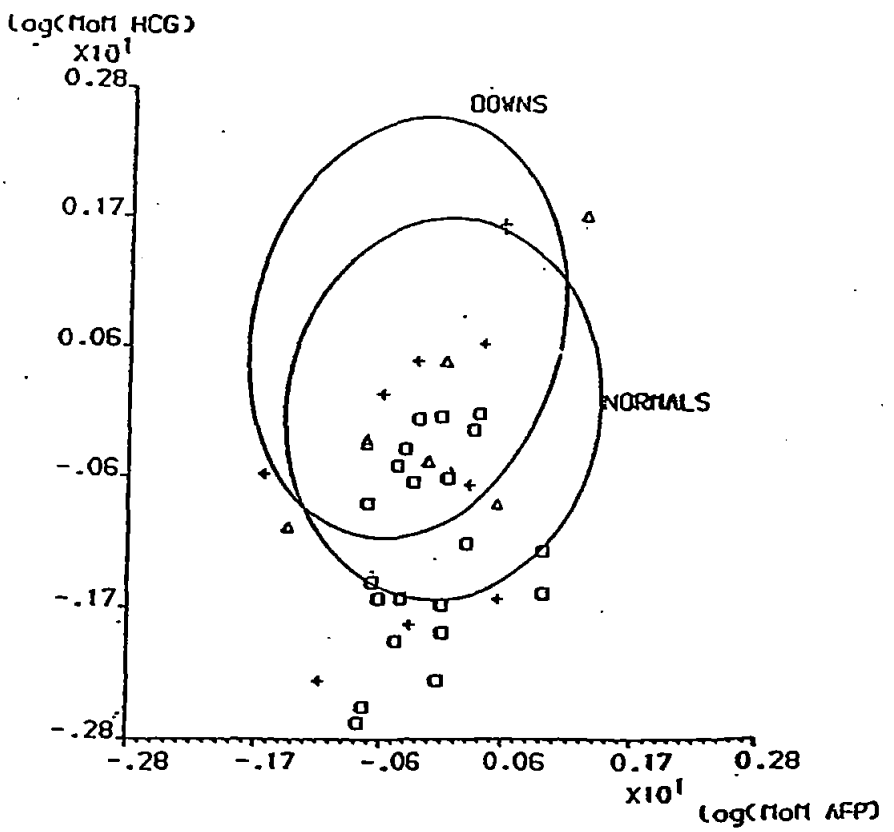

Figure 5.1: Plot showing cases in Table 5.2 and 99 % contours of the atypicality indices for the unaffected and Down's syndrome pregnancies., Δ= Trisomy 13, □ = Trisomy 18, + = others.

Chapter 6

A nonparametric alternative: the kernel method of density estimation

6.1 Introduction

The algorithm employed by Wald *et al* (1988) adopts a parametric approach to the problem of estimating the class conditional densities of transformed MoM analyte values. The class conditional distributions of MoM values are assumed to be adequately represented by multivariate lognormal distributions with differing mean and covariance matrices. Their 1988 and 1993 papers, however, highlight evidence of non-normality in the marginal distributions of AFP, UE3 and HCG which is particularly pronounced in the tails of the distributions. The distributions of UE3 demonstrate the most deviation from a Gaussian form. The problem is addressed by truncating the distributions to remove extreme values that fall outside a linear range on a normal probability plot. However, unreliable distributional assumptions affect the likelihood ratio based risks and the performance of the algorithm as a whole. With such rigid distributional assumptions surrounding parametric techniques of density estimation, nonparametric approaches based on more flexible methods may be more appropriate.

Nonparametric methods of estimating probability density functions are varied and there is a wealth of relevant literature (Hand (1981), Silverman (1986), Härdle(1991)). More common techniques include the traditional histogram, the nearest neighbour method, the kernel estimator, and the orthogonal series estimator. This chapter considers the kernel method of estimation which is certainly the most studied mathematically and for the purposes of this Ph.D., is sufficient. Section 6.2 covers the background and motivation in

using nonparamtetric methods of density estimation. Section 6.3 provides an initial discussion of the statistical framework of kernel methods as described by its founders. Section 6.4 of this chapter provides a concise overview of past and present research in this field. Attention is paid to recent developments in multivariate estimation techniques. It is the intention to provide a review of the results of published work in this area, theorems and proofs are omitted. Section 6.4.1 and section 6.4.2 review univariate and multivariate kernel methods of density estimation. Section 6.4.3 discusses automatic methods of window width selection. Some of the kernel estimators described at each stage of this chapter are applied to the clinical trials data for unaffected pregnancies which is summarized in Chapter three of this thesis. One and two dimensional density estimates are fitted to the class conditional distributions of residual analyte values for the models *ALIN2*, *UNLIN2* and *HNLIN2*, which are equivalent to log(MoM) values.

The application of nonparametric density estimation to the affected data is discussed as a separate issue in section 6.5. In section 6.6, the techniques used to compute parametric density estimates described by Wald *et al* (1988) are applied to the affected and unaffected distributions of residuals and the parametric density estimates are graphically compared to those constructed by nonparametric methods. The two approaches are compared through the detection rates achieved in two dimensions, and these are discussed in a concluding section, section 6.7. Copies of the software used in this chapter are given in Appendices E, F and G.

6.2 Background

The univariate kernel density estimator, was first described by Fix and Hodges (1951) who employed this technique in a discriminant analysis that was conducted to assist with medical diagnosis. A specific nonparametric kernel density estimate of the populations

under study was used in a discriminant rule that was subsequently used to classify new subjects. A more general estimator of this type was introduced by Rosenblatt (1956) and Parzen (1962). Initial extensions to multivariate data were supplied by Cacoullos (1966) and Epanechnikov (1969).

Let $\{X_i\}$ denote an independently and identically distributed sample of size n, $X_i \in \Re$ with pdf f and $\{\underline{X}_i\}$ denote an independently and identically distributed sample of size n, $\underline{X}_i \in \Re^d$ with pdf f.

The kernel method of estimation is an adaptation of the more restrictive naive estimator. Define a weight function such that

$$w(x) = \begin{cases} \frac{1}{2} & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

then the naive estimator becomes

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} w\left(\frac{x - X_i}{h}\right) \tag{6.2}$$

Unlike the histogram, the naive estimator is not affected by the choice of origin. The parameter h governs the window width of estimation. However, undesirable properties of this estimator are apparent in its jagged presentation. Figures 6.1- 6.3 plot the histograms of the distributions of residual AFP, UE3 and HCG values. Figure 6.4 plots the naive density estimate of the distribution of HCG. The plot demonstrates the crudeness of the naive estimator. The effect is exaggerated by the use of a small window width (h = 0.04). A generalized weighting function would be less artificial.

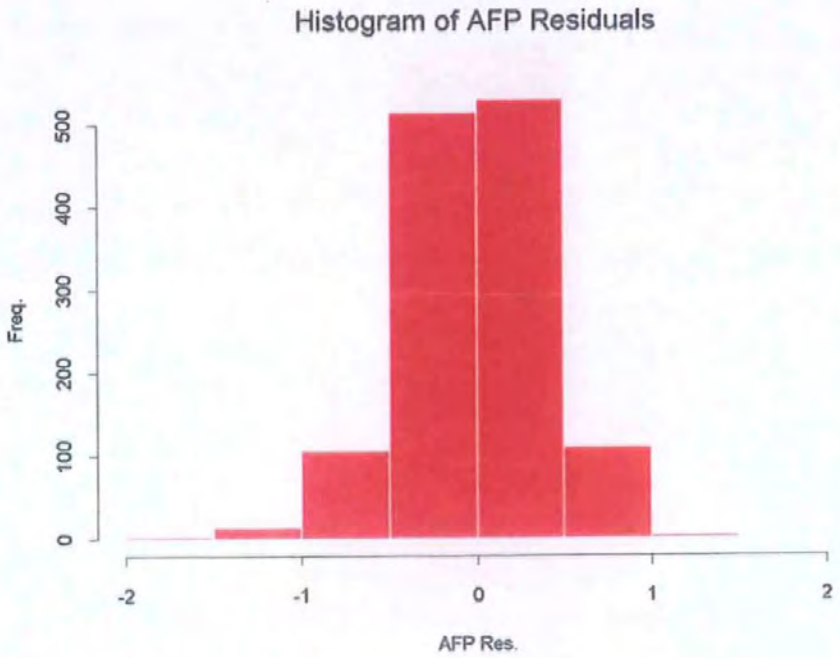

Figure 6.1: Histogram of AFP residuals.

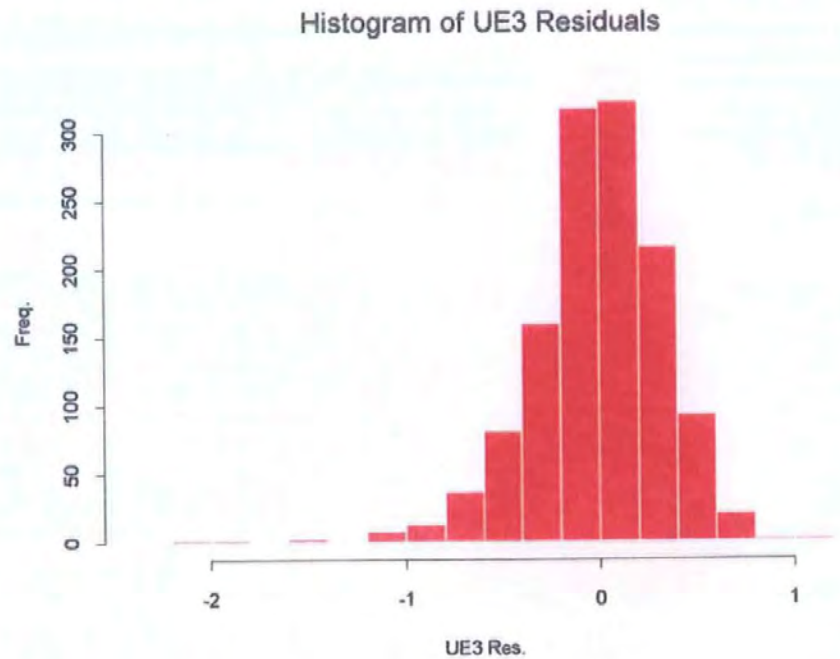

Figure 6.2: Histogram of UE3 residuals.

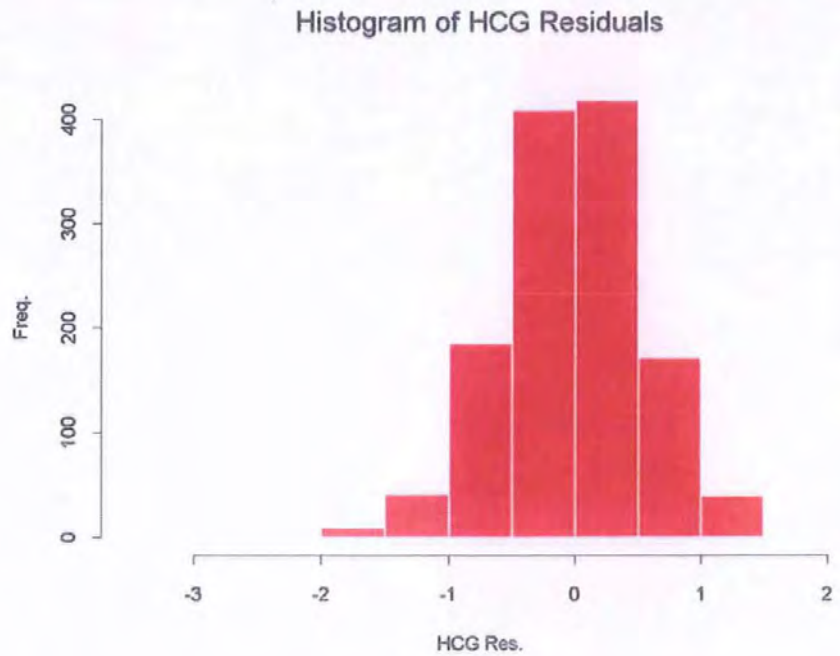

Figure 6.3: Histogram of HCG residuals.

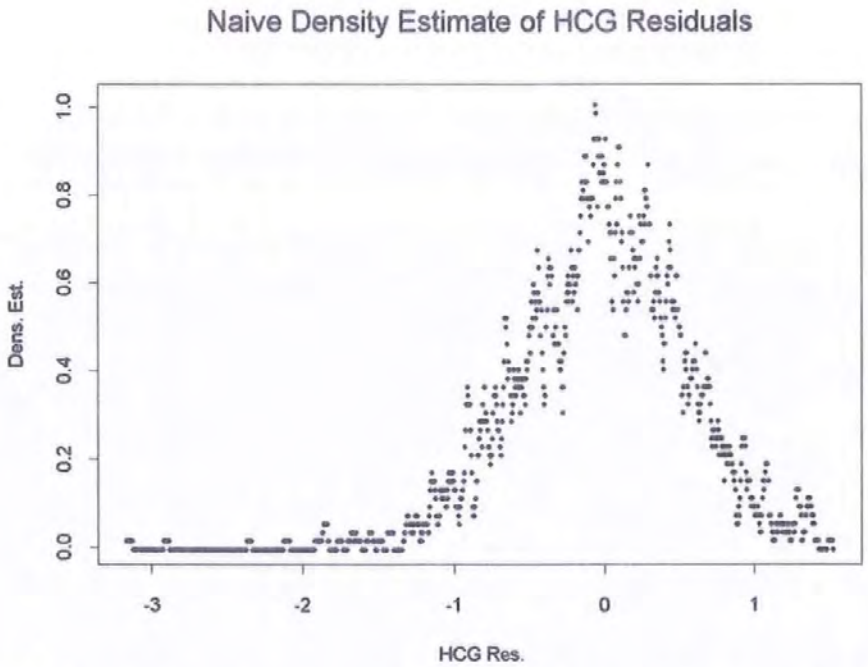

Figure 6.4: Naive density estimate constructed from HCG residual data with $h = 0.04$.

In essence, the kernel estimator provides a smooth version of the density estimate given by the naive estimator. The univariate kernel estimator with kernel function K is defined as

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{6.3}$$

K is a symmetrical function centered at zero that integrates to unity and h is the smoothing parameter. Essentially, as expressed by Silverman (1986), the kernel estimator is a weighted sum of individual symmetrical 'bumps' centered over each observation. The choice of kernel function determines the form of the density estimate and the degree of smoothing imposed on the estimate is determined through the choice of h.

With extensions to higher dimensions the general multivariate kernel density estimator is defined below.

$$\hat{f}(\underline{x}, H) = n^{-1}\sum_{i=1}^{n} K_H(\underline{x} - \underline{X}_i) \tag{6.4}$$

where K is a d-variate probability density function, H is a symmetrical positive definite $(d \times d)$ matrix, and $K_H(\underline{x}) = |H|^{-1/2} K(H^{-1/2}\underline{x})$. There are many permissible classes for H and these are discussed in section 6.4. Often it is satisfactory to replace H with a diagonal matrix yielding

$$\hat{f}(x_1, x_2 \ldots, x_d) = n^{-1}\sum_{i=1}^{n} \frac{1}{h_1 h_2 \ldots h_d} \cdot \prod_{j=1}^{d} K_j\left(\frac{x_j - X_{i,j}}{h_j}\right)$$

which permits different smoothings in each coordinate direction. Such multivariate kernel estimates are called product kernels. Even simpler estimates are attained using a fixed global smoothing parameter. If $h_j = h$, $j = 1,...,d$, then

$$\hat{f}(\underline{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{x_j - X_{i,j}}{h}\right) \tag{6.5}$$

With a single dimension, a subjective choice of smoothing parameter is influenced by the sample size and data variability. If h is too small the estimate presents itself as a series of probability peaks over the original observations. Alternatively, if h is too large information is lost through the severity of smoothing. With bivariate and multivariate density estimation selecting the window widths can be problematic. Optimal parameterizations depend on the criteria of optimization used and on the choice of kernel function (Hand (1981)).

Silverman's revival of the theoretical foundations of density estimation in 1986, rekindled interest in this area of study. Many workers have updated old, and developed new methodologies to address such problems. A variety of routes to achieve the most efficient smoothing parameters have been explored by experts such as Habbema *et al* (1974), and Terrell and Scott (1992), and a diverse range of optimization criteria's have been used. More recent publications have considered all possible classes of parameterizations to achieve optimal estimation (Wand and Jones (1993)). Kernel estimation methods have again been adopted in discriminant analysis for the purposes of medical diagnosis (Titterington (1981), Rossiter (1991), Boys (1992)).

6.3 Kernel functions and smoothing parameters

For each distinct data set, kernel density estimates are uniquely specified by the choice of smoothing parameters and kernel function adopted. Since many probability density functions are symmetrical and integrate to unity most behave suitably as kernel functions. Some univariate kernel functions are defined in Table 6.1. The kernel estimates derived from this class of functions are themselves densities and, unlike the more traditional histogram, are independent of origin choice. Gaussian kernels, being continuous with derivatives of all orders, have desirable analytical properties which are inherited by the estimate, although bounded kernels, such as piecewise kernels, may be computationally quicker since extreme points have density estimates of zero (Hand (1981)). Kernel estimates have attractive mathematical properties and their potential effectiveness in higher dimensions provides practical appeal in medical research and allocation theory (Habbema *et al* (1974), Rossiter *et al* (1991)).

The degree of smoothing in single dimensional kernel estimates is generally controlled by a fixed smoothing parameter, or window width h. Selection may be subjective or automatic. A subjective choice can be attained by plotting a series of density estimates with varying window widths. An over detailed density that forms probability spikes at each observation suggests h is too small. An oversmoothed estimate that takes the form of the original kernel, K, indicates h is too large. Figures 6.5-6.7 demonstrate the effects of varying the window widths of the kernel density estimate constructed using HCG residuals and a Gaussian kernel function, (see Table 6.1). With $h = 0.04$, the estimate shows too much spurious noise. Using $h = 0.12$ information is lost through oversmoothing. With $h = 0.08$ the fine detail is removed and the general trend of the distribution is clear. Illustrations of these types of effects are also offered in Silverman (1986) and Härdle (1991).

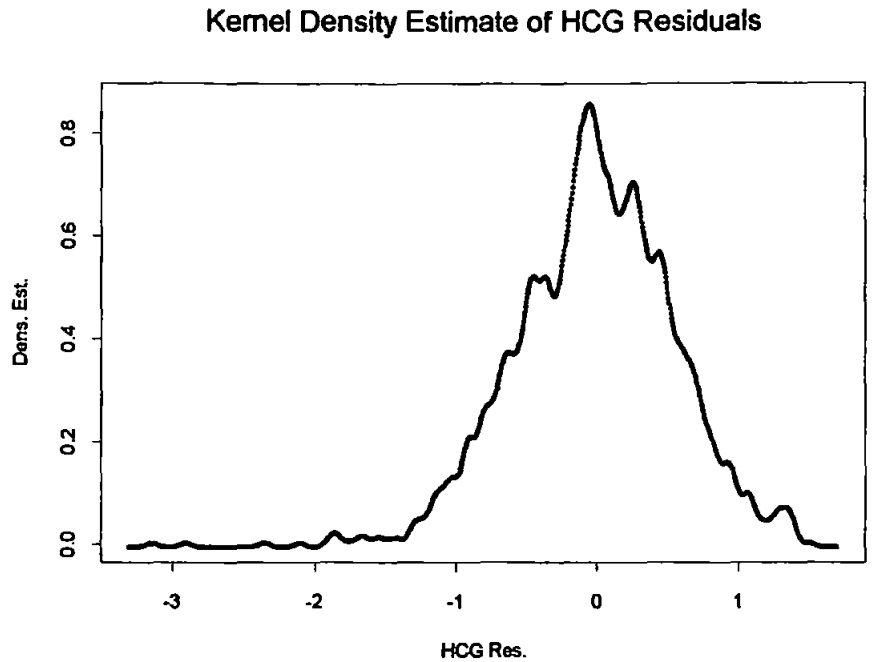

Figure 6.5: Kernel density estimate constructed from HCG residual data with $h = 0.04$.

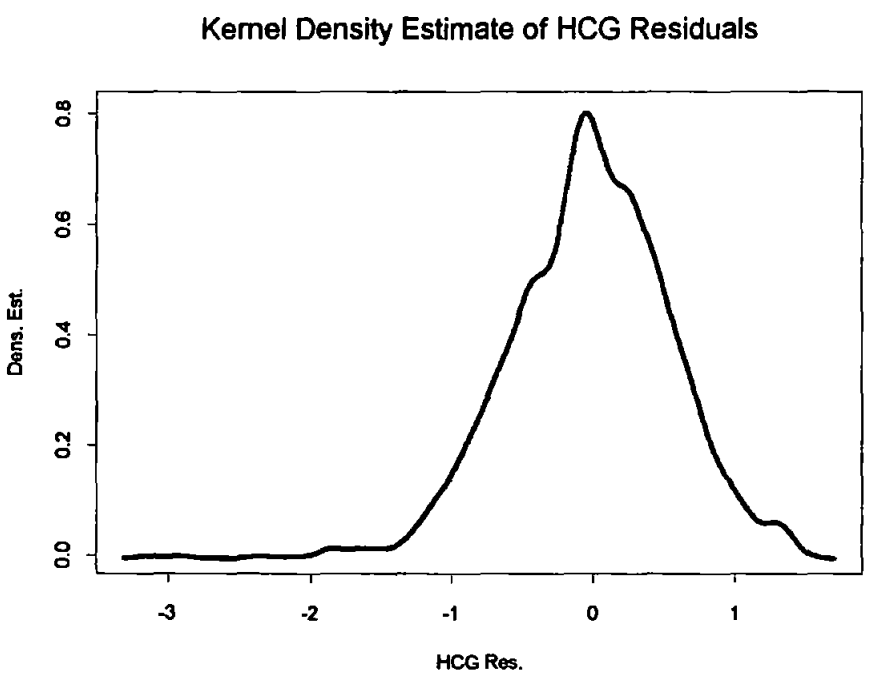

Figure 6.6: Kernel density estimate constructed from HCG residual data with $h = 0.08$.

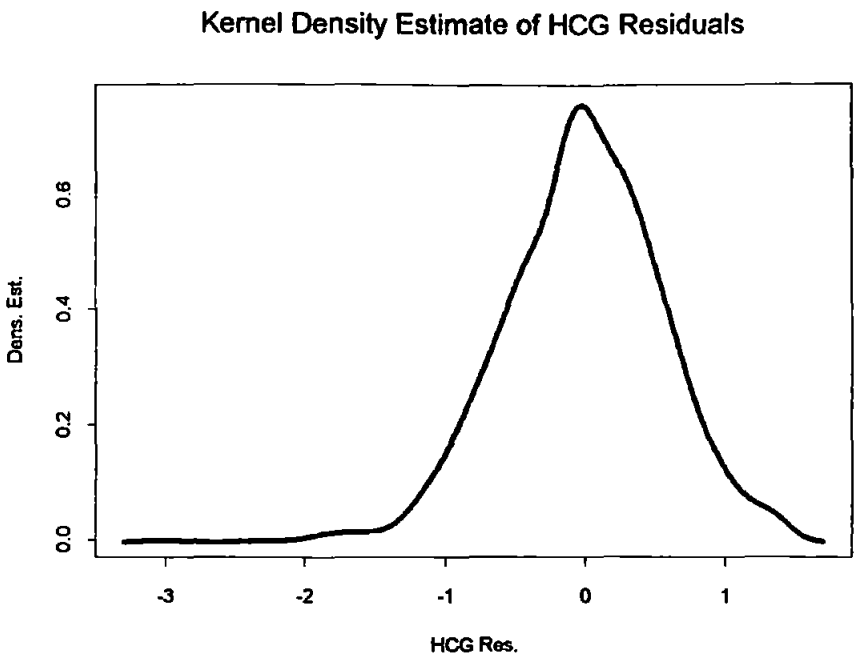

Figure 6.7: Kernel density estimate constructed from HCG residual data with $h = 0.12$.

Global smoothing parameters are less effective with data from a long tailed distribution. Disturbances are apparent in the extremes of the estimate. A variable window width that adjusts the degree of smoothing over regions of differing densities would achieve parameter estimates that were closer to their theoretical optimum. This technique has developed with advancing years of research (Terrell and Scott (1992)).

Automatic window width selection exhibits many forms and degrees of complexity. Efficient computational routes towards optimal estimation has captured much of the research involving density estimation. Automatic selection of smoothing parameters for bivariate and multivariate data sets has proved to be even more involved. Sample data from a mixture distribution may have optimal parameters in one class that may not be optimal in another (Marron and Wand (1992)). Again, the inconsistency of smoothing parameters to remain optimal over the global space of estimation causes concern (Terrell and Scott (1992)). More recent advances have introduced orientational smoothings as well as different dimensional smoothings to deal with correlated variables (Wand and Jones (1993)). The following section of this chapter reviews the significant publications that have lead to the key

developments in methods of kernel density estimation, dating from the original declaration by Fix and Hodges (1951).

6.4 Review

The initial proposal for an alternative approach to existing nonparametric methods of density estimation was offered by Fix and Hodges (1951). Fix employed the naive estimator to estimate the forms of unknown univariate densities for the purposes of discriminant analysis. Fryer (1977) later describes this estimator as a 'running histogram' that eliminates disputes over origin position. The advantages of such an estimate prompted Rosenblatt (1956) to investigate the generalized class of univariate estimators, defined as kernel or window estimators, with the kernel function K satisfying the conditions for a probability density function. Parzen (1962) discusses the asymptotic properties of this class and, by imposing further restrictions on the kernel function, illustrates asymptotic unbiasedness when the kernel K takes on particular densities. The adaptation of specific forms of kernel estimators to multivariate estimators was introduced by Cacoullos (1966) in the cases of Borel scalar kernels and product kernels. Desirable properties of more general multivariate density estimates with kernels of arbitrary form are set out by Epanechnikov (1969).

Much of the theoretical groundwork involving kernel density estimation is investigated in these publications. The underlying statistical properties provide a basis for the optimal selection of smoothing parameters and kernel functions. Silverman (1986) points out that most of the important applications of density estimation are concerned with multivariate data. However, since multivariate methods are generalizations of univariate methods it is necessary to develop an understanding of both. An overview of the statistical properties in their univariate form, and the modifications needed for extensions to multivariate density estimation follows.

6.4.1 Statistical aspects of the kernel method for univariate data

According to Hand (1981) an acceptable smoothing parameter attempts to find a satisfactory compromise between bias and random fluctuation. As given by Silverman (1986) for finite samples, the bias and variance of the estimator $\hat{f}(x)$ is defined

$$b[\hat{f}(x)] = E[\hat{f}(x)] - f(x)$$

$$= \int \frac{1}{h} K\left\{\frac{x-y}{h}\right\} f(y)dy - f(x) \qquad (6.6)$$

$$Var[\hat{f}(x)] = E[\hat{f}(x)^2] - (E[\hat{f}(x)])^2$$

and

$$Var[\hat{f}(x)] = Var\left(\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\right) = \frac{1}{nh^2} Var\left[K\left(\frac{x-y}{h}\right)\right]$$

$$= \frac{1}{nh^2}\left[\int K\left(\frac{x-y}{h}\right)^2 f(y)dy - \left(\int K\left(\frac{x-y}{h}\right) f(y)dy\right)^2\right]$$

$$= \frac{1}{nh^2}\int K\left(\frac{x-y}{h}\right)^2 f(y)dy - \frac{1}{n}\left\{b[\hat{f}(x)] + f(x)\right\}^2 \qquad (6.7)$$

The bias of $\hat{f}(x)$ depends on the window width h, and the choice of kernel function K. The bias does not directly depend on the sample size. However, if h is selected as a function of the sample size, n, the bias indirectly depends on n. Increasing the sample size alone will not succeed in reducing the bias. Approximate expressions for equations (6.6) and (6.7) are given in Silverman (1986) and written below.

$$b[\hat{f}(x)] = \frac{1}{2}h^2 f''(x)k_2 + o(h^2) \tag{6.8}$$

where $k_2 = \int x^2 K(x)dx \neq 0$

$$Var[\hat{f}(x)] = \frac{1}{nh} f(x)\int K(x)^2 dx + o(nh^{-1}) \tag{6.9}$$

The variance of the kernel estimator is nearly proportional to $(nh)^{-1}$, thus a reduction in variation is achieved through increasing h, which leads to an unavoidable increase in bias. The apparent 'trade off' problem between random and systematic error poses questionable debate over the most efficient criteria for selecting the smoothing parameter. A natural measure of discrepancy between the density estimator and the true density at a single point is the mean square error, MSE. An alternative settlement was first proposed by Rosenblatt (1956) that measured the global performance of $\hat{f}$ as an estimator of f. Rosenblatt (1956) employed the method of minimizing the mean integrated square error, $MISE$, to assess the global accuracy of the density estimate. This method combines the effects of both the bias and variance in the selection of h.

The mean square error of a point estimate, MSE, is defined

$$MSE[\hat{f}(x)] = E[\hat{f}(x) - f(x)]^2$$

$$= \left(E\,\hat{f}(x) - f(x)\right)^2 + Var(\hat{f}(x)) \tag{6.10}$$

and the mean integrated square error for global estimation, $MISE$ is

$$MISE[\hat{f}(x)] = E\int \{\hat{f}(x) - f(x)\}^2 dx$$

$$= \int \{E\,\hat{f}(x) - f(x)\}^2 dx + \int Var[\hat{f}(x)]dx$$

Equations (6.8)-(6.9) give the approximated integrated square bias and integrated variance as

$$\int b[\hat{f}(x)]^2 dx \approx \frac{1}{4}h^4 k_2^2 \int f''(x)^2 dx$$

$$\int Var[\hat{f}(x)]dx \approx \frac{1}{nh}\int K(x)^2 dx$$

yielding

$$MISE[\hat{f}(x)] = \frac{1}{4}h^4 k_2^2 \int f''(x)^2 dx + \frac{1}{nh}\int K(x)^2 dx \qquad (6.11)$$

Rosenblatt (1956) demonstrates that minimizing the *MISE*, equation (6.11), leads to an optimal choice of smoothing parameter which itself is a function of the unknown density and its derivatives. This led to Rosenblatt's (1956) disappointing theorem that for all continuous densities, there does not exist a uniformly unbiased estimator.

Parzen (1962) defines the optimal smoothing parameter found by minimizing the approximated *MISE* to be

$$h_{opt} = k_2^{-\frac{2}{5}}\left\{\int K(x)^2 dx\right\}^{\frac{1}{5}}\left\{\int f''^2(x)dx\right\}^{-\frac{1}{5}} n^{-\frac{1}{5}} \qquad (6.12)$$

Substituting equation (6.12) into equation (6.11) gives the approximate value of the *MISE* for optimal selection of h, and provides support in the optimal choice of kernel function K.

$$MISE(h_{opt}) = \frac{5}{4}c(K)\left\{\int f''(x)^2 dx\right\}^{\frac{1}{5}} n^{-\frac{4}{5}}$$

where the constant $c(K)$ is defined

$$c(K) = k_2^{\frac{2}{5}} \left\{ \int K(x)^2 dx \right\}^{\frac{4}{5}}$$

According to Epanechnikov (1969) the optimal choice of kernel function is given by

$$K(x) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{x^2}{5}\right) & -\sqrt{5} \le x \le \sqrt{5} \\ 0 & \textit{elsewhere} \end{cases} \qquad (6.13)$$

which later became known as the Epanechnikov kernel.

KERNEL	$K(x)$	EFFICIENCY
Epanechnikov	$\frac{3}{4}\left(1-\frac{1}{5}x^2\right)\Big/\sqrt{5}$ for $\lvert x\rvert < \sqrt{5}$ 0 otherwise	1
Biweight	$\frac{15}{16}\left(1-x^2\right)^2$ for $\lvert x\rvert < 1$ 0 otherwise	$\left(\frac{3087}{3125}\right)^{\frac{1}{2}} \approx 0.9939$
Triangular	$1-\lvert x\rvert$ for $\lvert x\rvert < 1$ 0 otherwise	$\left(\frac{243}{250}\right)^{\frac{1}{2}} \approx 0.9859$
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2}$	$\left(\frac{36\pi}{125}\right)^{\frac{1}{2}} \approx 0.9512$
Rectangular	$\frac{1}{2}$ for $\lvert x\rvert < 1$ 0 otherwise	$\left(\frac{108}{125}\right)^{\frac{1}{2}} \approx 0.9295$

Table 6.1: Some kernel density functions and their efficiencies relative to the Epanechnikov kernel estimator.

Parzen (1962) sets out the necessary conditions for estimates of the form in equation (6.3) to be asymptotically unbiased at all points x and for the probability density to be continuous. Kernels satisfying these conditions include the rectangular, triangular, Gaussian and Cauchy density functions. Table 6.1 gives a list of some kernel density functions along with their efficiencies.

6.4.2 Statistical aspects of the kernel method for multivariate data

The first multivariate extensions to the univariate kernel density estimators were set out by Cacoullos (1966) and Epanechnikov (1969). Cacoullos (1969) modifies Parzen's work (1962) to derive results concerning consistency, asymptotic unbiasedness, and bounds for bias and mean square error for estimation based on multivariate kernel functions. Epanechnikov (1969) discusses asymptotic properties and considers the optimal values of h and K, for an arbitrary choice of multivariate kernel function.

Results concerning the approximated bias, variance and the *MISE* of kernel estimates in higher dimensions, as given in Epanechnikov (1969) and revised by Silverman (1986), are stated for completeness.

Define $\alpha = \int x_1^2 K(\underline{x})\, d\underline{x}$ and $\beta = \int K(\underline{x})^2 d\underline{x}$

then

$$b[\hat{f}(\underline{x})] \approx \frac{1}{2} h^2 \alpha \nabla^2 f(\underline{x}) \qquad (6.14)$$

$$Var[\hat{f}(\underline{x})] \approx \frac{1}{nh^d}\beta f(\underline{x}) \tag{6.15}$$

and $$MISE[\hat{f}(\underline{x})] \approx \frac{1}{4}h^4\alpha^2\int\left(\nabla^2 f(\underline{x})\right)^2 d\underline{x} + \frac{1}{nh^d}\beta \tag{6.16}$$

The approximate optimal window width achieved through minimizing the *MISE* is given by

$$h_{opt}{}^{d+4} = d\beta\alpha^{-2}\left\{\int\left(\nabla^2 f\right)^2\right\}^{-1} n^{-1} \tag{6.17}$$

Again substituting equation (6.17) into equation (6.16) yields the approximate value of the *MISE* achieved with optimal smoothing parameters.

6.4.3 Methods of automatic window width selection

Multivariate density estimation permits several options for smoothing parameterizations. Possibilities extend to many classes of parameterizations. Distinct directional window widths are frequently more appropriate than a single global smoothing parameter (Epanechnikov (1969)). Kernel density estimates constructed from correlated variables may perform well with a matrix of window widths that provide orientations other than those of the coordinate direction (Deheuvels (1977)). Data transformations such as scaling, (equating the sample variances), and sphereing, (reducing the covariance matrix to unity), often enhances estimation (Fukunaga (1972)).

The use of many smoothing parameters adds flexibility to the density estimation but also increases the difficulty of parameter optimization. Automatic window width selection based on minimizing the *MISE* requires prior knowledge of the true density. Although selection may be subjective, this is particularly undesirable for estimation in higher dimensions. Several methods for automatic window width selection have been offered

suitable for both univariate and multivariate data. These include simple procedures such as Silverman's 'rule-of-thumb' approach and more complicated selection procedures employing cross-validatory methods based on either least squares assumptions or the classical maximum likelihood technique (Bowman (1984)). Alternative kernel approaches such as variable methods that incorporate nearest neighbour techniques and adaptive kernel methods have been utilized to deal with the problems encountered with data from a long tailed distribution (Terrell and Scott (1992), Sheather and Jones (1991)). These approaches will now be reviewed along with more recent innovations.

6.4.3.1 Rule of thumb method

A simple approach to select the most suitable smoothing parameter, other than by a purely subjective choice is to use a standard family of distributions to assign a value to the term $\int f''(x)^2 dx$ in equation (6.12) for univariate data and $\int \left(\nabla^2 f\right)^2$ in equation (6.17) for multivariate data. Silverman describes this approach as the rule-of-thumb method.

Consider the univariate case. Silverman (1986) uses the Gaussian distribution with variance σ^2 as an example. The standard normal density is defined ϕ and

$$\int f''(x)^2 dx = \sigma^{-5} \int \phi''(x)^2 dx = \frac{3}{8}\pi^{-1/2}\sigma^{-5} \approx 0.212\sigma^{-5} \qquad (6.18)$$

Now, using a Gaussian kernel and substituting equation (6.18) into equation (6.12) gives

$$h_{opt} = (4\pi)^{-1/10} \frac{3}{8}\pi^{-1/2}\sigma n^{-1/5} = 1.06\sigma n^{-1/5} \qquad (6.19)$$

σ can be estimated from the data and substituted into equation (6.19) to give the optimal smoothing parameter. Silverman points out that this approach works well if the population is normally distributed but may cause oversmoothing if the population is multimodal. Improved results are obtained when a more robust measure of spread is used. If the interquartile range, R, replaces σ, equation (6.19) becomes

$$h_{opt} = 0.79Rn^{-1/5} \tag{6.20}$$

The smoothing parameter in equation (6.20) gives better results with data from long tailed and skewed distributions. The adaptive estimate of spread

A= min(standard deviation, interquartile range/1.34),

instead of σ in equation (6.19), works well with unimodal densities and moderately well with bimodal densities. Silverman also suggests reducing the factor 1.06 in equation (6.19). With a Gaussian kernel

$$h = 0.9An^{-1/5} \tag{6.21}$$

gives a mean integrated square error within 10% of the optimum for the lognormal distribution with skewness up to 1.8 and many others. Silverman concludes that using equation (6.21) as a smoothing parameter works very well with many densities and is a good starting point for optimal parameter selection with others.

The smoothing parameter given in equation (6.21) is used to construct density estimates from AFP, UE3 and HCG residuals and these are given in Figures 6.8-6.10.

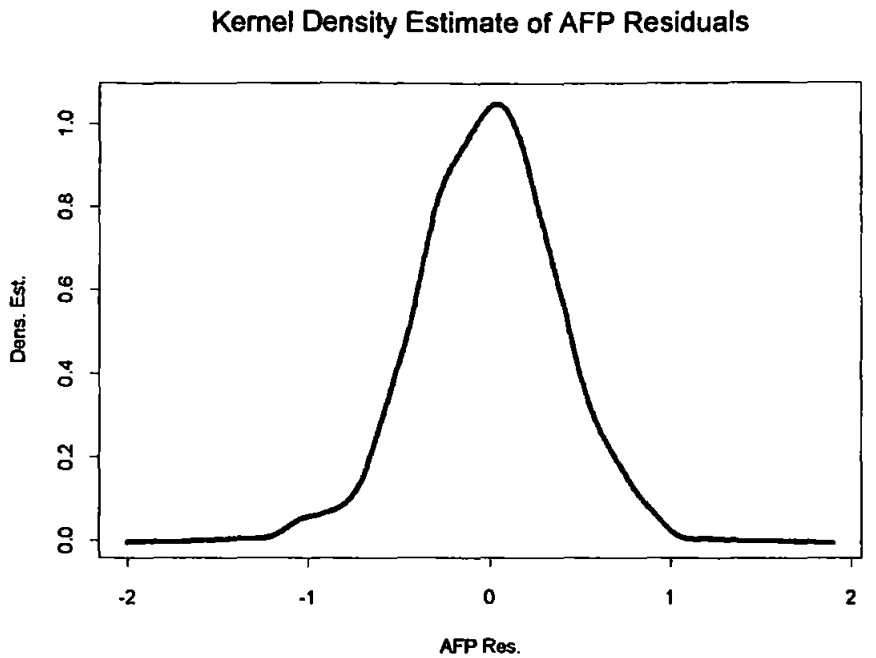

Figure 6.8: Kernel density estimate constructed from AFP residual data with the smoothing parameter determined by equation(6.21), $h = 0.08$.

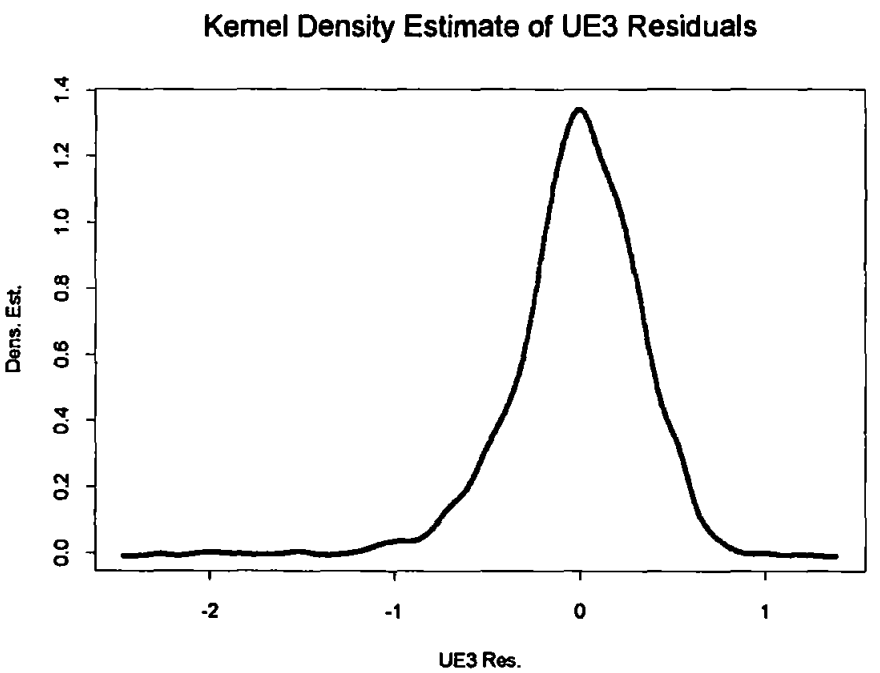

Figure 6.9: Kernel density estimate constructed from UE3 residual data with the smoothing parameter determined by equation(6.21), $h = 0.06$.

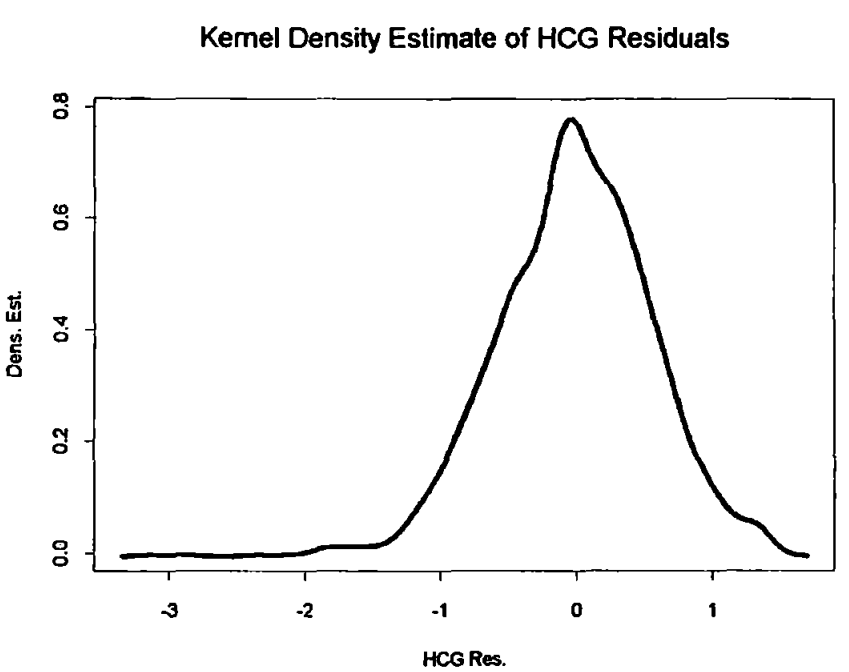

Figure 6.10: Kernel density estimate constructed from HCG residual data with the smoothing parameter determined by equation (6.21), $h = 0.1$.

The rule-of-thumb technique can be applied to multivariate data. Equation (6.17) can be used to compute a global window width when f is a standard density such as the multivariate normal. The multivariate Gaussian kernel function is defined as

$$K(\underline{x}) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\underline{x}^T \underline{x}\right) \tag{6.22}$$

and define ϕ as the unit d - variate normal density then

$$\int \left(\nabla^2 \phi\right)^2 = \left(2\sqrt{\pi}\right)^{-d}\left(\frac{1}{2}d + \frac{1}{4}d^2\right) \tag{6.23}$$

Substituting equation (6.23) into equation (6.17) gives the optimal window width, h_{opt}, for normally distributed data with unit variance

$$h_{opt} = A(K)n^{-1/(d+4)} \tag{6.24}$$

$$\text{where } A(K) = \left[d\beta\alpha^{-2}\left\{\int\left(\nabla^2\phi\right)^2\right\}^{-1}\right]^{1/(d+4)} \tag{6.25}$$

depends on the kernel function K. Silverman (1986) provides a list of the appropriate constants $A(K)$ with their respective kernels. With a general untransformed data set and a radially symmetrical kernel the value σh_{opt} is used as the window width, where σ is a single scale parameter. One suitable selection for σ is $\sigma^2 = d^{-1}\sum_i s_{ii}$ where s_{ii} is the appropriate entry in the covariance matrix.

Figures 6.11-6.13 show the bivariate density estimates constructed from the residual data using σh_{opt} where σ is the average marginal standard deviations of the appropriate analytes. The estimates perform well in the main body of the distributions but the tails of the distributions show spurious noise. This is particularly noticeable in Figure 6.11. The use of a global window width with data from a long tailed distribution leads to too much fine detail in the tails of the distribution. Increasing the window width is one way of dealing with this difficulty but this creates an oversmoothing problem in the main part of the distribution. This issue is addressed in section 6.4.3.3.

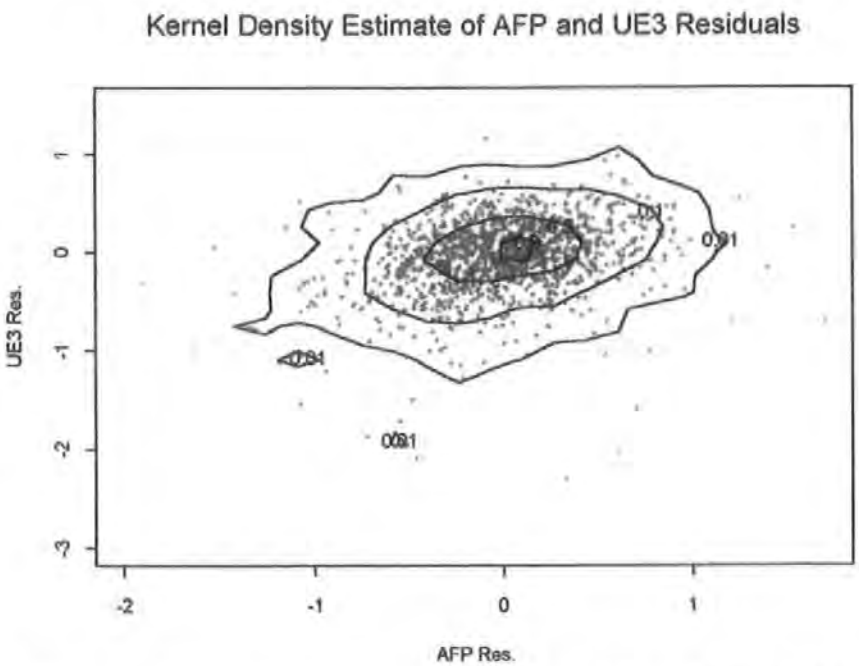

Figure 6.11: Bivariate Kernel density estimate constructed from AFP and UE3 residual data with the smoothing parameter determined by equation (6.24), $h = 0.10$.

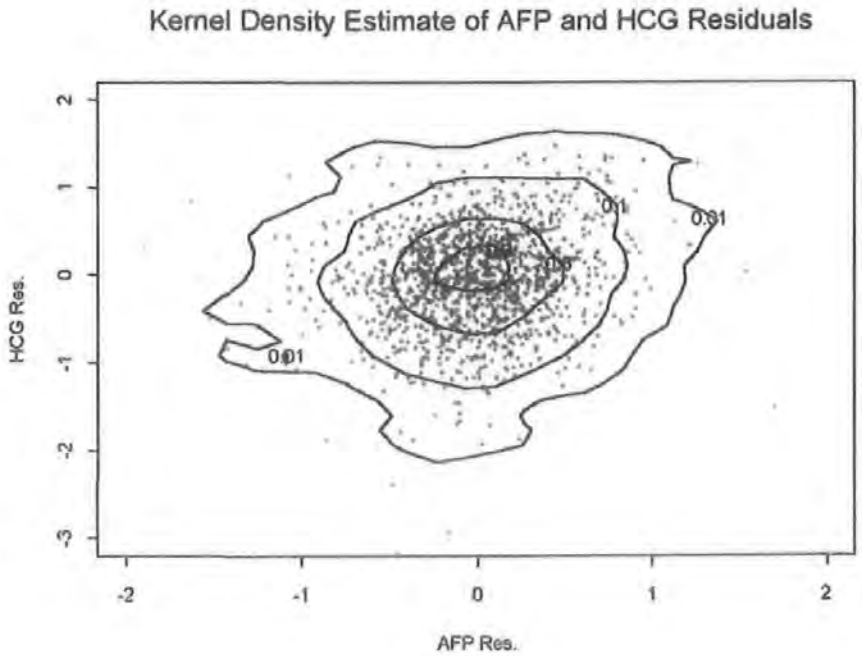

Figure 6.12: Bivariate Kernel density estimate constructed from AFP and HCG residual data with the smoothing parameter determined by equation (6.24), $h = 0.14$.

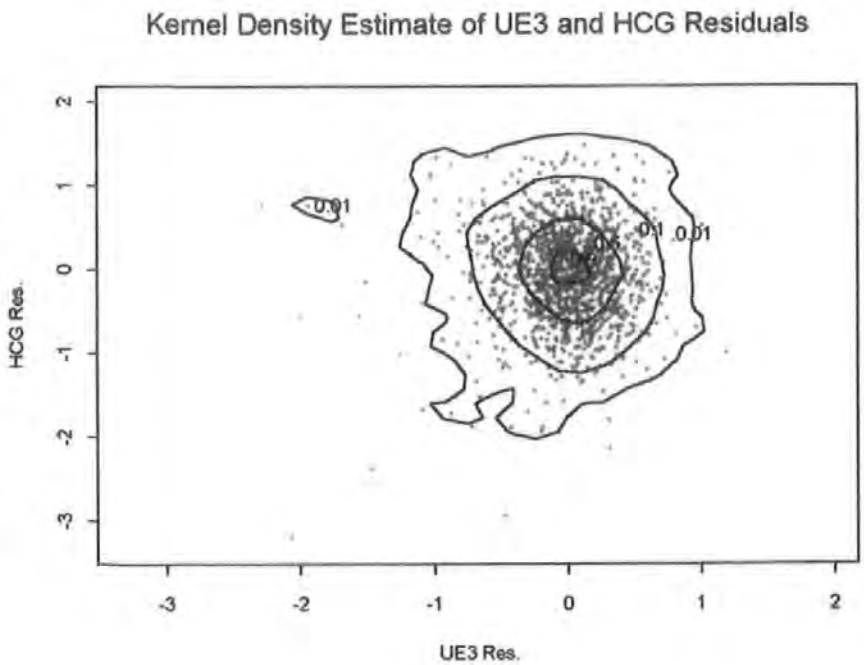

Figure 6.13: Bivariate Kernel density estimate constructed from UE3 and HCG residual data with the smoothing parameter determined by equation (6.24), $h = 0.13$.

6.4.3.2 Cross validation techniques

The inclusion of cross validation techniques with density estimation was initiated by Habbema *et al* (1974) as a means of deriving density estimates explicitly from sample data. Essentially the idea was developed from a modification of the classical maximum likelihood based principles of parameter selection. Optimal smoothing parameter estimates are selected using the criteria of minimizing an appropriate loss function derived from cross validatory approaches. Habbema *et al* (1974) use this technique to construct multivariate density estimates for the purposes of discriminant analysis. The optimal allocation rule and the most efficient combination of variables is achieved by minimizing an expected loss function which is evaluated by the leaving one out technique. By considering the reasoning of likelihood, the rationale behind the method becomes clear.

If an independent observation Y is available in addition to the independent observations $X_1, X_2, ..., X_n$, from the density f, the smoothing parameter is selected so as to maximize the likelihood, or log likelihood of Y belonging to the density $\hat{f}$ where $\hat{f}$ is regarded as a parametric family of densities depending on the window width h. The log likelihood becomes $\log(\hat{f}(Y))$. In the absence of such an observation, Y is replaced with one of the original observations, X_i, and a kernel density estimate is constructed from the remaining data points.

Define $\hat{f}_{-i}$ as the univariate density estimate constructed from all observations except X_i. Then

$$\hat{f}_{-i}(x) = \frac{1}{(n-1)h} \sum_{j \neq i}^{n} K\left(\frac{x - X_j}{h}\right) \qquad (6.26)$$

The cross validation function, $CV(h)$, is determined by averaging over each log likelihood function with one distinct observation X_i omitted.

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n} \log \hat{f}_{-i}(X_i) \tag{6.27}$$

By the following reasoning, Bowman (1984) demonstrates that selecting the maximizer of $CV(h)$ as the optimal smoothing parameter is equivalent to selecting the minimizer of the Kullback-Leibler loss function defined in equation (6.29).

Let $\hat{f}_{n-1}$ be a density estimate based on $(n-1)$ observations, then

$$E[CV(h)] = E[\log \hat{f}_{-n}(X_n)] = E\int f(x) \log \hat{f}_{n-1}(x)dx$$

$$\approx E\int f(x) \log \hat{f}(x)dx = -E\left\{I(f,\hat{f})\right\} + \int f \log f \tag{6.28}$$

where $I(f,\hat{f})$ is the Kullback-Leibler loss function and is defined

$$I(f,\hat{f}) = \int f(x) \log\left\{f(x)/\hat{f}(x)\right\}dx \tag{6.29}$$

illustrating that $-CV(h)$, up to a constant, is an unbiased estimator of the expected Kullback Leiber error function based on a sample size of $(n-1)$. Since the latter term of equation (6.28) is independent of h the expression is maximized by minimizing this error function, (equation (6.29)).

However, likelihood cross validation techniques are subject to criticism. Scott and Factor (1981) report reduced performance with data containing outliers. If a kernel function has bounded support and an observation is parted from any other observation by a distance

greater than h, $\hat{f}_{-i}$ becomes zero and $CV(h)$ will be infinity. The maximizer of $CV(h)$ must be large enough to overcome this difficulty, which again leads to possible over smoothings. Silverman (1986) points out that although using a kernel function with heavier tails may deal with this problem Chow *et al* (1983) report that this may lead to undersmoothed densities. Further, an infinite value of $CV(h)$ may be attained when the data contains identical observations (Härdle (1990)). Schuster and Gregory (1981) demonstrate that if the tails of f are eventually monotonic and die off at an exponential rate the cross validation method of estimation may lead to inconsistent estimates of the density since the distance between the outliers in the tails will not decrease as the sample size increases. Hall (1983) stresses the need for more robust estimation procedures.

Rudemo (1982) and Bowman (1984) investigate the chance to improve estimation using cross validation techniques of automatic selection. Satisfactory smoothing parameters are achieved through minimizing an estimate of a quadratic risk function which is based on the integrated square error, rather than the Kullback-Leibler information. The minimization process maintains computability through the risk function's detachment from the true density. Define the integrated square error (ISE) of any density estimate $\hat{f}$ as

$$ISE[\hat{f}] = \int (\hat{f} - f)^2 = \int \hat{f} - 2\int \hat{f} f + \int f^2 \qquad (6.30)$$

Since the latter term is independent of h it can be excluded from the minimization process. Let

$$R(\hat{f}) = \int \hat{f}^2 - 2\int \hat{f} f \qquad (6.31)$$

Least squares cross validation constructs an estimate of $R(\hat{f})$ from the data. This estimate is minimized over h to select the smoothing parameter.

Now, $\int \hat{f}^2$ is computable from the data. Consider the function $M(h)$ where

$$M(h) = \int \hat{f}^2 - 2n^{-1}\sum_i \hat{f}_{-i}(X_i) \tag{6.32}$$

Since

$$E\left[\frac{1}{n}\sum_i \hat{f}_{-i}(X_i)\right] = E\int \hat{f}(x)f(x)dx \tag{6.33}$$

substituting equation (6.33) into equation (6.32) gives

$$M(h) = \int \hat{f}^2 - 2\int \hat{f}(x)f(x)dx$$

Now $E[M(h)] = E[R(\hat{f})]$ and hence $M(h) - \int f^2$, for all h, is an unbiased estimator of the mean integrated square error and minimizing $E[M(h)]$ is equivalent to minimizing the *MISE*.

Stone (1984) shows that under mild conditions, asymptotically, least squares cross validation achieves the best possible choice of smoothing parameter in the sense of minimizing the integrated square error. The method is well justified in the cases of large samples. Silverman (1986) reports that Stone's theorem says that the score function $M(h)$ tells us, asymptotically, as much about the optimal smoothing parameter, from the integrated square error point of view, as if the underlying density f was known.

Both methods of parameter selection have multivariate extensions that are formed by simply replacing K with an appropriate multivariate function. Estimating densities in multidimensions amplifies the difficulties experienced from outliers with the likelihood cross validation techniques. These extreme points are less detectable and are more frequent in larger spaces. Although Habbema *et al* (1974) reported adequate results with kernels of unbounded support, problems such as these suggests the method of maximum likelihood cross validation is less adequate than the method of least squares cross validation when used to estimate multivariate densities (Silverman (1986)).

Bowman (1984) conducts a simulation study to examine the small sample properties of both cross validation methods. The study shows little evidence of any differences between the two approaches with bivariate data from normal or normal mixture distributions, although the Kullback Leibler cross validation is marginally superior to the least squares cross validation with data from a standard normal distribution. Least squares cross validation, however, outweighs the Kullback Leibler method, in terms of performance, with data from a long tailed distribution.

A further application of cross validation methods to density estimation was proposed by Scott and Terrell (1987). The paper introduces some biased cross validation criteria for the selection of smoothing parameters of kernel density estimates which is comparable to the unbiased method of least squares. Equation (6.11) defines the approximated $MISE$, for the univariate case which involves estimating $f''(x)^2$. The biased cross validation function proposes the most natural estimate to be $\hat{f}''(x)^2$, where $\hat{f}$ is a kernel estimator. Define the biased cross validation function as

$$BCV(h) = \frac{1}{4}h^4 k_2^2 \int \hat{f}''(x)^2 dx + \frac{1}{nh}\int K(x)^2 dx \qquad (6.34)$$

Now $Var[\hat{f}''(x)] = \text{var}\left(\frac{1}{h^3}\sum_{i=1}^{n} K''\left\{\frac{x-X_i}{h}\right\}\right) \approx \frac{1}{nh^5}\int K''(x)^2 dx$

Parzen (1962) suggests a sequence of bandwidths which are proportional to $n^{-1/5}$ should be used to minimize the $MISE$. If this choice of bandwidth is used for the optimization of $BCV(h)$ then $Var[\hat{f}''(x)]$ will not converge to zero and $BCV(h)$ cannot approximate the $MISE$. However Scott and Terrell (1987) provide a formula for the expectation of $\hat{f}''(x)^2$

when the kernel K and density f are at least twice continuously differentiable. The formula is

$$E\left[\int \hat{f}''(x)^2\right]dx = \int f''(x)^2\,dx + \frac{1}{nh^5}\int K''(x)^2\,dx + O(h^2)$$

which is correct for density estimates based on quartic, triweight or Gaussian kernels. The bias in equation (6.34) can be corrected by

$$\int f''(x)^2\,dx = \int \hat{f}''(x)^2\,dx - \frac{1}{nh^5}\int K''(x)^2 dx$$

which is asymptotically unbiased when $h \approx n^{-1/5}$. The biased cross validation estimator becomes

$$BCV(h) = \frac{1}{4}h^4 k_2^{\,2}\left[\int \hat{f}''(x)^2\,dx - \frac{1}{nh^5}\int K''(x)^2\,dx\right] + \frac{1}{nh}\int K''(x)^2 \qquad (6.35)$$

Scott and Terrell (1987) further show that $\hat{h}_{BCV} = \arg[\min BCV(h)]$ is asymptotically optimal in the case of minimizing the ISE. Simulation studies are conducted to examine small sample properties of the biased and unbiased estimators. The studies indicate the biased cross validation estimator oversmooths the density estimate and the unbiased estimator has a very large variance. Large gains in asymptotic efficiency are observed with sample sizes beyond 500-1000, with the exception of the lognormal density which requires several thousand observations to construct an adequate density estimate. More recent work involving the further development of cross validation based parameter selection is reported in Stute (1992) and Feluch and Koronacki (1992).

6.4.3.3 Variable and adaptive kernel methods

Most of the techniques for smoothing parameter estimation discussed so far cause a problem of oversmoothing in density estimates that are constructed from distributions with outliers. If h is selected to avoid a multiple peaked estimate in areas of low density information in the main body of the distribution is obscured. Oversmoothing inevitably leads to poor performance, particularly when constructing multivariate densities. The difficulty is caused by the use of a fixed global window width over regions of differing density. Variable and adaptive methods of window width estimation were introduced specifically to address this problem. The techniques allow h to vary over the domain of estimation so that any spurious effects with long tailed distributions are removed without masking vital information in areas of high density.

The variable kernel method, proposed by Loftsgaarden and Quesenberry (1965), is an extension of the nearest neighbour approach to density estimation. Silverman (1986) reviews the nearest neighbour method and its association with kernel methods. The estimate is constructed by standard kernel techniques, but the window width is free to vary with each point from which the estimate is taken, thus the degree of smoothing over regions of extreme mass is controlled. The generalized k^{th} nearest neighbour estimator and the variable kernel estimator are defined below.

Let K be a kernel function and k be a positive integer. Define $d_k(x)$ to be the Euclidean distance from x to the k^{th} nearest sample point in the set comprising of the other $(n-1)$ observations. The generalized k^{th} nearest neighbour estimator is defined as

$$\hat{f}(x) = \frac{1}{n d_k(x)} \sum_i K\left(\frac{x - X_i}{d_k(x)}\right) \tag{6.35}$$

$\hat{f}(x)$ is the kernel estimate evaluated at x with window width $d_k(x)$.

The variable kernel density estimator is defined as

$$\hat{f}(x) = \frac{1}{n}\sum_{i} \frac{1}{hd_{i,k}} K\left(\frac{x - X_i}{hd_{i,k}}\right)$$

where $d_{i,k}$ is the distance between the i^{th} observation and the k^{th} nearest sample point. The degree of smoothing depends on the local density of data and is proportional to $d_{i,k}$ ensuring that regions of low density have more widespread kernels than regions of higher density. Define the multivariate variable kernel estimator to be

$$\hat{f}(\underline{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{[hd_{i,k}]^d} K\left(\frac{\underline{x} - \underline{X}_i}{hd_{i,k}}\right) \qquad (6.36)$$

Breiman *et al* (1977) define an alternative estimator that also enables variable smoothing over the domain of estimation. In its general form, the adaptive kernel estimator consists of a fixed series of kernel functions K with varying window widths that depend on the point of estimation rather than the observed values.

An initial pilot estimate of the true density over each observed value, $\tilde{f}(\underline{x})$, is constructed that satisfies $\tilde{f}(\underline{X}_i) > 0 \ \forall \ i$. This requires the use of another density estimation method. Breiman *et al* (1977) suggest that since the method of adaptive smoothing is insensitive to the detailed information of the pilot estimate a natural choice of kernel function, in the multivariate case, would be the Epanechnikov kernel. Define local window width factors λ_i by

$$\lambda_i = \left\{\frac{\tilde{f}(\underline{X}_i)}{g}\right\}^{-\alpha}$$

where g is the geometric mean of $\tilde{f}(\underline{X}_i)$ and α is the sensitivity parameter satisfying $0 \le \alpha \le 1$. The univariate and multivariate adaptive kernel estimators are respectively defined as

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h\lambda_i} K\left(\frac{x - X_i}{h\lambda_i}\right) \tag{6.37}$$

$$\hat{f}(\underline{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d \lambda_i^{\,d}} K\left(\frac{\underline{x} - \underline{X}_i}{h\lambda_i}\right) \tag{6.38}$$

where K is an appropriate kernel function satisfying the conditions set out in section 6.1 and h is the smoothing parameter. Larger values of α lead to increased variation in the smoothing applied over the domain of estimation. The adaptive kernel estimate ensures that the constructed density is free from heavy tails. The estimate also has all the differential properties of the fixed width kernel estimate.

A special case of the adaptive kernel estimate is also defined by Breiman *et al* (1977) with $\alpha = 1/d$ where d is the dimensionality of the space in which the density is being estimated. Abramson (1982) proposes using $\alpha = 1/2$ for all dimensions, since this gives a density estimate with a bias of smaller order than that of the fixed width kernel estimate. Silverman (1986) illustrates the use of the adaptive kernel estimator with $\alpha = 1/2$ and reports great improvements in the performance of this estimator, in both the tails and the main part of the density, when compared to fixed width kernel estimators with data from a shifted lognormal distribution. Silverman concludes that the adaptive kernel estimator with $\alpha = 1/2$ is worth serious consideration when constructing densities that require more accuracy in the tails.

Since the residual analyte data is contaminated by outliers which produce longtailed effects in the distributions, the method of adaptive smoothing seems to be a suitable method to estimate the class conditional densities of residual values. Univariate and multivariate adaptive kernel density estimates are constructed from the analyte residual data. Abramson's choice of $\alpha = 1/2$ is selected as the sensitivity parameter and Silverman's rule of thumb method is applied to calculate values of h which are used in the construction of the pilot

estimate. Gaussian kernel functions are employed. Figure 6.14 plots the adaptive kernel density estimate constructed from the HCG residual data. This can be compared to the fixed width kernel estimate in Figure 6.10 where h is again selected by Silverman's rule of thumb method. Information is lost in the main part of the density estimate using the fixed width kernel approach and the tails of the estimate are over detailed. The adaptive kernel method smoothes over the detail in the tails of the density and illustrates more information in the main body of the density. Figures 6.15-6.17 illustrate bivariate kernel density estimates of the residual data using adaptive methods of smoothing. The performance of the adaptive kernel estimator outweighs the performance of the fixed width kernel estimator, the estimates of which are shown in Figures 6.11-6.13. Again, the fixed width estimator oversmooths the main body of the densities and illustrates spurious effects in the tails of the density. The spurious detail in the tails of Figure 6.11 are removed in Figure 6.15, and the improvement in the density estimate is dramatic. The problem of oversmoothing is particularly apparent in Figure 6.11. Figure 6.16 shows how the adaptive kernel estimator smoothes over the effects of noise in the tails of the density without losing detail in the main part of the density.

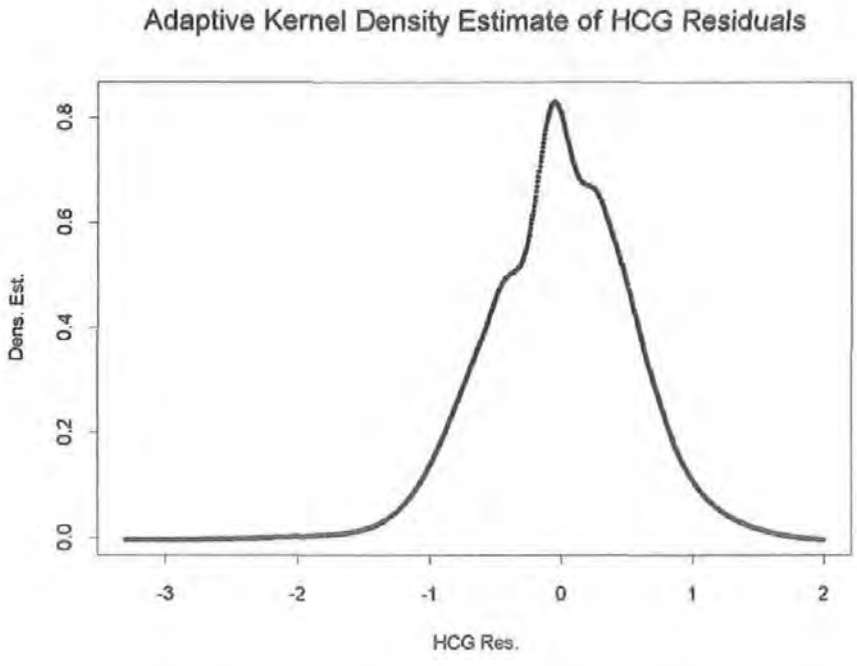

Figure 6.14: Univariate adaptive kernel density estimate constructed from HCG residual data

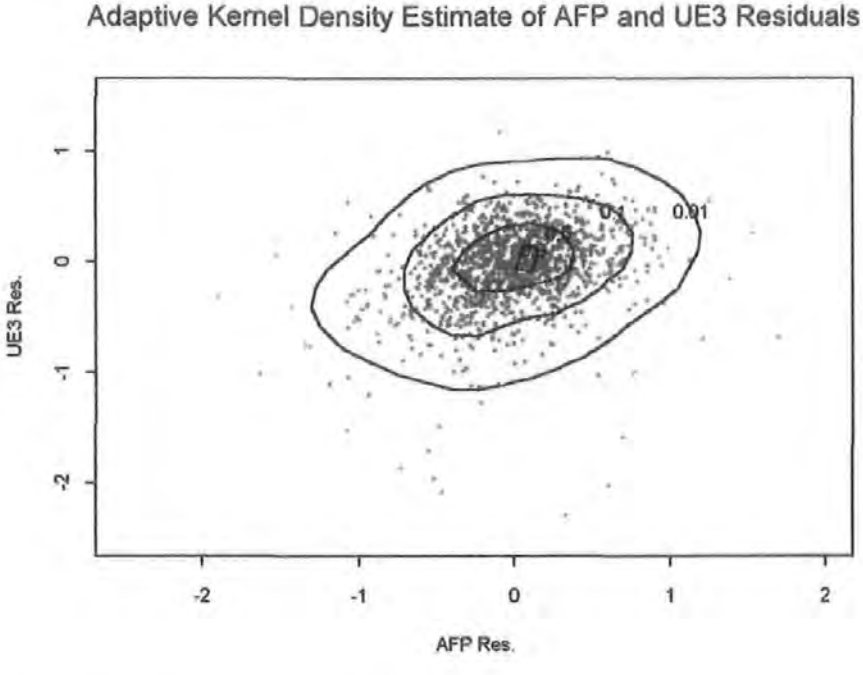

Figure 6.15: Bivariate adaptive kernel density estimate constructed from AFP and UE3 residual data.

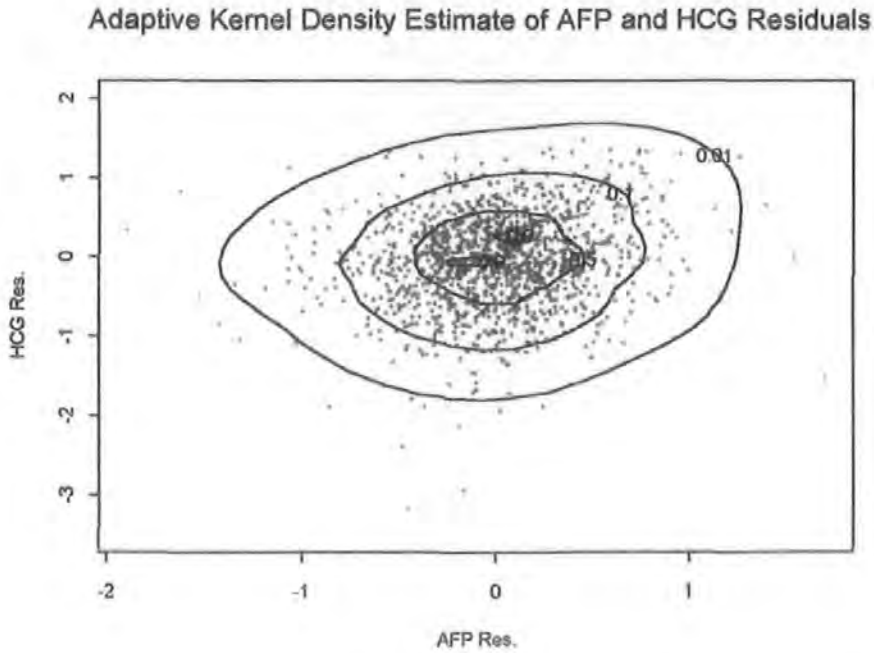

Figure 6.16: Bivariate adaptive kernel density estimate constructed from AFP and HCG residual data.

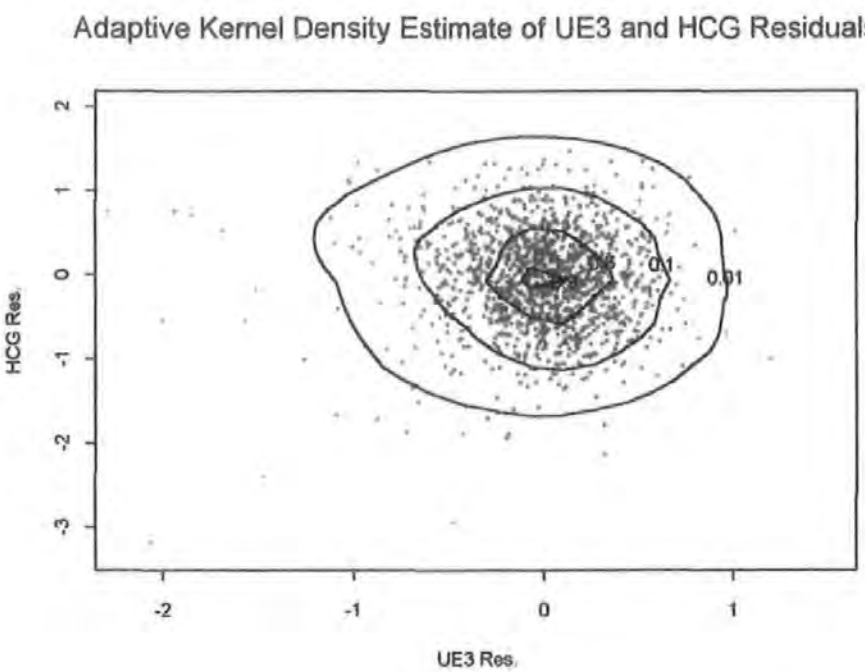

Figure 6.17: Bivariate adaptive kernel density estimate constructed from HCG and UE3 residual data.

Terrell and Scott (1992) investigated these estimators further to examine the possibility of improving univariate and multivariate kernel density estimates pointwise and

globally. The performance is based on minimizing the *MSE* and the *MISE* using asymptotic and finite sample results from simulated studies. The paper states that for pointwise estimation of the density, the window width of the kernel may depend only on the point of estimation, x, or the sample point X_i. These two approaches give rise to the balloon estimators, equation (6.39), and sample smoothing estimators, equation (6.40), respectively. The multivariate forms of these estimators are defined as

$$\hat{f}_1(\underline{x}) = \frac{1}{nh(\underline{x})^d} \sum_{i=1}^{n} K\left(\frac{\underline{x} - \underline{X}_i}{h(\underline{x})}\right) \tag{6.39}$$

$$\hat{f}_2(\underline{x}) = \frac{1}{nh(\underline{X}_i)^d} \sum_{i=1}^{n} K\left(\frac{\underline{x} - \underline{X}_i}{h(\underline{X}_i)}\right) \tag{6.40}$$

where $h(\underline{x})$ is the distance between the point of estimation, $\underline{x}$, and the k^{th} sample point, and $h(\underline{X}_i)$ is the distance between the $\underline{X}_i^{th}$ observation and the k^{th} sample point. Terrell and Scott (1992) report that nearest neighbour estimators are asymptotically of type $\hat{f}_1(\underline{x})$. When applied globally, this estimate does not integrate to unity so is not a probability density function. Moreover, there is little improvement, in common cases, with this type of estimator over the fixed kernel approach.

The sample smoothing estimator, akin to the Abramson estimator, and Breiman estimator in the asymptotic sense, is a mixture of identical but individually scaled kernels at each observation (Terrell and Scott (1992)). This estimator satisfies the conditions required for a probability density function , provided K is a density. In all dimensions though, the estimate at any point is highly influenced by extreme observations and not just by nearby points which increases in severity with increasing sample size.

6.4.3.4 Recent developments in kernel density estimation

Most recent developments to achieve optimal estimation have addressed the problem of estimating all possible parameterizations of the kernel density estimator in higher dimensions. The simplest multivariate kernel estimator has a fixed global bandwidth (Cacoullos (1966)). The next level of parameterization allows different bandwidths in each co-ordinate direction (Epanechnikov (1969)). Full parameterization enables a matrix of bandwidths, H, defined in equation (6.4), that permits smoothings in orientations other than those of the co-ordinate direction (Deheuvels (1977)). Wand and Jones (1993) consider the effects of different smoothings in each co-ordinate direction coupled with additional orientational parameters involving data scaling and sphereing. All possible classes of parameterizations for H which are used in the kernel density estimation of a bivariate density f are defined in Wand and Jones (1993) using the multivariate Gaussian kernel, equation (6.22). The classes are written below.

$$\mathcal{H}_1 = \left\{h_1^2 I : h_1 > 0\right\},\ \mathcal{H}_2 = \left\{diag(h_1^2, h_2^2) : h_1, h_2 > 0\right\},\ \mathcal{H}_3 = \left\{\begin{bmatrix} h_1^2 & h_{12} \\ h_{12} & h_2^2 \end{bmatrix} : h_1, h_2 > 0, |h_{12}| < h_1 h_2\right\}$$

(I is the identity matrix). Choosing $H \in \mathcal{H}_1$ means the kernel density estimate will always be spherically symmetrical, having circular contours. $H \in \mathcal{H}_2$ permits different smoothing parameters in each co-ordinate direction so the contours may be elliptical but with elliptical axes parallel to the co-ordinate axis. Full parameterization, $H \in \mathcal{H}_3$, produces elliptical contours with arbitrary orientation.

Wand and Jones (1993) consider other classes of parameterizations that involve scaling and sphereing the data (ℓ_1 and ℓ_2 respectively). These are defined as

$$\ell_1 = \{h^2 D : h > 0\},\ \ell_2 = \{h^2 C : h > 0\}, \text{ and } \gamma = \left\{ \begin{bmatrix} h_1^2 & \rho_{12} h_1 h_2 \\ \rho_{12} h_1 h_2 & h_2^2 \end{bmatrix} : h_1, h_2 > 0 \right\}$$

where C is the covariance matrix corresponding to the density f with (i, j) entry c_{ij}, ρ_{12} is the correlation coefficient of the density f, $D = diag(c_{11}, c_{22})$. Wand and Jones (1993) define γ as the hybrid parameterization. This class allows for sphereing and independent smoothings in each co-ordinate direction. Optimal window widths for 12 differing mixtures of normal distributions are derived by the criteria's of minimizing the asymptotic and exact mean integrated squared error. Each class of smoothing parameter is compared, in terms of efficiency, to the class $\hbar_3$. The results indicate the importance of different window widths for each co-ordinate direction, and show that generally this type of parameterization is sufficient. Arbitrary orientations may often be appropriate for densities with high amounts of curvature. Sphereing data to select orientation parameters is generally detrimental.

The preceding sections of this chapter reviewed most of the major publications involving nonparametric methods of kernel density estimation and illustrated some of these techniques by constructing one and two dimensional kernel density estimates of the residual data for unaffected pregnancies. The illustrations in these sections demonstrate the high performance of the adaptive kernel estimator which certainly performs better than the fixed kernel estimator, when used to construct univariate and bivariate density estimates from the trials data for unaffected pregnancies. In the following section, the adaptive kernel method is used to construct nonparametric density estimates of the residual AFP UE3 and HCG data for Down's syndrome pregnancies. The adequacy of the estimates are discussed along with circumstances that may affect the suitable use of these estimation methods.

6.5 Analysis of Down's syndrome residuals data

Nonparametric methods of density estimation are applied to the Down's syndrome residual AFP, UE3 and HCG data to provide estimates of the univariate and bivariate class conditional densities of logarithmic transformed MoM values. The adaptive kernel estimator that is discussed in section 6.4.3.3 is used to construct bivariate density estimates of the affected data that are then used to compute likelihood ratios.

Section 6.4 illustrates the benefits of using nonparametric methods to estimate the class conditional densities for unaffected pregnancies. Since there is a vast amount of data available for unaffected pregnancies reliable estimates can easily be formed in higher dimensions. However the rarity of data for affected pregnancies questions whether multivariate nonparametric methods of density estimation are suitable. Silverman (1986) discusses the required sample sizes for a given degree of accuracy of a nonparametric density estimate at a single point. Silverman considers estimating f at the point 0 when the true density is unit multivariate normal, the kernel function is normal and the window width h is chosen to minimize the mean square error at this point. Silverman provides the sample sizes required to ensure that the relative mean square error $E\left\{\hat{f}(0)-f(0)\right\}^2 / f(0)^2$ is less than 0.1. Some of these sample sizes are shown in Table 6.2. This kind of accuracy can be achieved in one, two, and three dimensions with relatively small sample sizes.

n	4	19	67	223	768	2790	10700
dim	1	2	3	4	5	6	7

Table 6.2: Sample sizes, **n**, required to ensure that the relative mean square error at zero is less than 0.1, when estimating a standard multivariate normal density in **dim** dimensions using a normal kernel and the window width that minimizes the mean square error at zero.

Univariate and bivariate kernel estimates of the densities for affected pregnancies are constructed from a sample of 93 pregnancies. Figures 6.18-6.20 plot histograms constructed from AFP, UE3 and HCG residuals. The longtailed effects in the distribution of UE3 are accentuated. Figures 6.21-6.23 illustrate fixed width univariate kernel density estimates constructed from the data. Silverman's rule of thumb method, given in equation (6.21), is used to calculate the smoothing parameters, h. The estimates generally show more spurious noise than those constructed from the larger samples of unaffected data. The multivariate rule of thumb method of calculating smoothing parameters, equation (6.24), is used to construct the bivariate density estimates in Figures 6.24-6.26. The fixed kernel approach demonstrates poor performance when used to construct density estimates from the affected data. Detail in the tails of the densities is exaggerated. Adaptive methods of smoothing are applied to the data, in Figures 6.27-6.29 to deal with this problem. The adaptive kernel estimates smooth out the fluctuations in the tails of the densities and illustrate more detail in the main body of the densities.

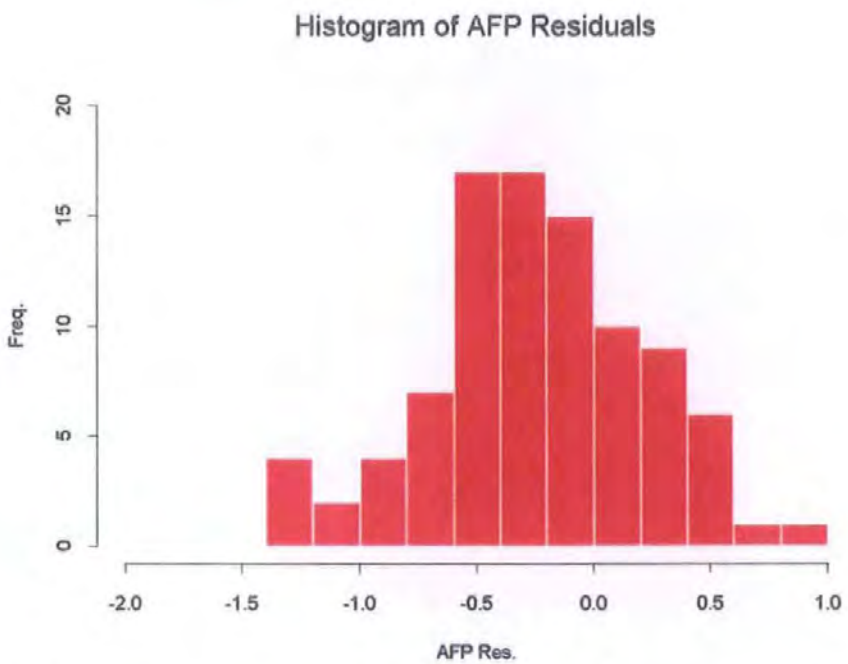

Figure 6.18: Histogram of AFP residuals for Down's syndrome pregnancies.

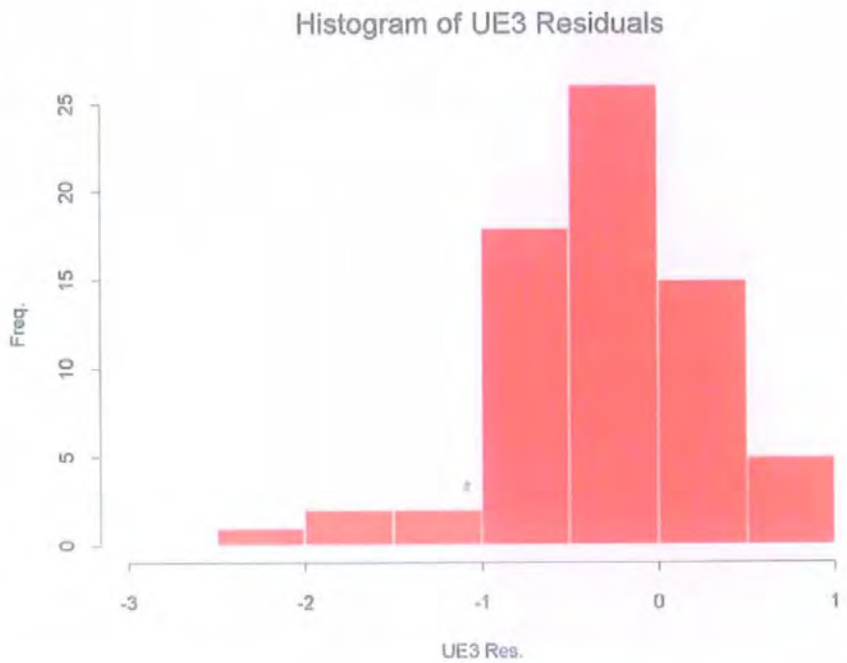

Figure 6.19: Histogram of UE3 residuals for Down's syndrome pregnancies.

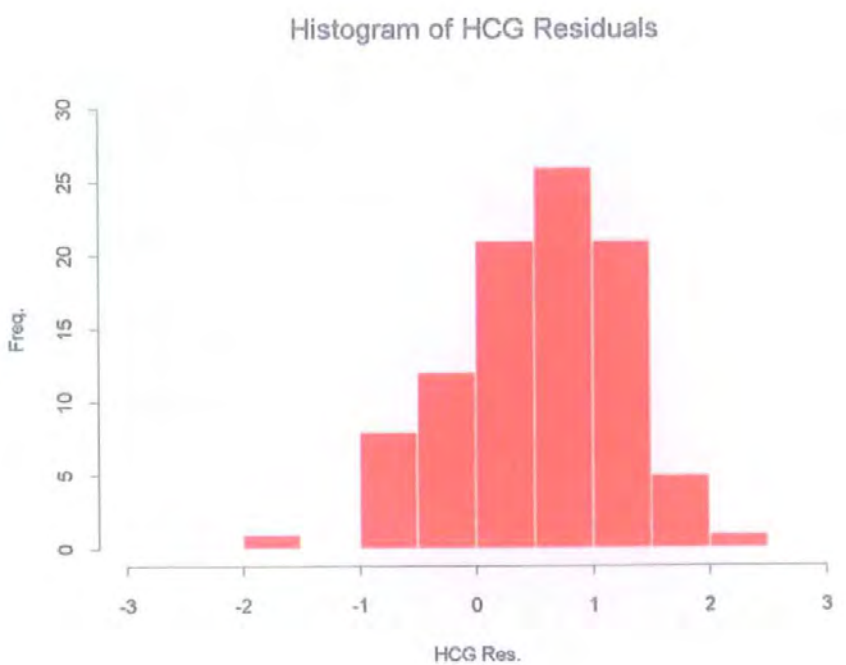

Figure 6.20: Histogram of HCG residuals for Down's syndrome pregnancies.

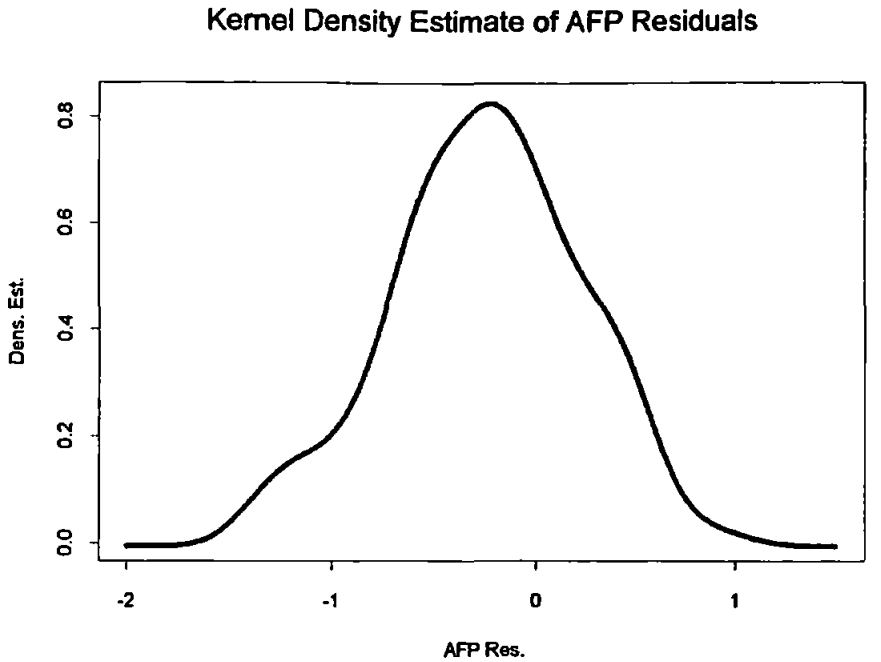

Figure 6.21: Kernel density estimate constructed from AFP residual data for Down's syndrome pregnancies with the smoothing parameter determined by equation (6.21), h=0.17.

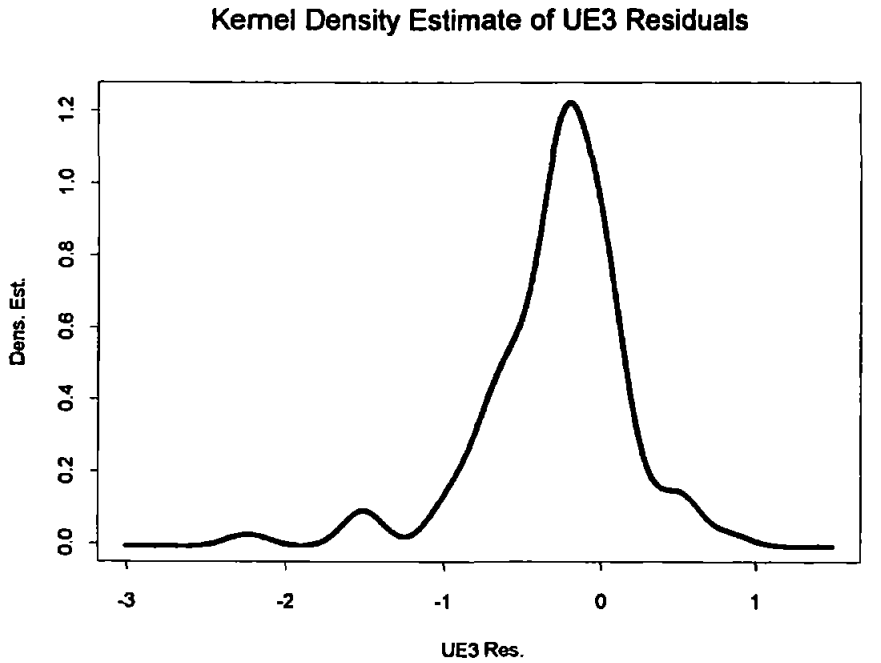

Figure 6.22: Kernel density estimate constructed from UE3 residual data for Down's syndrome pregnancies with the smoothing parameter determined by equation (6.21), h=0.12.

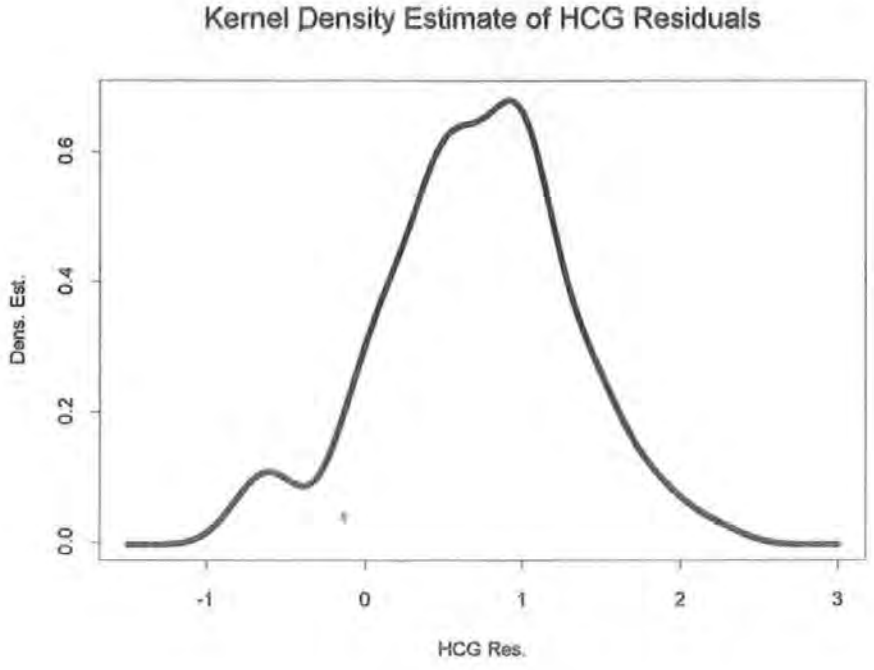

Figure 6.23: Kernel density estimate constructed from HCG residual data for Down's syndrome pregnancies with the smoothing parameter determined by equation (6.21), h =0.24.

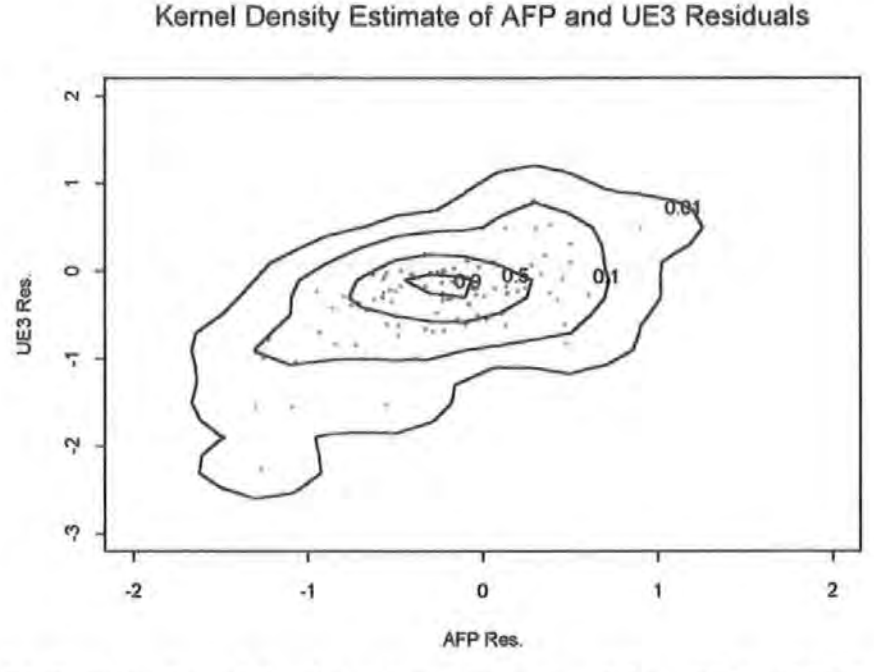

Figure 6.24: Bivariate kernel density estimate constructed from AFP and UE3 residual data for Down's syndrome pregnancies with the smoothing parameter determined by equation (6.24), h = 0.19.

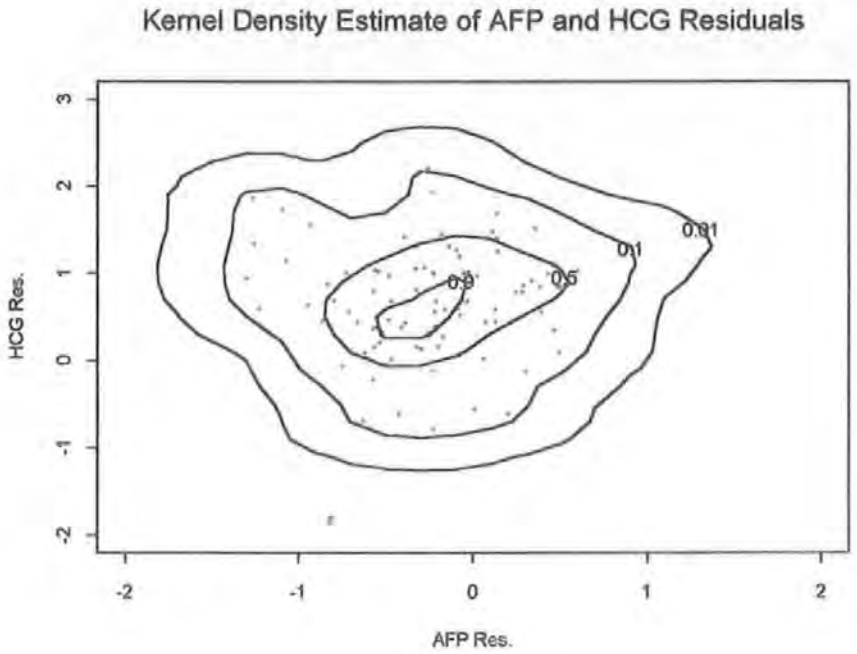

Figure 6.25: Bivariate kernel density estimate constructed from AFP and HCG residual data for Down's syndrome pregnancies with the smoothing parameter determined by equation (6.24), $h = 0.27$.

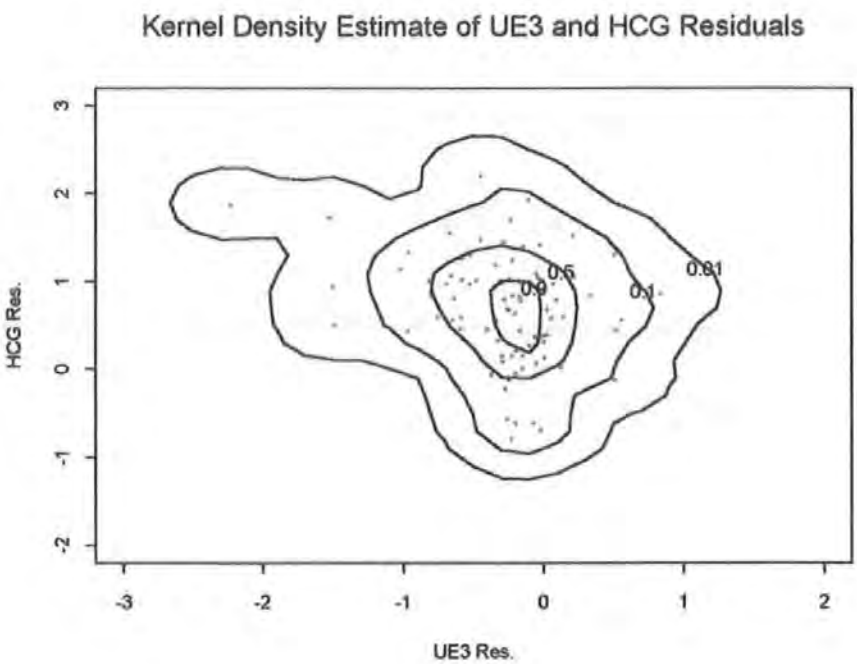

Figure 6.26: Bivariate kernel density estimate constructed from HCG and UE3 residual data with the smoothing parameter determined by equation (6.24), $h = 0.24$.

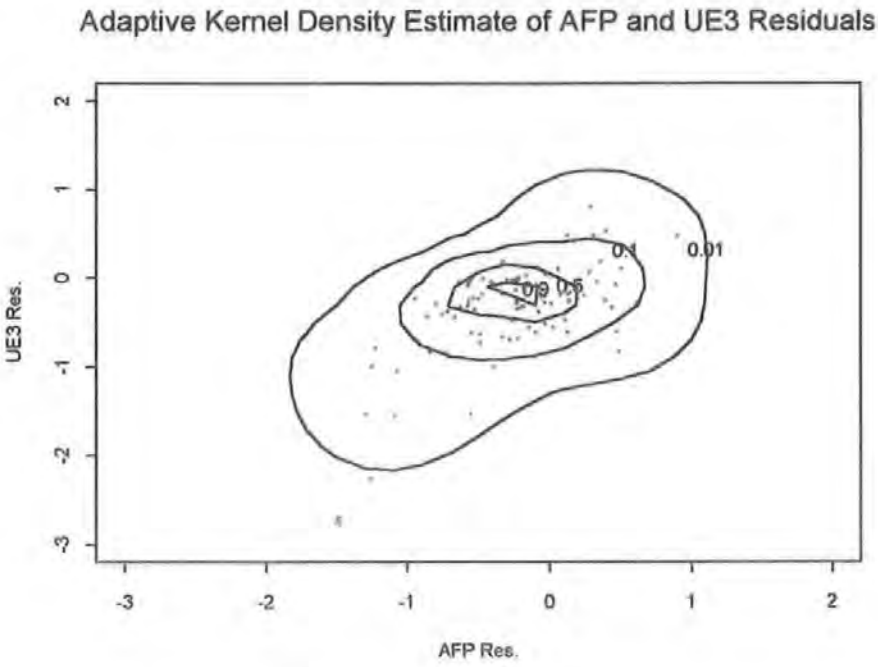

Figure 6.27: Adaptive kernel density estimate constructed from AFP and UE3 residual data.

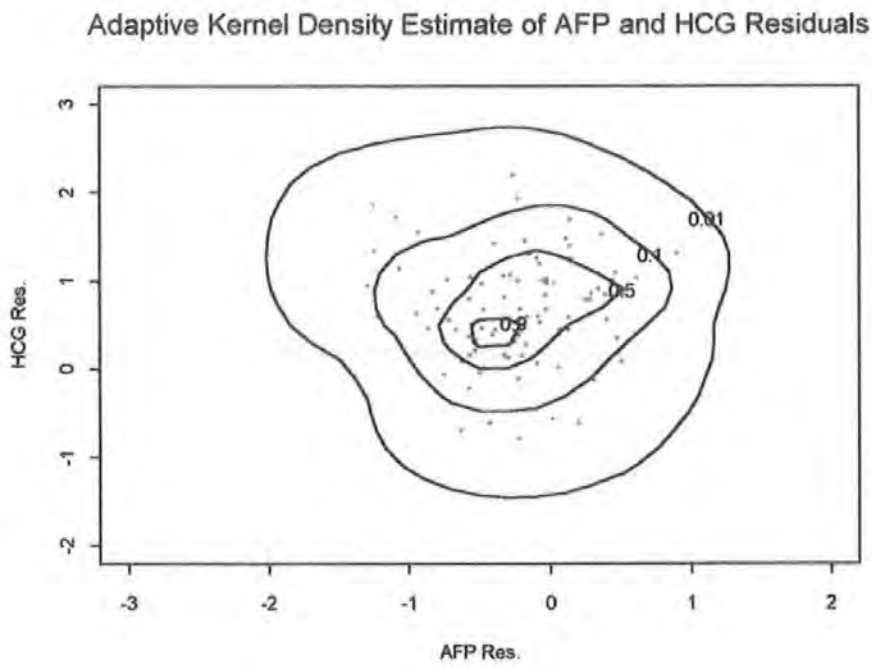

Figure 6.28: Adaptive kernel density estimate constructed from AFP and HCG residual data.

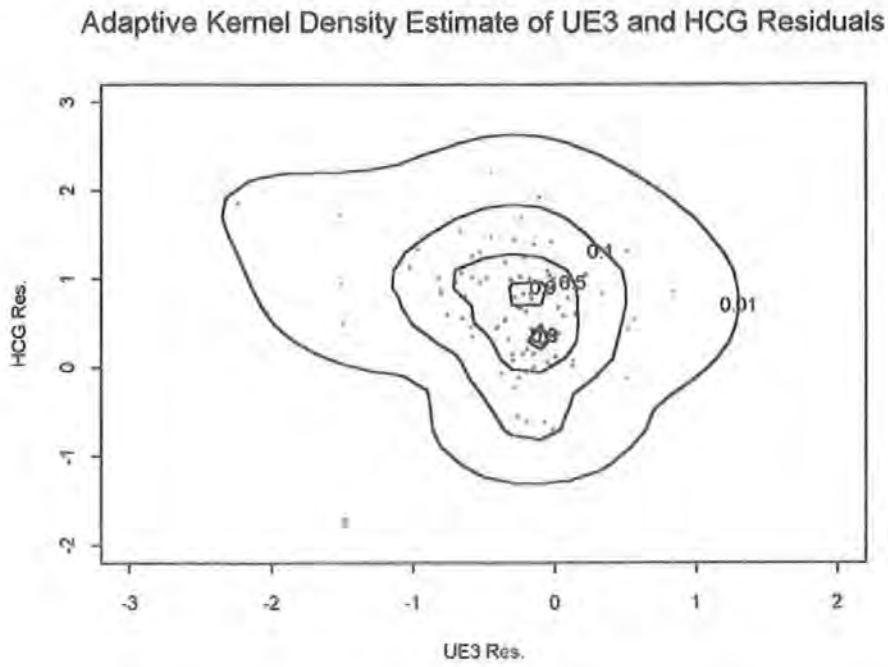

Figure 6.29: Adaptive kernel density estimate constructed from UE3 and HCG residual data.

The ratio of the class conditional density estimates of the trials data provides nonparametric likelihood ratios that can be used in the risk algorithm. Figures 6.30-6.32 illustrate contours of likelihood ratios of the bivariate class conditional densities, that are constructed using the adaptive kernel estimator, of unaffected residual data to affected residual data. The contours generally indicate the ratios expected for the analyte residual, or log(MoM) values given within the plotted range. However, the density estimates constructed from the affected data are based on a relatively small sample size and this prevents the estimation procedure achieving its optimal performance. This is reflected in the plotted contours. Although the bivariate density estimates that are constructed from this data are reasonably efficient the reliability of the density estimates and likelihood ratios formed in higher dimensions is questionable. Therefore, reliable estimates of risk can only be

constructed from nonparametric likelihood ratios that are formed from bivariate class conditional density estimates.

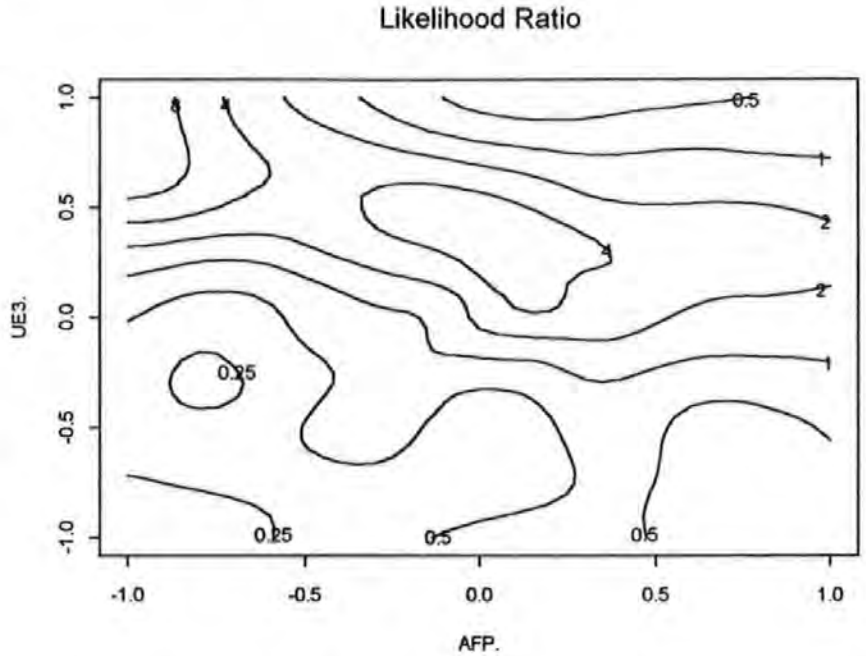

Figure 6.30: Likelihood ratio of fitted nonparametric densities constructed form AFP and UE3 residual .

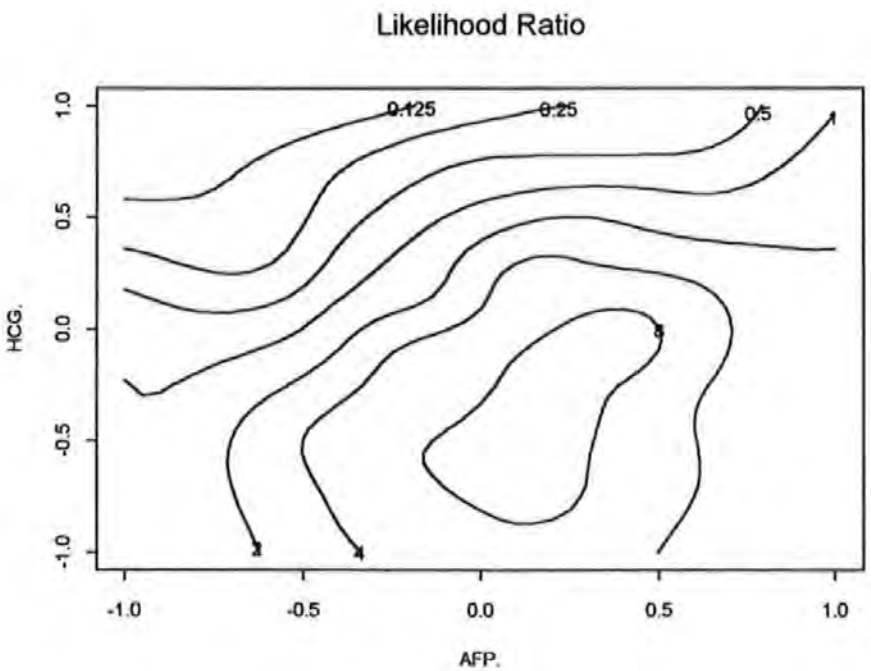

Figure 6.31: Likelihood ratio of fitted nonparametric densities constructed from AFP and HCG residuals.

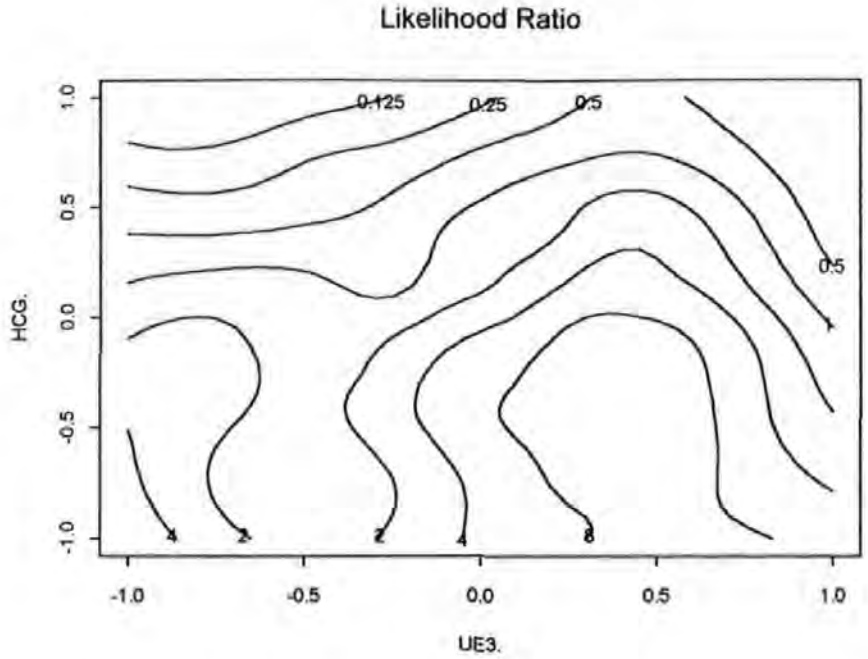

Figure 6.32: Likelihood ratio of fitted nonparametric densities constructed from UE3 and HCG residuals.

The problem of estimating class conditional densities when only relatively small samples are available from particular classes is addressed in Wright (1995). The report discusses the application of discriminant analysis to medical screening and investigates models in which the class conditional distributions are assumed to have a common distributional form. Wright applies nonparametric density estimation to model this common form, and uses parametric shifts in location and dispersion to model the differences between the classes. This technique is applied to Down's syndrome screening and the distributions of affected cases are represented as shifted and scaled versions of those of unaffected cases. Since there is a vast amount of data available for unaffected pregnancies, this approach can be used to provide higher dimensional models of the affected distributions which can then be used in the calculation of risk. The models given in Wright (1995) are discussed in the concluding chapter of this thesis.

In the following section the methods adopted by Wald *et al* (1988) to construct parametric density estimates are applied to the unaffected and affected distributions of residual analyte concentrations. These can be compared to the adaptive nonparametric estimates formed in this chapter. The parametric based likelihood ratios that are used in the risk algorithm are constructed from the trials data. The parameter estimates given in Wald *et al* (1992) and (1993) are shown in Table 3.1. In section 6.7, the detection rates based on the likelihood ratios that are constructed from the bivariate nonparametric density estimates using AFP and HCG residuals data are calculated and compared to those computed using the screening algorithm given by Wald *et al* (1988).

6.6 Parametric density estimation

The parameter estimates given in Wald *et al* (1992) are used to construct parametric density estimates of the trials data described in Chapter three. These can be compared to the parameter estimates derived under the models fitted in Chapter three, given in Table 4.1.

Figure 6.33 and Figure 6.34 plot the parametric density estimates constructed from the unaffected and affected residual AFP and HCG data respectively. In accordance with Wald *et al* (1988) and (1992), truncation limits are applied to the residual values. The application of truncation limits has a dramatic effect on the estimated densities. Figure 6.35 plots the contours of the parametric likelihood ratios of unaffected pregnancies to affected pregnancies. The contours are generally a lot smoother that those constructed from the nonparametric density estimates. However, the effects of applying truncation limits are clearly visible in the plot.

For comparison, Figures 6.36 and Figure 6.37 plot the parametric density estimates constructed from the unaffected and affected AFP and HCG residual data respectively

without applying truncation limits to the log(MoM) values. The contours are more precisely defined and are more representative of the true distributions of residual values.

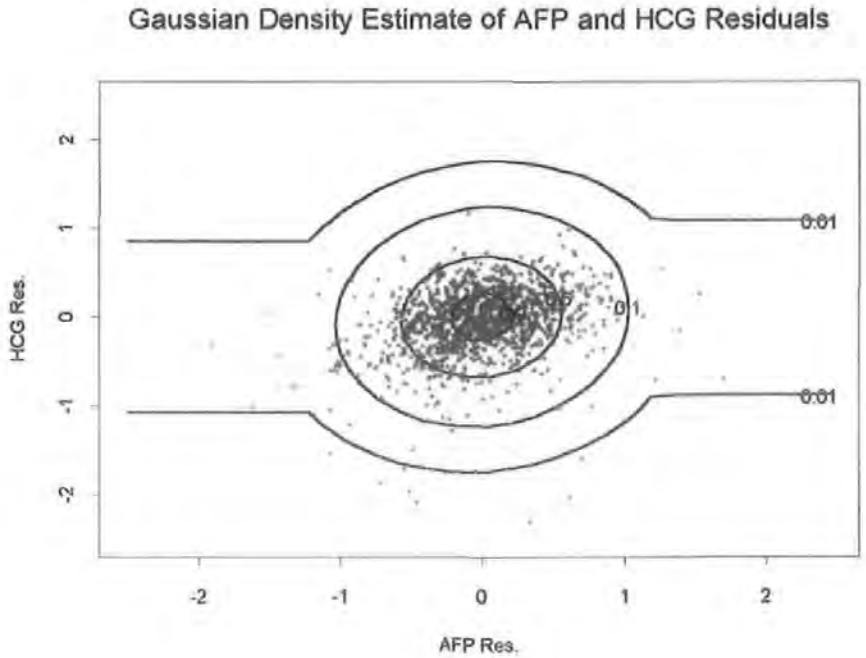

Figure 6.33: Gaussian density estimate of truncated AFP and HCG residual data for unaffected pregnancies.

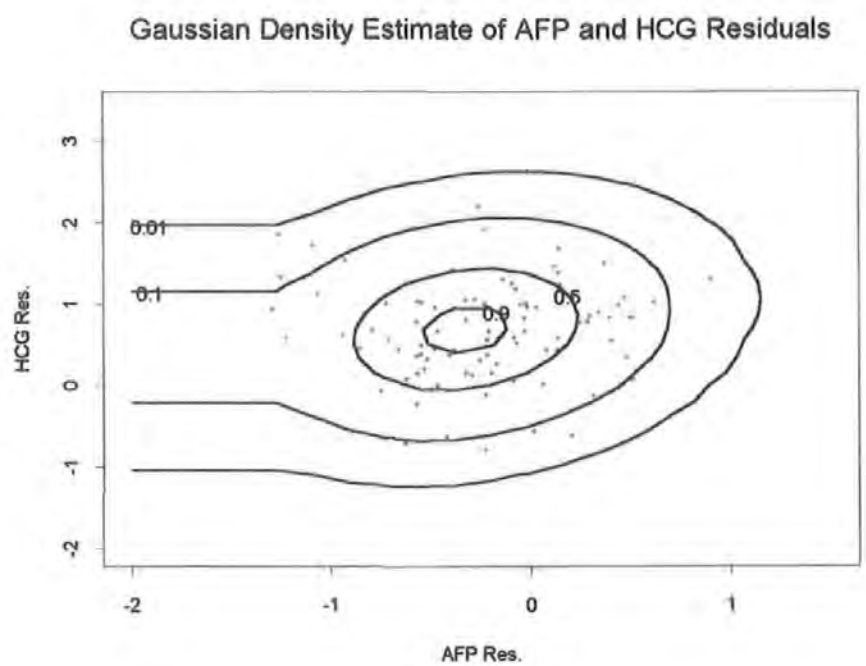

Figure 6.34: Gaussian density estimate of truncated AFP and HCG residual data for Down's syndrome pregnancies.

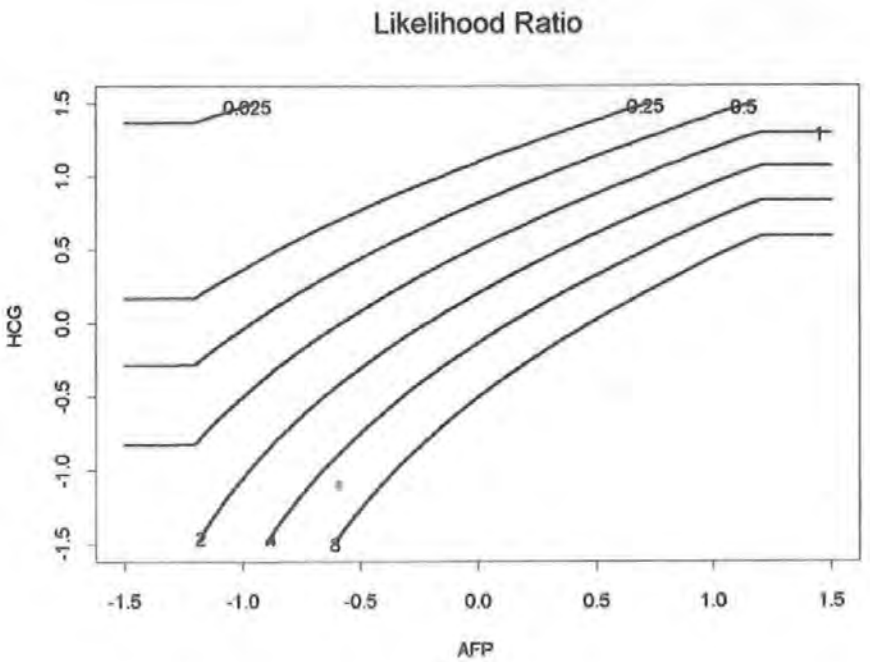

Figure 6.35: Likelihood ratio of fitted parametric densities constructed from truncated AFP and HCG residuals.

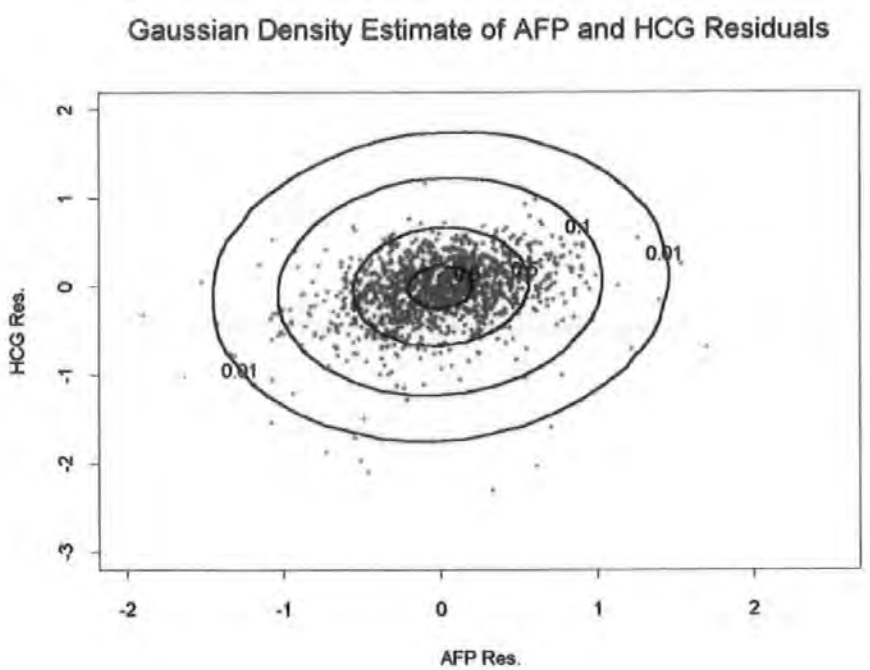

Figure 6.36: Gaussian density estimate of AFP and HCG residual data for unaffected pregnancies (without truncation).

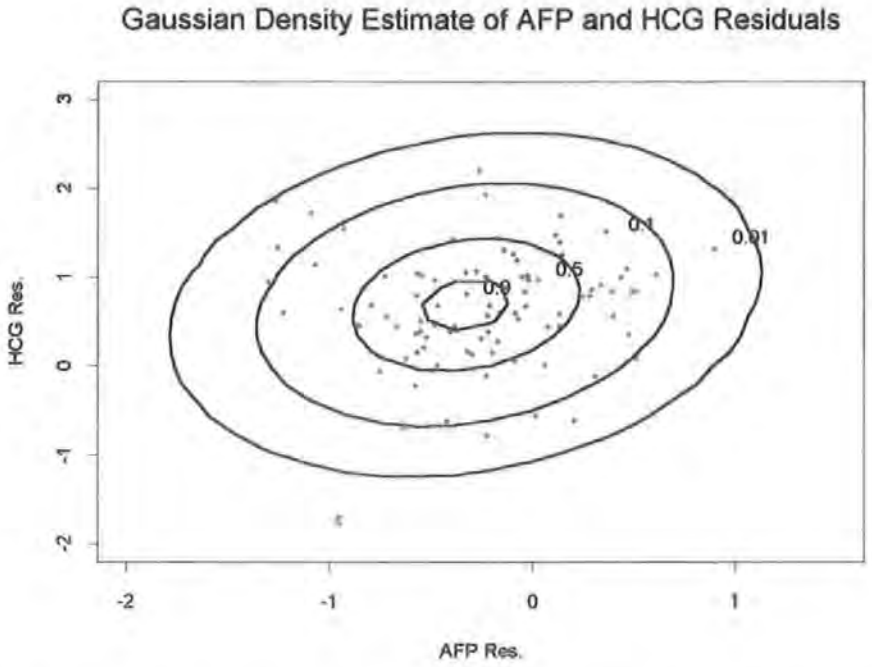

Figure 6.37: Gaussian density estimate of AFP and HCG residual data for Down's syndrome pregnancies (without truncation).

The following section compares the performance of the adaptive kernel density estimates with the parametric density estimates constructed using the approach given in Wald *et al* (1992). The conclusions of this chapter are summarized below.

6.7 Conclusion

This chapter proposes the use of nonparametric methods to estimate the densities of the class conditional distributions of transformed MoM analyte values. Various techniques for the construction of nonparametric univariate and multivariate density estimates are applied to the trials data described in chapter three. The adaptive kernel estimator provides high quality density estimates of the data for unaffected pregnancies. The problem of estimating nonparametric densities in higher dimensions from the relatively small samples of

affected data is discussed and one solution to this problem is set out in Chapter eight of this thesis.

The likelihood ratios computed from the bivariate adaptive kernel density estimates and from the parametric density estimates, that are based on the truncated distributions, of the AFP and HCG residual data, are combined with maternal age related risks which are derived from a standardized age distribution to provide estimates of detection rates and false positive rates. The rates are shown in Table 6.3.

	CUT-OFF	**DR**	**FPR**
Nonparametric Likelihood Ratios	1:230	60.0%	5.1%
Parametric Likelihood Ratios (with truncation limits)	1:310	57%	5.1%

Table 6.3: Detection rates and false positive rates based on nonparametric and parametric likelihood ratios constructed form AFP and HCG residual data. Wald *et al* (1988) report a detection rate of 54% at a false positive rate of 5% using analytes AFP and HCG in combination with maternal age.

The detection rate of the screening algorithm is greater when the parametric likelihood ratios are replaced with nonparametric estimates based on the adaptive kernel estimator using the residual data described in Chapter three. Both of the detection rates quoted in Table 6.3 are higher than the rate reported in Wald *et al* (1988), for a 5% false positive rate. However, as pointed out in Chapter four of this thesis, detection rates are prone to considerable sampling error, therefore, there is little difference between the detection rates when they were considered along with their standard errors. Moreover, detection rates that are based on nonparameteric methods of density estimation are more prone to bias than those based on parametric techniques. This is simply because the

nonparametric density estimates are constructed entirely from the original data values. The similarity in the performance of the two type of estimators is not surprising since, after removing the outliers, the residual data provides a fairly good approximation to a Gaussian form. The performance of screening algorithms that are based on data with distributions that showed large deviations from a Gaussian form, such as screens involving nuchal translucency, would benefit from the use of nonparametric density estimates. However, the study demonstrates that the adaptive kernel estimator is a strong competitor to the parametric estimator when applied to the trials data to calculate bivariate estimates. Further, it is believed that by combining the use of the adaptive kernel estimator with the use of the nonparametric shift model suggested by Wright (1995), higher quality density estimates can be constructed from the data for affected pregnancies and the performance of the screening algorithm based on nonparametric techniques can be improved.

Chapter 7

The effects of errors recorded in gestational dating methods on the current screening algorithm for Down's syndrome

7.1 Introduction

The screening algorithm described in Wald *et al* (1988) and MacDonald *et al* (1991) is based on the assumption that fetal age is recorded without error. However, as mentioned in Chapter one of this thesis, all methods for assessing gestational age are subject to errors. Imprecision in the recorded fetal ages may impact the estimated parameters and the performance measures given by the screening algorithm. In 1991 DiPietro and Allen published an article on child development. The article gives details of the procedures used to record fetal age. These procedures are reviewed in Section 7.2. The limitations of each procedure are also discussed.

To date, linear and non-linear regressions have been used to model the relationships between the analyte levels and recorded gestational ages. The form of the fitted regression equations published in Wald *et al* (1988) and (1992) were discussed in Chapter three. An alternative approach for modelling the relationship between gestational age and analyte concentrations is to use a model that assumes the explanatory variable is observed with error. There is a large amount of literature on these types of error-in-variables models, see for example Fuller (1980), Gleser (1981).

This chapter investigates the use of functional models in which the recorded gestational age comprises of the true gestational age plus an error term. Section 7.3 discusses how errors in gestational dates may provide misleading information concerning the outcome of pregnancy. Section 7.4 introduces the form of the functional model and the assumptions incorporated in the model. An expression is formulated for the difference between the covariance matrices that can be attributed to the error in the explanatory variable.

Section 7.5 applies the loglinear regression models that are fitted in Chapter three to estimate the standard deviations and correlations that would arise if the error in gestational age could be removed. Section 7.6 discusses the results of this analysis.

7.2 A review of the procedures for recording fetal age

It has been noted that each method for dating pregnancies records fetal age with an element of imprecision. Wald *et al* (1992a) discusses the changes in the performance of the algorithm when different gestational dating methods are used. In their trial, 12063 women were screened for Down's syndrome and the gestational ages were recorded by either ultrasound at the time of screening, (65%), or by LMP dating methods if no ultrasound was available. In the event of either methods being unavailable a clinical assessment was used. If a positive result was obtained by the screen for pregnancies whose gestational dates were recorded by LMP dating methods and the difference between dates and scans were as much as 17 days, the result of the screen was revised. The trial detected 12 out of 25 Down's syndrome pregnancies. According to Wald *et al* (1992a), the low level of detection was due to the selective use of ultrasound scans among women with positive results. An assessment

of the gestational dating methods are given in DiPietro and Allen (1991). This section reviews these procedures and their limitations.

Sonography, which involves an abdominal ultrasound scan, can be used to date pregnancies in the first and second trimesters of pregnancy. Different parts of the fetus are measured and these are compared to gestational age standards. Such measurements include fetal crown rump, biperietal diameter of the head, femur length and chest diameter. Before sonography was introduced, gestational dating relied on a clinical assessment, such as evaluating the size of the uterus. This type of assessment is most accurate in early pregnancy.

Dating by the last recorded menstrual period calculates the expected date of birth from the first day of the last menstrual period, including the two weeks prior to conception. Forty weeks are projected from this point to determine the estimated date of confinement.

Each method of estimating gestational age has its own degree of error. According to DiPietro and Allen (1991) accuracy of early second trimester dating by sonography has been accomplished using confirmation of ovulation by basal temperature records. Ultrasound is most reliable in first and early second trimesters as fetal growth rates are more invariant in early pregnancy (Campbell *et al* (1985)).

The accuracy in LMP dating methods can be impaired by irregular menstrual cycles, the use of oral contraceptives and unrecognised spontaneous abortions. The main problem with LMP dating is caused by the degree of inaccuracy in the dates recalled by the mothers.

7.3 The effects of the errors in gestational dating methods on the performance of the risk algorithm.

It is now well established that in unaffected pregnancies the levels of HCG decrease with advancing pregnancy whilst the levels of AFP and UE3 increase. Also higher levels of HCG and lower levels of AFP and UE3 are associated with Down's syndrome (Wald *et al* (1988), Merkatz *et al* (1984), Canick *et al* (1988)). However errors in the recorded gestational dates filter through to the transformed MoM values and this can subsequently effect the classification of the pregnancy. An example of this effect can be illustrated using the analytes HCG and AFP: If the recorded gestation of a pregnancy is younger than the true gestation, the mother's HCG concentration will be lower than that predicted and her AFP concentration will be higher. This will affect the log(MoM) values associated with the pregnancy and the estimated risk will be reduced. If Down's syndrome pregnancies have sufficiently large enough errors of this type in their gestational dates, the estimated detection rate of the algorithm could be reduced. Alternatively, if the recorded gestation is more advanced than the true gestation, then the recorded HCG concentrations will be higher than that predicted and the AFP concentrations will be lower. These type of errors will have the effect of increasing the mother's risk of Down's syndrome. If the error is large and the pregnancy is actually unaffected by any abnormality, it could be classified as Down's syndrome.

Consider the further effects on the risk distribution. The 5% of pregnancies with the highest risks are screened positive and referred for an amniocentesis. Some of these pregnancies may be unaffected but may have risks that are contaminated by gestational age errors. Therefore, it is possible that the false positive rate amongst women with accurate

dates is lower than the rate amongst women with inaccurate dates. Moreover, with LMP based gestational ages, it could be argued that women with accurate dates suffer as a consequence of those who have inaccurate dates. By removing these errors, the risks are reduced and fewer women are referred for an amniocentesis. In order to maintain a 5% screen positive value, the selected cut-off on the risk would need to be decreased. One consequence of this is that the detection rate would increase.

7.4 The functional and structural model

It has been suggested that models with errors-in-variables may be appropriate for the regression of the logarithmic analyte values on gestational age. Two forms of these errors-in-variables models are the functional and structural models. These models assume the variables in the regression are observed with an element of random error (Fuller (1980)). The form of the model relates the correct values of the response variable ($Y_{i,j}$) and the explanatory variable (X_i) by a linear regression. If the X_i's are regarded as fixed, the model is a functional model. Alternatively, if the X_i's are drawn independently from a population with mean μ_x and finite variance σ_x^2 the model is a structural model (Nyquist (1987)).

This section uses a simplified model that assumes the log analyte values are linearly related to gestational age. Denote the true gestational age for subject i , by X_i and the recorded gestational age by x_i. The true log transformed analyte values for subject i and analyte j are denoted by $Y_{i,j}$ and the recorded log analyte values by $y_{i,j}$. In the discussion presented below, the values of α_j and β_j are assumed to be known.

The true values of the variables $(X_i, Y_{i,j})$ are related through

The true values of the variables $(X_i, Y_{i,j})$ are related through

$$Y_{i,j} = \alpha_j + \beta_j X_i$$

The observed pairs $(x_i, y_{i,j})$ have errors of the form η_i and $\epsilon_{i,j}$ such that

$$x_i = X_i + \eta_i$$

$$y_{i,j} = Y_{i,j} + \epsilon_{i,j}$$

The errors-in-variables model assumes the errors $(\eta_i, \epsilon_{i,j})$ are uncorrelated with mean zero and have finite variances σ_η^2, and $\sigma_{\epsilon_j}^2$ respectively. The pairs $(X_i, Y_{i,j})$ are independent of the errors $(\eta_i, \epsilon_{i,j})$. X_i has mean μ_X and variance σ_x^2.

The log(MoM) values that are based on true gestational age are defined as

$$\epsilon_{i,j} = y_{i,j} - Y_{i,j}$$

$$\epsilon_{i,j} = y_{i,j} - (\alpha_j + \beta_j X_i) \qquad i = 1,2,..,n \quad j = 1,2,..,p$$

The log(MoM) values, $e_{i,j}$, that are based on recorded gestational age are given as

$$
\begin{aligned}
e_{i,j} &= y_{i,j} - (\alpha_j + \beta_j x_i) \\
&= (\alpha_j + \beta_j X_i + \epsilon_{i,j}) - (\alpha_j + \beta_j (X_i + \eta_i)) \\
&= \epsilon_{i,j} - \beta_j \eta_i
\end{aligned}
$$

Hence the recorded log(MoM) values are the true values plus an additional error, $-\beta_j \eta_i$. The variances and covariances of the recorded log(MoM) values contain components that result from gestational age error. Therefore

$$
\begin{aligned}
V(e_{i,j}) &= V(\epsilon_{i,j} - \beta_j \eta_i) \\
&= \sigma^2_{\epsilon_j} + \beta_j^2 \sigma^2_{\eta} \qquad (7.1)
\end{aligned}
$$

$$
COV(e_{i,j}, e_{i,k}) = COV(\epsilon_{i,j}, \epsilon_{i,k}) + \beta_j \beta_k \sigma^2_{\eta} \qquad (7.2)
$$

Thus the effect of errors in gestational age inflate the standard deviations of the log(MoM) values and bias the correlation coefficients. The following section uses the differences in the parameter estimates, given in the expressions in equation (7.1) and equation (7.2), to investigate the effect of the bias on these estimates.

7.5 Application of the functional model in the screening algorithm

This thesis has discussed the procedures for recording fetal age and has highlighted the problem of error in the recorded dates. The weeks in which gestational ages are recorded with the most precision are reported in DiPietro and Allen (1991).

The effects of errors on the parameter estimates given in Table 7.2 is demonstrated in this section. The regressions involving AFP and UE3 match the form of those used by Wald *et al* (1992). The possible advantages of the functional regression model over the linear regression models are assessed by means of a simple study that examines the changes in the parameter estimates for each outcome, when the bias associated with an gestational age error standard deviation of up to 7 days is extracted from the parameter estimates for the linear models. The slope parameters for the regression equations are given in Table 7.1. The statistics used in the study are given in Table 7.2. Table 7.3 and Table 7.4 show the changes in the standard deviations for unaffected and Down's syndrome outcomes when the bias is removed. Table 7.5 and Table 7.6 show the changes in the correlation coefficients. A concluding overview of the results is provided.

Slope parameters of the loglinear regression models fitted in Chapter three		
Model ALIN2	*Model ULOGLIN2*	*Model HLIN2*
β_1=0.0081	β_2=0. 0134	β_3= -0.0086

Table 7.1: Slope parameters of the loglinear regression equations fitted in Chapter three of this thesis.

7.6 Conclusion

Unaffected pregnancies

The effect of the bias on the standard deviations for each analyte is extremely small and removing the bias leads to only a slight reduction in the standard deviations. The greatest reduction is seen in the standard deviation for log(UE3), which reduces by 0.0128 of a standard deviation if the bias associated with a 7 day error standard deviation is removed. The effect of the bias on the correlation coefficients is also marginal. There is a slight decrease in the positive correlation between AFP and UE3 and an increase in the positive correlation between AFP and HCG. The negative correlation between UE3 and HCG marginally reduces. However, the effect of the bias is extremely small.

Down's syndrome pregnancies

Again, the effect of the bias on the standard deviations for each analyte is extremely small and removing the bias leads to a slight reduction in the standard deviations. The standard deviation for UE3 reduces the most, if the bias associated with a 7 day error standard deviation is removed. The reduction is 0.0125 of a standard deviation. The effect of the bias on the correlation coefficients is also very small. The changes in the correlation coefficients are very slight and the changes move in the same directions, for the same pairs of analytes, as described above.

This study has used functional models to investigate the extent of the bias caused by errors in recorded gestational ages on the parameter estimates for the transformed analyte values. The results show that, for both unaffected and Down's syndrome pregnancies, the

bias in the standard deviations and correlation coefficients is small. Also it has been established that the bias is a function of the slope parameters of the fitted linear regression models. In this study, the slope parameters were assumed to be known, however, the slope parameters are actually biased downwards because of the errors in the recorded gestational ages, therefore, the results of this study probably underestimate the effect of the bias. Moreover, the R^2 values of the fitted regression models that are quoted in Table 7.1 are small, indicating that only a small amount of the variation in the transformed analyte values is explained by gestational age. The bias would have more impact on the parameter estimates under models that are associated with greater R^2 values. Under these conditions, errors-in-variables models would be more appropriate.

ESTIMATED COVARIANCE MATRICES UNDER THE MODELS IN TABLE 7.1

Unaffected pregnancies

	AFP	UE3	HCG
AFP	0.1404	0.0489	0.0208
UE3		0.1221	-0.0215
HCG			0.3053

Down's syndrome pregnancies

	AFP	UE3	HCG
AFP	0.2132	0.0815	0.0156
UE3		0.1265	-0.0451
HCG			0.3144

STANDARD DEVIATIONS

Unaffected pregnancies

AFP	UE3	HCG
0.3747	0.3494	0.5525

Down's syndrome pregnancies

AFP	UE3	HCG
0.4617	0.3556	0.5607

CORRELATION COEFFICIENTS

Unaffected pregnancies

	AFP	UE3
AFP		0.3732
HCG	0.1006	-0.1112

Down's syndrome pregnancies

	AFP	UE3
AFP		0.4963
HCG	0.0602	-0.2263

Table 7.2: Covariance matrices, standard deviations and correlation coefficients under the models given in Table 7.1.

σ_η (days)	CORRECTED LOG(MoM) STANDARD DEVIATIONS		
	AFP	**UE3**	**HCG**
0	0.3747	0.3494	0.5525
1	0.3746	0.3492	0.5525
2	0.3743	0.3484	0.5523
3	0.3739	0.3471	0.5519
4	0.3733	0.3453	0.5515
5	0.3725	0.3429	0.5509
6	0.3715	0.3401	0.5501
7	0.3704	0.3366	0.5493

Table 7.3: Standard deviations of log(AFP), log(UE3) and log(HCG) for unaffected pregnancies, with bias caused by gestational age errors removed.

σ_η (days)	CORRECTED LOG(MoM) STANDARD DEVIATIONS		
	AFP	**UE3**	**HCG**
0	0.4617	0.3556	0.5607
1	0.4617	0.3554	0.5606
2	0.4615	0.3547	0.5605
3	0.4612	0.3534	0.5601
4	0.4606	0.3516	0.5597
5	0.4600	0.3493	0.5591
6	0.4592	0.3465	0.5583
7	0.4582	0.3431	0.5575

Table 7.4: Standard deviations of log(AFP), log(UE3) and log(HCG) for Down's syndrome pregnancies, with bias caused by gestational age errors removed.

σ_η (days)	CORRECTED CORRELATION COEFFICIENTS		
	AFP-UE3	**AFP-HCG**	**UE3-HCG**
0	0.3732	0.1006	-0.1112
1	0.3730	0.1008	-0.1109
2	0.3716	0.1020	-0.1093
3	0.3692	0.1038	-0.1068
4	0.3659	0.1065	-0.1032
5	0.3615	0.1099	-0.0986
6	0.3561	0.1140	-0.0928
7	0.3496	0.1190	-0.0857

Table 7.5: Correlation coefficients for log(AFP), log(UE3) and log(HCG) for unaffected pregnancies, with bias caused by gestational age errors removed.

σ_η (days)	CORRECTED CORRELATION COEFFICIENTS		
	AFP-UE3	**AFP-HCG**	**UE3-HCG**
0	0.4963	0.0602	-0.2263
1	0.4960	0.0605	-0.2258
2	0.4953	0.0614	-0.2246
3	0.4942	0.0628	-0.2226
4	0.4925	0.0648	-0.2198
5	0.4904	0.0674	-0.2162
6	0.4877	0.0706	-0.2117
7	0.4846	0.0744	-0.2063

Table 7.6: Correlation coefficients for log(AFP), log(UE3) and log(HCG) for Down's syndrome pregnancies, with bias caused by gestational age errors removed.

Chapter 8

Conclusion

This thesis has reviewed the screening algorithm that is currently used to calculate risks of fetal Down's syndrome. The algorithm uses a discriminant analysis to classify pregnancies as unaffected or Down's syndrome and this thesis has focused on improving the methodology applied in the analysis. Alternatives to the conventional methods of modelling the clinical trials data, and methods of estimating the class conditional densities, have been offered and illustrated. This thesis has also discussed the reported use of the estimated performance rates and the possibility of bias in these estimated rates has been researched. The interpretation of risks that are associated with abnormalities other than Down's syndrome has been questioned and a non-specific classification scheme has been introduced to increase sensitivity with these pregnancies. This chapter summarises the results of the research conducted in this thesis and points to areas of possible further research.

8.1 Modelling the analyte concentration values.

Current procedures used to estimate the risk of Down's syndrome are based largely on the risk algorithm described in Wald *et al* (1988). A probability model for maternal age related risks is combined with a probability model based on maternal serum samples to produce risks of Down's syndrome. Wald *et al* (1988) report the most useful analytes for predicting risks are AFP, UE3 and HCG. A risk cut-off is selected and a discriminant rule is applied that classifies pregnancies as either unaffected or Down's syndrome, depending on

whether their associated risk is lower or higher than the cut-off. Median maternal serum concentrations for unaffected pregnancies are regressed against gestational age which is recorded to the nearest completed week. The maternal serum concentrations are generally standardized to multiples of the unaffected median concentrations for the same gestations. Research has indicated that expressing concentration values as MoM values for completed weeks of gestation does not provide an efficient means of standardizing data. Different centres use different approaches to model the median analyte values so MoM values calculated from different clinical trials should not be pooled and subsequently used as part of a discriminant analysis (Bishop (1994)).

Chapter two states that the log(MoM) values are equivalent to the residuals of the fitted models. Therefore, the residuals of these models can be used directly in the discriminant analysis. This approach avoids the need to group gestational dates into completed weeks and standardize the data.

In Chapter three, the form of the regression models that are currently used by Wald *et al* (1992) to model the median concentrations of transformed AFP, UE3 and HCG values against gestational age were applied to the ungrouped data. Non-linear regression models were also used to model the separate regressions of log(UE3) and log(HCG) on gestational age. The models assume that the distribution about the location is lognormal. Centre was used as a factor in the analysis. The analysis has shown that there is little difference, in terms of fit, between the models used by Wald *et al* (1992), and the non-linear models proposed in this thesis. However, the effects of centre were overwhelmingly significant to the regression.

The gestational dates of the data used in the analysis were recorded by LMP dating methods. The motivation for using this data was that the sample based on LMP dates was larger for both unaffected and affected pregnancies than the sample based on dates that

were recorded by sonography. It would be informative to apply the same analysis to data with the gestational ages recorded by sonography, since it is well recognized that dating by sonography has less error.

8.2 Detection rates and false positive rates

Chapter four has discussed the debate over the differences between the published detection rates. The problem is partly caused by sampling error in the estimates of detection rates and false positive rates. Simulation studies were conducted in this chapter to calculate the standard errors of these rates. The standard errors are large and illustrate why clinical trials that are conduct under similar conditions, may obtain different performance estimates. The possibility of bias in the reported error rates was also investigated. It is well known that error rates which are estimated from the design set that was used to formulate the discriminant rule are overrated. The nonparametric bootstrap was applied to correct the bias in the detection rates and false positive rates. The extent of the bias caused by reclassifying the design set when screening with unrepresentative maternal age distributions was also assessed. The results indicate that the bias in the estimated performance rates is extremely small. Also, there are no real differences in the bias that is specific to each maternal age.

8.3 The interpretation of risk in non-Down's abnormalities

The risk algorithm given by Wald *et al* (1988) often fails to recognize abnormalities other than Down's syndrome and frequently associates these abnormalities with low risks (Heyl *et al* (1990)). The risks should, therefore, not be used to reassure a mother that her pregnancy is unaffected by any disorder. This problem can be addressed by incorporating a

non-specific classification into the existing screen. The Mahalanobis distance can be used to create an index of atypicality, to identify those pregnancies with analyte concentration values that are unlike trisomy 21 pregnancies but are highly atypical of unaffected pregnancies (Wright *et al* (1993)). These pregnancies can then be classified as non-specific to either outcome. This approach can easily be implemented into the existing algorithm. The approach provides an effective screen for non-Down's syndrome abnormalities and reduces the false positive rate of the screen.

8.4 Estimation of the class conditional densities

Chapter six of this thesis compared the parametric approach adopted by Wald *et al* (1988) to estimate the class conditional densities of transformed MoM analyte values with a nonparametric approach. A review of the methods of kernel density estimation has been given and illustrations of nonparametric density estimates of the trials data were provided. The adaptive kernel estimator provided high quality density estimates of the data for unaffected pregnancies. Adaptive kernel density estimates for the affected data were also constructed. Nonparametric likelihood ratios were calculated from bivariate density estimates, and these were used as part of the discriminant analysis to discriminate between affected and unaffected pregnancies.

Estimating densities in higher dimensions from relatively small samples can lead to estimates that are inefficient. Wright (1995) investigates models in which the class conditional distributions are assumed to have a common distributional form and applies nonparametric density estimation to model this common form. Wright (1995) uses parametric shifts in location and dispersion to model the differences between the classes. In the report, the technique is applied to Down's syndrome screening and the distribution of

affected cases are represented as shifted and scaled versions of those of unaffected cases. The univariate model is defined below. The notation is of the style given in Wright (1995).

Let $X_{g1}, X_{g2}, ..., X_{gn_g}$ denote the n_g observations comprising of the training data on the g^{th} class $(g = 1,2,...,G)$, which are used to estimate the probability density functions. The nonparametric approach involves fitting a density of the form

$$\hat{f}_g(x) = \frac{1}{n_g h_g} \sum_{i=1}^{n_g} K\left(\frac{x - X_{gi}}{h_g}\right)$$

where h_g is the smoothing parameter for group g and K is the kernel function. Define the class conditional densities as $f_2, f_3, ..., f_G$ for affected classes and f_1 for the unaffected class. Then, for affected classes, the density is of the form

$$f_g(x) = a_g f_1(a_g x + b_g)$$

where $a_g > 0$ is the scale parameter and b_g is the location parameter for group g.

Wright (1995) states that if these parameters are fixed, the combined sample of $n = n_1 + n_2 + ... + n_g$ observations can be used to obtain the density estimate

$$\hat{f}_1(x; \underline{a}, \underline{b}) = \frac{1}{n} \sum_{g=1}^{G} \sum_{i=1}^{n_g} \frac{1}{h} K\left(\frac{x - (X_{gi} - b_g)/a_g}{h}\right)$$

where for convenience $a_1 = 1$ and $b_1 = 0$.

Since there is a vast amount of data available for unaffected pregnancies, this approach can be used to provide higher dimensional models for the affected distributions which can then be used in the calculation of risk. This nonparametric shift model can be used with the adaptive kernel estimator to produce high quality density estimates. There is scope for future research involving the application of this model to screening. The methods of correcting bias in the error rates can be applied to the rates associated with this model.

Also, a nonparametric atypicality index can be applied to the estimated densities to establish a preliminary screen for non-Down's syndrome abnormalities.

8.5 Errors in the gestational dating methods

Chapter seven has discussed how the models that are currently used for the relationship between analyte concentration values and gestational age do not account for the errors in the recorded gestational dates. The chapter shows how errors-in-variables models can be used for these relationships. One such errors-in-variables model is the functional model that assumes the explanatory variable is recorded with a random error. The functional model was used to investigate the possibility of bias in the estimated standard deviations and correlation coefficients for the linear models. The bias was found to have very little effect on the standard deviations and the correlation coefficients that were derived for the loglinear models given in Chapter three. This could be due to the fact that the bias is a function of the slope parameters of the fitted models, and the R^2 values for the fitted regression models were particularly small. Therefore, only a small proportion of the variation in the transformed analyte values is explained by gestational age.

Appendix A

```
C
C  PROGRAM DOWN1ES.FOR
C  ********************
C
C  LAST MODIFIED 1/1/94
C
C  THIS PROGRAM CONDUCTS MONTE CARLO SIMULATIONS OF THE DOWN SYNDROME
C  SCREENING ALGORITHM USING GAUSSIAN LIKELIHOOD RATIOS TO MODIFY AGE
C  RELATED RISKS. RANDOM NUMBER GENERATORS ARE USED TO INITIALIZE
C  THE SIMULATIONS. THE CURRENT VERSION APPLIES TRUNCATION
C  LIMITS OUTSIDE (.3,3.3),(.2,5.0) AND (0.5,2.) FOR AFP, HCG AND UE3
C  RESPECTIVELY.
C
C
$LARGE
$TITLE :'DOWNS SCREENING SIMULATION SOFTWARE'

      PROGRAM DOWNS
      INTEGER MAXP,MAXSAM,MAXCAT,MXSIMG,MXSIMP
      PARAMETER (MAXP=5,MAXSAM=50000,MAXCAT=50,MXSIMG=1,MXSIMP=1,NOG=10)
      INTEGER IFREQ(MAXCAT),NCATS,INCODE(MAXP),COUNT,NTIMES,NP
      INTEGER ICODES(14,MAXP),IOUT
      INTEGER  NSAMPN,NSAMPD,NSSIZE,DSSIZE,IX,IY,IZ,I,J,K,SIMNO
      REAL AGE(MAXCAT),PANORM(MAXCAT),PADOWN(MAXCAT),
     *     PR(MAXCAT),PDOWN,NMEAN(MAXP),NCOV(MAXP,MAXP)
      REAL NCOVL(MAXP,MAXP),CVINVN(MAXP,MAXP),DETCVN,
     *     THELRN(MAXSAM),NMEANL(MAXP),CHOLN(MAXP,MAXP)
      REAL NMCOPY(MAXP),DMCOPY(MAXP),NCCOPY(MAXP,MAXP),
     *     DCCOPY(MAXP,MAXP),DAT(3000,MAXP)
      REAL NML1(MAXP),DML1(MAXP),NCVL1(MAXP,MAXP),
     *     DCVL1(MAXP,MAXP),XBARN(MAXP),XBARD(MAXP),SN(MAXP,MAXP)
      REAL SD(MAXP,MAXP),DMEAN(MAXP),DCOV(MAXP,MAXP),
     *     CHOLD(MAXP,MAXP),DCOVL(MAXP,MAXP),CVINVD(MAXP,MAXP),DETCVD
      REAL THELRD(MAXSAM),DMEANL(MAXP),XMIN(5),XMAX(5),
     *     G(NOG),PM1,THISG,PN(NOG),PD(NOG),LPM(NOG),LPM1,
     *     PN100,PD100,PM100
      REAL SCREEN,R2,SLOPE,INTER
      REAL PM(NOG),GMEAN,PNMEAN,PDMEAN,PMMEAN
      CHARACTER*31 FILEIN,LOGFILE,ROCFILE,DETFILE
      COMMON /RAND/IX,IY,IZ
      COMMON /SIMDAT/ DAT
      COMMON /LIMITS/ XMIN,XMAX
      DATA G/200,220,240,260,280,300,320,340,360,380,400/


****OPEN I/O FILES
      WRITE(*,3211)
 3211 FORMAT(20(1X/),27X,'DOWNS SCREENING SIMULATION'/
     + 27X,26('_'),10(1X/),
     +  3X,'INPUT FILE : ')
      open(2,file = 'runfile')
      do 8888 iii=1,1
      READ(*,'(A)') FILEIN
      OPEN(1,STATUS='OLD',FILE=FILEIN)
      READ(1,'(A)') LOGFILE
      READ(1,'(A)') ROCFILE
```

```
      READ(1,'(A)') DETFILE
      OPEN(9,FILE= LOGFILE)
      OPEN(6,FILE= ROCFILE)
      OPEN(7,FILE= DETFILE)


****READ IN THE MATERNAL AGE DISTRIBUTION
      CALL READAG(AGE,IFREQ,MAXCAT,NCATS)

****WORK OUT THE MASTER TABLE
      CALL WTABLE(NCATS,AGE,IFREQ,PANORM,PADOWN,PR,PDOWN)

****FIND OUT DIMENSIONALITY REQUIRED
      READ(1,*) NP
      WRITE(*,9997) NP
 9997 FORMAT(3X,'NO PARAMS=',I5)

****READ NO OF REPEATS
      READ(1,*) NTIMES

****READ SAMPLE SIZE FOR XBAR AND S
      READ(1,*) NSAMPN
      READ(1,*) NSAMPD

****READ TRUNCATION LIMITS FOR CALCULATION OF LIKELIHOOD RATIO
       READ(1,*)(XMIN(I),I=1,NP)
       READ(1,*)(XMAX(I),I=1,NP)
****INITIALISE INCODES
      DO 10 I=1,NP
   10 INCODE(I)=1

****FIND OUT PARAMETERS REQUIRED FOR NORMAL POPULATION
      CALL PARAMS('Normal Population',NMEAN,NCOV,MAXP,NP,
     *  CHOLN,MAXP,NML1,NCVL1,MAXP,CVINVN,MAXP,DETCVN,
     *           NSSIZE,MAXSAM,INCODE)

****FIND OUT PARAMETERS REQUIRED FOR DOWNS POPULATION
      CALL PARAMS('Downs Population',DMEAN,DCOV,MAXP,NP,
     *  CHOLD,MAXP,DML1,DCVL1,MAXP,CVINVD,MAXP,DETCVD,
     *           DSSIZE,MAXSAM,INCODE)

****COPY FOR SAMPLING
      DO 20 I=1,MAXP
      NMCOPY(I)=NMEAN(I)
      xbarn(i) = nmean(i)
      xbard(i) = dmean(i)
      DMCOPY(I)=DMEAN(I)



      DO 21 J=1,MAXP
      sn(i,j) = ncov(i,j)
      sd(i,j) = dcov(i,j)
      NCCOPY(I,J)=CHOLN(I,J)
   21 DCCOPY(I,J)=CHOLD(I,J)

   20 CONTINUE

****ASK FOR SEEDS TO SIMULATION
      CALL SEED
```

```
****ASK FOR VALUE OF PM (SCREEN RATe)

    READ(1,*) PM1
****READ SIMULATION SPECIFICATION
    DO 333 I=1,100
    READ(1,*,END=777) (ICODES(I,J),J=1,NP)
    write(*,*) (ICODES(I,J),J=1,NP)
 333 CONTINUE
 777 CONTINUE
    NCODES = I-1

****DO "NTIMES" COMPLETE SIMULATIONS


    DO 999 ICOUNT=1,NTIMES



    IF(NTIMES.GT.1) THEN
    CALL SAMCOV(NSAMPN,NMCOPY,NCCOPY,XBARN,SN,MAXP,NP,INCODE)
    CALL SAMCOV(NSAMPD,DMCOPY,DCCOPY,XBARD,SD,MAXP,NP,INCODE)
    ENDIF
****RUN THROUGH ALL COMBINATIONS OF MARKERS
    DO 400 I=1,NCODES
    DO 402 J =1,NP
    INCODE(J) = ICODES(I,J)
 402 CONTINUE
    K = 0
    DO 401 J =1,NP

    IF((INCODE(J).NE.1).AND.(INCODE(J).NE.0))GOTO 400
    K = K+INCODE(J)
 401  CONTINUE
    IF(K.LE.0) GOTO 400

    CALL SELECT(XBARN,SN,MAXP,INCODE,NP,NP1,NMEAN,NCOV)
    CALL SELECT(XBARD,SD,MAXP,INCODE,NP,NP1,DMEAN,DCOV)
    CALL SELECT(NML1,NCVL1,MAXP,INCODE,NP,NP1,NMEANL,NCOVL)
    CALL SELECT(DML1,DCVL1,MAXP,INCODE,NP,NP1,DMEANL,DCOVL)
    CALL CHOLES(NCOV,MAXP,NP1,CHOLN,MAXP)
    CALL CHOLES(DCOV,MAXP,NP1,CHOLD,MAXP)
    CALL INVERT(NCOVL,MAXP,NP1,CVINVN,MAXP,0.1E-10,DETCVN)
    CALL INVERT(DCOVL,MAXP,NP1,CVINVD,MAXP,0.1E-10,DETCVD)

****COMPUTE ROC


    DO 500 SIMNO=1,MXSIMG


****SIMULATE THE LIKELIHOOD RATIO ASSUMING NORMALS
    CALL SIMUL(THELRN,NSSIZE,NP1,NMEAN,CHOLN,MAXP,
    *       NMEANL,CVINVN,MAXP,DETCVN,DMEANL,CVINVD,MAXP,
    *    DETCVD,INCODE)

****SIMULATE THE LIKELIHOOD RATIO ASSUMING DOWNS
    CALL SIMUL(THELRD,DSSIZE,NP1,DMEAN,CHOLD,MAXP,
    *       NMEANL,CVINVN,MAXP,DETCVN,DMEANL,CVINVD,MAXP,
    *    DETCVD,INCODE)
```

```
****DO THE CALCULATIONS
    CALL ROC(NOG,G,PANORM,PADOWN,PR,NCATS,PDOWN
   *  ,THELRN,NSSIZE,THELRD,DSSIZE,PN,PD,PM)
****WRITE VALUES FOR THIS RUN

         IF (NP.EQ.1) ASSIGN 9901 TO IOUT
         IF (NP.EQ.2) ASSIGN 9902 TO IOUT
         IF (NP.EQ.3) ASSIGN 9903 TO IOUT
         IF (NP.EQ.4) ASSIGN 9904 TO IOUT
         IF (NP.EQ.5) ASSIGN 9905 TO IOUT

9901 FORMAT(I5,1X,1I2,1X,F4.0,3(1X,F5.2))


9902 FORMAT(I5,2I2,1X,F4.0,3(1X,F5.2))
9903 FORMAT(I5,3I2,1X,F4.0,3(1X,F5.2))
9904 FORMAT(I5,4I2,1X,F4.0,3(1X,F5.2))
9905 FORMAT(I5,5I2,1X,F4.0,3(1X,F5.2))
    DO 31 LL = 1,NOG
    PN100 = PN(LL)*100
    PD100 = PD(LL)*100
    PM100 = PM(LL)*100
    WRITE(6,IOUT)ICOUNT,(INCODE(J),J=1,NP),G(LL),PN100,PD100,PM100
    WRITE(*,IOUT)ICOUNT,(INCODE(J),J=1,NP),G(LL),PN100,PD100,PM100
 31 LPM(LL)  = LOG(PM(LL)+0.00001)
    LPM1= LOG(PM1)
    CALL REGRESS(NOG,LPM,PD,LPM1,SCREEN,SLOPE,INTER,R2)

         IF (NP.EQ.1) ASSIGN 9801 TO IOUT
         IF (NP.EQ.2) ASSIGN 9802 TO IOUT
         IF (NP.EQ.3) ASSIGN 9803 TO IOUT
         IF (NP.EQ.4) ASSIGN 9804 TO IOUT
         IF (NP.EQ.5) ASSIGN 9805 TO IOUT
9801 FORMAT(I5,I2,1X,2(1X,F5.2))
9802 FORMAT(I5,2I2,1X,2(1X,F5.2))
9803 FORMAT(I5,3I2,1X,2(1X,F5.2))
9804 FORMAT(I5,4I2,1X,2(1X,F5.2))
9805 FORMAT(I5,5I2,1X,2(1X,F5.2))

    WRITE(7,IOUT)ICOUNT,(INCODE(J),J=1,NP),100*PM1,100*SCREEN



 500 CONTINUE

 400 CONTINUE
 999 CONTINUE
8888 continue
7777 STOP


    END

*************************************************************************************

    SUBROUTINE PARAMS(TITLE,MEAN,COV,MAXP,NP,CHOL,ICH,
   *  MEANL,COVL,ICL,COVINV,ICI,DETCOV,NSSIZE,MAXSAM,INCODE)
    CHARACTER*(*) TITLE
```

```
      INTEGER NP,MAXP,ICH,ICI,NSSIZE,MAXSAM,ICL,INCODE(5)
      INTEGER MCOP,NP1,IROW,ICOL
      PARAMETER(MCOP=5)
      REAL COVCOP(MCOP,MCOP),MEAN(NP),MEANL(NP)
      REAL COV(MAXP,NP),CHOL(ICH,NP),COVINV(ICI,NP),
     *     DETCOV,COVL(ICL,NP),CORR(MCOP,MCOP)

****FIND OUT MEAN AND COVARIANCE MATRIX OF POPULATION
****NOTE 2 COVARIANCE MATRICES
****ONE FOR SIMULATION, OTHER TO BE PLUGGED INTO LIKELIHOOD RATIO
      CALL ASK(TITLE,MEAN,COV,MAXP,NP,MEANL,COVL,ICL,INCODE,NP1)

****FIND CHOLESKY DECOMPOSITION OF COVARIANCE MATRIX (SIMULATION)
      CALL CHOLES(COV,MAXP,NP1,CHOL,ICH)


****COPY COVARIANCE MATRIX
      DO 1 IROW=1,NP1
           DO 2 ICOL=1,NP1
            COVCOP(IROW,ICOL)=COVL(IROW,ICOL)
    2 CONTINUE
    1 CONTINUE
****WRITE MEAN VECTOR, COV AND CORR MATRIX
      WRITE(9,9004) TITLE
      WRITE(9,9000)
 9004 FORMAT(////1X,70('*')//10X,'POPULATION : ',A20//)
 9000 FORMAT(///30X,'SIMULATION'//)
      WRITE(9,9001) (MEAN(J),J=1,NP1)
 9001 FORMAT(1X,70('*')//5X,'MEAN VECTOR'//5(3X,F10.5))
      WRITE(9,9002)
 9002 FORMAT(//5X,'COVARIANCE MATRIX'//)
      DO 901 I =1,NP1
      WRITE(9,9003)( COV(I,J),J=1,NP1)
      DO 904 J =1,NP1
      CORR(I,J) = COV(I,J)/SQRT(COV(I,I)*COV(J,J))
  904 CONTINUE
  901 CONTINUE
      WRITE(9,9006)
 9006 FORMAT(//5X,'CORRELATION MATRIX'//)
      DO 905 I = 1,NP1
  905 WRITE(9,9003)(CORR(I,J),J=1,NP1)
 9003 FORMAT(5(3XF10.5))
      WRITE(9,9005)
 9005 FORMAT(////1X,70('*')//30X,'ALGORITHM'//)
      WRITE(9,9001) (MEANL(J),J =1,NP1)
      WRITE(9,9002)
      DO 902 I = 1,NP1
      WRITE(9,9003)( COVL(I,J),J =1,NP1)
      DO 906 J = 1,NP1
      CORR(I,J) = COVL(I,J)/SQRT(COVL(I,I)*COVL(J,J))
  906 CONTINUE
  902 CONTINUE
      WRITE(9,9006)
      DO 907 I = 1,NP1
  907 WRITE(9,9003)(CORR(I,J),J=1,NP1)
****FIND INVERSE & DET OF COVARIANCE MATRIX IN L.R. STATISTIC
      CALL INVERT(COVCOP,MCOP,NP1,COVINV,ICI,0.1E-10,DETCOV)

****ASK FOR THE SAMPLE SIZE REQUIRED FROM THIS POPULATION
       CALL ASKSAM(NSSIZE,MAXSAM)
```

```
      WRITE(*,*)NSSIZE

      RETURN
      END

*******************************************************************************

      SUBROUTINE ASK(TITLE,MEAN,COV,MAXP,NP,MEANL,COVL,ICL,INCODE,
     *  NP1)
      CHARACTER*(*) TITLE
****THIS SUBROUTINE ASKS FOR THE MEAN & COVARIANCE MATRIX
      INTEGER MAXP,NP,ICL,NP1,I,J,K
      REAL MEAN(NP), COV(MAXP,NP),MEANL(NP),COVL(ICL,NP)
      REAL  DUMEAN(5),DUMCOV(5,5)

****READ APPROPRIATE POPULATION PARAMETERS

      READ(1,*)(MEAN(I),I=1,NP)
      READ(1,*)((COV(I,J),J=1,I),I=1,NP)
****READ L-R MATRICES
      READ(1,*)(MEANL(I),I=1,NP)
      READ(1,*)((COVL(I,J),J=1,I),I=1,NP)

****COPY OVER THE COVARIANCE MATRIX
      DO 10 I=1,NP
       DO 9 J=1,I
            COV(J,I)=COV(I,J)
            COVL(J,I)=COVL(I,J)
    9 CONTINUE
   10 CONTINUE

****SELECT REQUIRED VARIABLES
      CALL SELECT(MEAN,COV,MAXP,INCODE,NP,NP1,MEAN,COV)
      CALL SELECT(MEANL,COVL,MAXP,INCODE,NP,NP1,MEANL,COVL)

      RETURN
      END

*******************************************************************************

      SUBROUTINE CHOLES(V,IV,N,T,IT)
      IMPLICIT REAL(A-H,O-Z)
      INTEGER IV,N,IT
      REAL V(IV,N),T(IT,N)
      REAL TE,S,EPS
      INTEGER I,I1,J,J1,K

****CHOLESKY SQUARE ROOT OF V
****V ASSUMED TO BE POSITIVE DEFINITE
****WANT TO FIND T (LOWER TRIANGULAR)
****SUCH THAT (T)*(T)TR = (V)

      EPS=0.1E-11
      DO 100 I=1,N
      DO 101 J=1,N
      T(J,I)=0.0
  101 CONTINUE
  100 CONTINUE
      TE=V(1,1)
      IF(TE.LT.EPS) GO TO 50
```

```
      T(1,1)=SQRT(TE)
      IF(N.EQ.1)RETURN
      T(2,1)=V(2,1)/T(1,1)
      TE=V(2,2)-T(2,1)**2
      IF(TE.LT.EPS) GO TO 50
      T(2,2)=SQRT(TE)
      IF(N.EQ.2)RETURN
      DO 31 I=3,N
      T(I,1)=V(I,1)/T(1,1)
      I1=I-1
      DO 29 J=2,I1
      S=0.0D0
      J1=J-1
      DO 28 K=1,J1
   28 S=S+T(I,K)*T(J,K)
   29 T(I,J)=(V(I,J)-S)/T(J,J)
      S=0.0D0
      DO 30 J=1,I1
   30 S=S+T(I,J)**2
      TE=V(I,I)-S
      IF(TE.LT.EPS) GO TO 50
   31 T(I,I)=SQRT(TE)
      RETURN

****ERROR MESSAGES
   50 CONTINUE
      WRITE(6,76)
   76 FORMAT(' ***ILL-CONDITIONED VARIANCE MATRIX***')
      STOP
      END

*******************************************************************************

      SUBROUTINE SEED
      INTEGER IX,IY,IZ,I,ISEED(3)
      COMMON /RAND/IX,IY,IZ
      LOGICAL OK

****READS SEEDS FOR RANDOM NO. GENERATOR
      READ(1,*)(ISEED(I),I=1,3)
      OK=.TRUE.
      DO 3 I=1,3
      IF((ISEED(I).LT.1) .OR. (ISEED(I).GT.30000))OK=.FALSE.
    3 CONTINUE
      IF(.NOT. OK) THEN
                              WRITE(6,10)
   10 FORMAT(' *** SEED OUTSIDE RANGE 1 - 30000 ***')
                              STOP
                         ENDIF
      IX=ISEED(1)
      IY=ISEED(2)
      IZ=ISEED(3)
      RETURN
      END

*******************************************************************************

      SUBROUTINE NRAND(SIM,N)
      IMPLICIT REAL(A-H,O-Z)
      INTEGER N,I,NHALF
```

```
      REAL SIM(N),DUM,U,V,VV,RANDOM

****THIS SUBROUTINE GENERATES A NORMAL RANDOM VARIABLE
****FROM TWO INDEPENDENT UNIFORMS BY THE POLAR MARSAGLIA-BRAY METHOD

      NHALF=(N+1)/2
      DO 1 I=1,NHALF
    2 U=RANDOM(DUM)
      V=RANDOM(DUM)

****U AND V ARE IND. U(0,1),TRANSFORM TO U(-1,1)
      U=(2.0D0*U)-1.0D0
      V=(2.0D0*V)-1.0D0

****TRANSFORM TO TWO IID N(0,1)'S
      VV=(U*U)+(V*V)
      IF(VV.GT.1.0D0)GO TO 2
      VV=SQRT((-2.0D0*LOG(VV))/VV)
      SIM(I)=U*VV
      SIM(N-I+1)=V*VV
    1 CONTINUE
      RETURN
      END

*******************************************************************************

      REAL FUNCTION RANDOM(DUM)

****ALGORITHM AS 183 APPL. STATIST. (1982) VOL.31, NO.2
****RETURNS A PSUDO-RANDOM NUMBER RECTANGULARLY DISTRIBUTED
****BETWEEN O AND 1.
****IX,IY AND IZ SHOULD BE SET TO INTEGER VALUES BETWEEN
****1 AND 30000 BEFORE FIRST ENTRY
****INTEGER ARITHMETIC UP TO 30323 IS REQUIRED

      INTEGER IX,IY,IZ
      REAL DUM
      COMMON /RAND/IX,IY,IZ
      REAL FLOAT
      INTEGER MOD
      REAL DMOD

      IX=171*MOD(IX,177)-2*(IX/177)
      IY=172*MOD(IY,176)-35*(IY/176)
      IZ=170*MOD(IZ,178)-63*(IZ/178)
      IF(IX.LT.0) IX=IX+30269
      IF(IY.LT.0) IY=IY+30307
      IF(IZ.LT.0) IZ=IZ+30323
      RANDOM=MOD(FLOAT(IX)/30269.0D0 +FLOAT(IY)/30307.0D0
     *          +FLOAT(IZ)/30323.0D0,1.0D0)
      RETURN
      END

*******************************************************************************

      SUBROUTINE INVERT(A,IA,N,B,IB,EPS,DEL)
      INTEGER IA,IB,N
      REAL A(IA,N),B(IB,N),EPS
      REAL DEL,AMAX,AMULT,ATMP,BTMP,DIV
      INTEGER I,J,K,IMAX,KP1
```

```
****MATRIX INVERSION BY ELIMINATION WITH PARTIAL PIVOTING
****ORIGINAL MATRIX=A INVERSE MATRIX =B
****NOTE THAT A IS CHANGED ON EXIT
****EPS IS A SMALL QUANTITY USED TO SEE IF MATRIX SINGULAR
****DEL IS VALUE OF DETERMINANT ON EXIT

****CONSTRUCT IDENTITY MATRIX B(I,J)= I
    DO 6 I=1,N
    DO 5 J=1,N
    B(I,J)=0.0
  5 CONTINUE
    B(I,I)=1.0
  6 CONTINUE

****LOCATE MAXIMUM MAGNITUDE A(I,K) ON OR BELOW MAIN DIAGONAL
    DEL=1.0
    DO 45 K=1,N
    IF (K-N) 12,30,30
 12 IMAX=K
    AMAX=ABS(A(K,K))
    KP1=K+1
    DO 20 I=KP1,N
    IF(AMAX - ABS(A(I,K)))15,20,20
 15 IMAX=I
    AMAX=ABS(A(I,K))
 20 CONTINUE

****INTERCHANGE ROWS IMAX AND K IF IMAX NOT EQUAL TO K
    IF(IMAX-K)25,30,25
 25 DO 29 J=1,N
    ATMP=A(IMAX,J)
    A(IMAX,J)=A(K,J)
    A(K,J)=ATMP
    BTMP=B(IMAX,J)
    B(IMAX,J)=B(K,J)
 29 B(K,J)=BTMP
    DEL=-DEL
 30 CONTINUE

****TEST FOR SINGULAR MATRIX
    IF(ABS(A(K,K))- EPS)93,93,35
 35 DEL=A(K,K)*DEL

****DIVIDE PIVOT ROW BY ITS MAIN DIAGONAL ELEMENT
    DIV=A(K,K)
     DO 38 J=1,N
    A(K,J)=A(K,J)/DIV
 38 B(K,J)=B(K,J)/DIV

****REPLACE EACH ROW BY LINEAR COMBINATION WITH PIVOT ROW
    DO 43 I=1,N
    AMULT=A(I,K)
    IF (I-K) 39,43,39
 39 DO 42 J=1,N
    A(I,J)=A(I,J) - AMULT*A(K,J)
 42 B(I,J)=B(I,J) - AMULT*B(K,J)
 43 CONTINUE
 45 CONTINUE
    RETURN
```

```
 93 WRITE(6,113)K
113 FORMAT(' SINGULAR MATRIX FOR K=',I3)
    STOP
    END

*******************************************************************************

    SUBROUTINE MVN(X,NPAR,MEAN,CHOL,IC,INCODE)
    INTEGER NPAR,IC,INCODE(5),MAXPAR,DIM,DIM2
    PARAMETER (MAXPAR=5)
    REAL X(NPAR),CHOL(IC,NPAR),MEAN(NPAR),N01(MAXPAR)

****GENERATE A RANDOM SAMPLE OF NPAR NORMAL(0,1)
    CALL NRAND(N01,NPAR)

****CONVERT TO MULTIVATIATE NORMAL WITH MEAN = MEAN(1...NPAR)
****AND A VARIANCE COVARIANCE MATRIX WHOSE CHOLESKY DECOMPOSITION = CHOL
    DO 10 DIM=1,NPAR
     X(DIM)=MEAN(DIM)
     DO 11 DIM2=1,NPAR
        X(DIM)=X(DIM) + CHOL(DIM,DIM2)*N01(DIM2)
 11 CONTINUE

****RETURN UE3 TO ORIGINAL SCALE
C    IF ((NPAR.EQ.3).AND.(DIM.EQ.3)) X(DIM)=EXP(X(DIM))
C    IF ((NPAR.EQ.2).AND.(INCODE(3).EQ.1).AND.(DIM.EQ.2))
C   *            X(DIM)=EXP(X(DIM))
C    IF ((NPAR.EQ.1).AND.(INCODE(3).EQ.1)) X(DIM)=EXP(X(DIM))
 10 CONTINUE
    RETURN
    END

*******************************************************************************  *

    SUBROUTINE SIMUL(SIM,NSIMS,N,MEAN,CHOL,IC,
   *    MEAN1,CV1INV,ICV1,DET1,MEAN2,CV2INV,ICV2,DET2,INCODE)
    INTEGER NSIMS,N,IC,ICV1,ICV2,INCODE(5),MAXPAR,ISIM
    PARAMETER (MAXPAR=5)
    REAL SIM(NSIMS),X(MAXPAR),MEAN(N),CHOL(IC,N),PDF
    REAL MEAN1(N),CV1INV(ICV1,N),DET1
    REAL MEAN2(N),CV2INV(ICV2,N),DET2

    DO 1 ISIM = 1, NSIMS

****GENERATE MVN WITH MEAN=MEAN & COVARIANCE MATRIX
****WHOSE CHOLESKY DECOMPOSTION IS CHOL
    CALL MVN(X,N,MEAN,CHOL,IC,INCODE)

****FORM A LR STATISTIC
    SIM(ISIM)=PDF(X,N,MEAN1,CV1INV,ICV1,DET1,INCODE)/
   *       PDF(X,N,MEAN2,CV2INV,ICV2,DET2,INCODE)
  1 CONTINUE
    RETURN
    END

*******************************************************************************

    REAL FUNCTION PDF(X,N,MEAN,CVINV,IC,DET,INCODE)
    INTEGER N,IC,MAXPAR,I,J
    PARAMETER (MAXPAR=5)
```

```
      REAL X(N),MEAN(N),CVINV(IC,N),DET,XMIN(5),XMAX(5)
      REAL XMMU(MAXPAR),CXMMU(MAXPAR),QUAD
      INTEGER INCODE(MAXPAR)

****RETURNS MVN PDF (NOT INCLUDING PI PARTS)

****SET VALUES FOR CENSORING

      COMMON /LIMITS/ XMIN,XMAX

****ALLOW FOR MISSING VARS
      J=1
      DO 10 I=1,3
      IF (INCODE(I).EQ.1) THEN
            XMIN(J)=XMIN(I)
            XMAX(J)=XMAX(I)
            J=J+1
      ENDIF
   10 CONTINUE

****CENSORING...
      DO 1 I=1,N
C     IF (X(I).LT.XMIN(I))  X(I) = XMIN(I)
C     IF (X(I).GT.XMAX(I))  X(I) = XMAX(I)
            XMMU(I)=X(I) - MEAN(I)
    1 CONTINUE

****EVALUATE PDF
      QUAD=0.0D0
      DO 2 I= 1,N
            CXMMU(I)=0.0D0
            DO 3 J=1,N
             CXMMU(I)=CXMMU(I) + CVINV(I,J)*XMMU(J)
    3 CONTINUE
            QUAD=QUAD + XMMU(I)*CXMMU(I)
    2 CONTINUE
      PDF=EXP(-0.5*QUAD)/SQRT(DET)
      RETURN
      END

*************************************************************************************** *

      SUBROUTINE READAG(AGE,IFREQ,MAXCAT,NCATS)
      INTEGER MAXCAT,IFREQ(MAXCAT),NCATS,I,NUMSO,IFR
      REAL AGE(MAXCAT),AGECAT
      CHARACTER*20 FNAME

****READS MATERNAL AGE DISTRIBUTION
      READ(1,21)FNAME
   21 FORMAT(A20)
      OPEN(5,STATUS='OLD',FILE=FNAME)
      NUMSO=0
      DO 10 I=1,MAXCAT
      READ(5,*,END=30,ERR=40)AGECAT,XIFR
      AGE(I)=AGECAT
      IFREQ(I)=3*XIFR
      NUMSO=I
   10 CONTINUE

****CHECK TO SEE NO MORE INFO IN FILE
```

```
      READ(5,*,END=30)AGECAT
      WRITE(6,50)MAXCAT
   50 FORMAT(' Program only allows for ',I4,' age categories')
      STOP
   30 NCATS=NUMSO
      IF (NUMSO.EQ.0)THEN
          WRITE(6,60)
   60 FORMAT(' NO INFO IN AGE FILE!!!!')
          STOP
      ENDIF
      RETURN

****ERROR MESSAGES
   40 WRITE(6,41)
   41 FORMAT(' ERROR IN READING MATERNAL AGE FILE!!!!!')
      STOP
      END

*******************************************************************************

      SUBROUTINE WTABLE(NCATS,AGE,IFREQ,PANORM,PADOWN,PR,PDOWN)
      INTEGER NCATS,IFREQ(NCATS),I,MTOT,MAXCAT
      PARAMETER (MAXCAT=50)
      REAL AGE(NCATS),PANORM(NCATS),PADOWN(NCATS),PR(NCATS)
      REAL PDOWN,PDAGE(MAXCAT),PAGE(MAXCAT),DMTOT,PNORM

****GIVEN MATERNAL AGE DISTRIBUTION IN NCATS GROUPS, CALCULATES :
****     MAT. AGE DISTN. FOR DOWNS (PADOWN)
****     MAT. AGE DISTN. FOR NORMALS (PANORM)
****     PROB RATIOS NORMALS OVER DOWNS FOR EACH MAT. AGE (PR)
****     OVERALL PROPORTION OF DOWNS (PDOWN)

      MTOT=0
      DO 1 I=1,NCATS
       PDAGE(I)=0.000627D0 + EXP(-16.2395D0 + 0.286D0*AGE(I))
       PR(I)=(1.0D0-PDAGE(I))/PDAGE(I)
       MTOT=MTOT+IFREQ(I)
    1 CONTINUE
      DMTOT=FLOAT(MTOT)
      PDOWN=0.0D0
      DO 2 I=1,NCATS
       PAGE(I)=FLOAT(IFREQ(I))/DMTOT
       PDOWN=PDOWN + PDAGE(I)*PAGE(I)
    2 CONTINUE
      PNORM=1.0D0-PDOWN
      DO 4 I=1,NCATS
       PADOWN(I)=PAGE(I)*PDAGE(I)/PDOWN
       PANORM(I)=PAGE(I)*(1.0D0-PDAGE(I))/PNORM
    4 CONTINUE
      RETURN
      END

*******************************************************************************

      SUBROUTINE DOCALC(G,PANORM,PADOWN,PR,NCATS,PDOWN,PN,PD,PM,
     *                  SNORM,NNORM,SDOWN,NDOWN)
      INTEGER NCATS,NNORM,NDOWN,I
      REAL G,PANORM(NCATS),PADOWN(NCATS),PR(NCATS),
     *              PDOWN,PN,PD,PM,SNORM(NNORM),SDOWN(NDOWN)
      REAL PROPN,P1I,P2I,THISV(50)
```

```
      INTEGER NTAL(50),DTAL(50)
****RETURNS PN,PD AND PM FOR GIVEN G

      PN=0.0D0

      PD=0.0D0
      DO 1 I=1,NCATS
      NTAL(I)=NNORM
      DTAL(I)=NDOWN
    1 THISV(I) = G/PR(I)

      DO 4 I=1,NNORM
      DO 3 J=NCATS,1,-1
      IF(SNORM(I).GT.THISV(J))THEN
      DO 2 K =1,J
    2 NTAL(K)=NTAL(K)-1
      GOTO 4
      END IF
    3 CONTINUE
    4 CONTINUE
      DO 7 I = 1,NDOWN
      DO 6 J =NCATS,1,-1
      IF(SDOWN(I).GT.THISV(J)) THEN
      DO 5 K =1,J
    5 DTAL(K) = DTAL(K)-1
      GOTO 7
      END IF
    6 CONTINUE
    7 CONTINUE
      DO 8 I=1,NCATS
      PN=PN+FLOAT(NTAL(I))*PANORM(I)/FLOAT(NNORM)
       PD=PD+FLOAT(DTAL(I))*PADOWN(I)/FLOAT(NDOWN)
    8 CONTINUE
      PM=PN*(1.0D0-PDOWN) + PD*PDOWN
      RETURN
      END

*******************************************************************************

      SUBROUTINE ROC(NOG,G,PANORM,PADOWN,PR,NCATS,PDOWN,
     *               SNORM,NNORM,SDOWN,NDOWN,PN,PD,PM)
      INTEGER NCATS,NNORM,NDOWN
      REAL G(NOG),PANORM(NCATS),PADOWN(NCATS),PR(NCATS),
     *             PDOWN,PM(NOG),SNORM(NNORM),SDOWN(NDOWN)
      REAL PN(NOG),PD(NOG),PMTOP,GBOT,GTOP,GMID,PMMID,ASZERO
      LOGICAL CONVGE,LOWER

**** FINDS DETECTION RATES PN(NOG), PD(NOG) AND PM(NOG) FOR NORMALS, DOWN
**** SYNDROME AND TOTAL POPULATION RESPECTIVELY FOR RISK CUT-OFFS STORED
**** IN G(NOG).

**** NOTE SHOULD BE ABLE TO MAKE THIS MUCH FASTER BY DEALING WITH ENTIRE
**** ARRAY OF G'S AND EXPLOITING MONOTONICITY

      DO 100 I = 1,NOG
      CALL DOCALC(G(I),PANORM,PADOWN,PR,NCATS,PDOWN,PN(I),PD(I)
     *,PM(I),SNORM,NNORM,SDOWN,NDOWN)
```

```
100 CONTINUE

    RETURN
    END

*******************************************************************************

    SUBROUTINE REGRESS(N,X,Y,XPRED,YPRED,SLOPE,INTER,R2)

    INTEGER N
    INTEGER I

    REAL X(N),Y(N),SLOPE,INTER,XPRED,YPRED
    REAL XD,YD,SXX,SXY,SYY

    SXX = 0
    SXY = 0
    SYY = 0

    XMEAN = 0
    YMEAN = 0

    DO 1 I =1,N
    XMEAN = XMEAN+X(I)/N
    YMEAN = YMEAN+Y(I)/N
  1 CONTINUE

    DO 2 I =1,N
    XD = X(I) - XMEAN
    YD = Y(I) - YMEAN
    SXX = SXX+XD*XD
    SXY = SXY+XD*YD
    SYY = SYY+YD*YD
  2 CONTINUE

    SLOPE = SXY/SXX
    INTER = YMEAN-SLOPE*XMEAN

    YPRED = INTER+SLOPE*XPRED

    R2 = SXY*SXY/(SXX*SYY)

    RETURN

    END

*******************************************************************************
    REAL FUNCTION PROPN(X,N,XNUM)
    INTEGER N,I,COUNT
    REAL X(N),XNUM

****THIS SUBROUTINE FINDS THE PROPORTION OF THE ARRAY WHICH IS
**** LESS THAN XNUM

    COUNT=0
    DO 1 I=1,N
  1 IF (X(I).LT.XNUM) COUNT=COUNT+1
    PROPN=FLOAT(COUNT)/FLOAT(N)
    RETURN
    END
```

```
*******************************************************************************

      SUBROUTINE MEANOF(X,N,XMEAN)
      INTEGER N,I
      REAL X(N),XMEAN,TOT,SUM

****CALCULATES MEAN OF ARRAY X

      TOT=0.0
      SUM=0.0
      DO 1 I=1,N
       TOT=TOT+X(I)
    1 CONTINUE
      XMEAN=TOT/FLOAT(N)
      RETURN
      END

*******************************************************************************

      SUBROUTINE ASKSAM(NSAMP,MAXSAM)
      INTEGER NSAMP,MAXSAM

****GET SAMPLE SIZES REQUIRED FOR SUBROUTINE SIMUL
      READ(1,*)NSAMP
      IF((NSAMP.LT.1).OR.(NSAMP.GT. MAXSAM)) THEN
                                             WRITE(6,10) MAXSAM
   10 FORMAT(' ***SAMPLE SIZE OUTSIDE RANGE 1 -',I6)
                                             STOP
                                             ENDIF
      RETURN
      END

*******************************************************************************

      SUBROUTINE SELECT(XBAR,S,MAXP,INCODE,NP,NP1,NEWXB,NEWS)
      REAL XBAR(NP),S(MAXP,NP),NEWXB(NP),NEWS(MAXP,NP)
      INTEGER INCODE(5),NP,NP1,I,J,K

****SELECTS ROWS AND COLS FROM XBAR AND S AS DEFINED IN INCODE AND
****STORES IN NEWXB AND NEWS

      K=1
      DO 1 I=1,NP
      IF (INCODE(I).EQ.1) THEN
          NEWXB(K)=XBAR(I)
          DO 2 J=1,NP
    2   NEWS(K,J)=S(I,J)
          K=K+1
          ENDIF
    1 CONTINUE
      K=1
      DO 3 J=1,NP
      IF (INCODE(J).EQ.1) THEN
          DO 4 I=1,NP
    4   NEWS(I,K)=NEWS(I,J)
          K=K+1
          ENDIF
    3 CONTINUE
      NP1=K-1
```

```
      RETURN
      END

*******************************************************************************

      SUBROUTINE SAMCOV(N,MCOPY,CCOPY,MEAN,S,MAXP,NP,INCODE)
      REAL MCOPY(MAXP),CCOPY(MAXP,MAXP),MEAN(NP),
     *     X(5),DAT(3000,5),S(MAXP,MAXP)
      INTEGER N,MAXP,NP,INCODE(MAXP),I,J
      COMMON/SIMDAT/DAT

****GENERATES SAMPLE OF SIZE N FROM APPROPRIATE MVN DISTRIBUTION, RETURNS
****UE3 TO ORIGINAL SCALE AND CALCULATES SAMPLE MEAN AND COVAR MATRICES

      DO 1 J=1,N
      CALL MVN(X,NP,MCOPY,CCOPY,MAXP,INCODE)
      DO 2 I=1,NP
C       IF ((NP.EQ.3).AND.(I.EQ.3)) X(I)=LOG(X(I))
C       IF ((NP.EQ.2).AND.(INCODE(3).EQ.1).AND.(I.EQ.2))
C     *                       X(I)=LOG(X(I))
C      IF ((NP.EQ.1).AND.(INCODE(3).EQ.1)) X(I)=LOG(X(I))
    2 DAT(J,I)=X(I)
    1 CONTINUE
      CALL SAMMAT(NP,N,MEAN,S,MAXP)
      RETURN
      END

*******************************************************************************

      SUBROUTINE SAMMAT(NP,NOBS,XBAR,S,MAXP)
      REAL X(3000,5),XBAR(NP),S(MAXP,MAXP)
      INTEGER NP,NOBS,I,J,K
      COMMON/SIMDAT/X

****CALCULATES MEAN AND COVARIANCE MATRIX FOR DATA IN X

      DO 10 I=1,NP
      XBAR(I)=0.0
      DO 10 J=1,NP
   10 S(I,J)=0.0
      DO 1 I=1,NP
      DO 2 J=1,NOBS
    2 XBAR(I)=XBAR(I)+X(J,I)
    1 XBAR(I)=XBAR(I)/NOBS
      DO 3 I=1,NP
      DO 3 J=1,NOBS
    3 X(J,I)=X(J,I)-XBAR(I)
      DO 4 I=1,NP
      DO 5 J=I,NP
      DO 6 K=1,NOBS
    6 S(I,J)=S(I,J)+X(K,I)*X(K,J)
      S(I,J)=S(I,J)/(NOBS-1)
    5 S(J,I)=S(I,J)
    4 CONTINUE
      RETURN
      END
```

Appendix B

Maternal age distribution for England and Wales (1986-1988)

11	1
12	2
13	23
14	189
15	1102
16	4412
17	10787
18	17506
19	23862
20	28686
21	33847
22	38958
23	43582
24	47234
25	49737
26	49847
27	48500
28	45507
29	41935
30	37269
31	31259
32	26333
33	21781
34	17800
35	14405
36	11216
37	8682
38	6752
39	5042
40	3463
41	2096
42	1218
43	691
44	388
45	201
46	113
47	68
48	46
49	29
50	19
51	15
52	13
53	8
54	6
55	8

Appendix C

```
$LARGE
      PROGRAM SAMPLING

      PARAMETER (MAXN=300,MAXP=3,MAXG=4,AGEMIN=11,AGEMAX=55,NCATS=60)
      COMMON /RAND/IX,IY,IZ
      REAL XN(MAXN,MAXP),XD(MAXN,MAXP),RANDOM,XBARN(MAXP),XBARD(MAXP)
      REAL SN(MAXP,MAXP),SD(MAXP,MAXP),INVN(MAXP,MAXP),INVD(MAXP,MAXP)
      REAL DETN,DETD,ESP,XMIN(MAXP),XMAX(MAXP),PDF,RISK(NCATS)
      REAL XNB(MAXN,MAXP),XDB(MAXN,MAXP),MRISKN(NCATS),MRISKD(NCATS)
      REAL BIAS(MAXG,NCATS),BIASC(MAXG,NCATS),OABERRORN,OABERRORD
      REAL ERROR,DS(MAXG,NCATS),AP(MAXG,NCATS),SUMBIAS(MAXG)
      REAL OABRIGHTD,BADRIGHTD,SUMAP(MAXG)
      REAL SUMBIASDR(MAXG),BIASDR(MAXG,NCATS),BIASCDR(MAXG,NCATS)
      REAL DSDR(MAXG,NCATS),SUMODR(MAXG),ODR(MAXG,NCATS)
      REAL SUMBIASC(MAXG),SUMBIASCDR(MAXG)
      REAL CONDPMN(ncats),CONDPMD(ncats),PM(NCATS),FREQM(NCATS),TALLY
      DOUBLE PRECISION LRN(MAXN),LRD(MAXN)
      INTEGER NTOTAL,DTOTAL,K,R,G,INDEX(MAXN),D,P,SET,C
      INTEGER OAN(MAXN,NCATS),BAN(MAXN,NCATS),OABN(MAXN,NCATS)

      INTEGER OAD(MAXN,NCATS),BAD(MAXN,NCATS),OABD(MAXN,NCATS)



****SAMPLING.FOR ****
****AUTHOR: CHRISTINE DONOVAN ****

****PURPOSE: This program is designed to calculate ****
****age specific bias corrected estimates of the group ****
****conditional error rates for a sample based discriminant ****
****rule. The comnputer intensive resampling technique, ****
****bootstrapping, is employed to generate a series of ****
****apparent error rates that are corrected for bias ****


****APPLICATION: The program is applied to Down's syndrome screening ****
****to provide age specific biased corrected estimates of the rates of ****
****misclassification of Wald's screening algorithm based on two ****
****or more markers. The sampling distributions of detection rates and false positive rates ****
****are summed over a standardized maternal age distribution. ****



****Data file read in is a:\data, error rates are written to ****
****a:\results ****

    OPEN(3,FILE='a:data2.txt')
    OPEN(4,FILE='a:results')
    OPEN(5,FILE='a:DATA.txt')
    OPEN(6,FILE='a:DDATA.txt')
    OPEN(7,FILE='A:AGE.DAT')

****G indicates the number of classesin the discriminant analysis, P sets the data dimension ****

      READ(3,*)G
      READ(3,*)P
```

```
****NTOTAL and DTOTAL denote samples sizes for normals and Downs data****

      READ(3,*)NTOTAL
      READ(3,*)DTOTAL

****R indicates the number of bootstrap samples required ****

      READ(3,*)R

****XN stores the normal data ****

      DO 10 J=1,NTOTAL
      READ(5,*)(XN(J,D),D=1,P)
 10   CONTINUE

****XD stores the Downs data ****

      DO 50 J=1,DTOTAL
      READ(6,*)(XD(J,D),D=1,P)
 50   CONTINUE


****sets truncation limits on markers ****

    DO 103 I=1,P
    READ(3,*)XMIN(I)
    READ(3,*)XMAX(I)
103  CONTINUE

****reads in maternal age distribution ***

    DO 121 J=AGEMIN,AGEMAX
    READ(7,*)FREQM(J)
    TALLY=TALLY+FREQM(J)
121 CONTINUE

****calculates relative frequencies ***

    DO 232 J=AGEMIN,AGEMAX
    PM(J)=FREQM(J)/TALLY
232 CONTINUE


****reads in selected risk cut off ****

    PRINT*,'RISK CUT OFF='
    READ*,C

****calculates age specific risks ****

      CALL AGERISK(MRISKN,MRISKD,RISK)

****calculates parameter estimates for the class conditional distributions****

      CALL SEED
      CALL SETCOV(NTOTAL,P,XBARN,SN,XN)
      CALL INVERT(SN,MAXP,P,INVN,MAXP,ESP,DETN)
      CALL SETCOV(DTOTAL,P,XBARD,SD,XD)
      CALL INVERT(SD,MAXP,P,INVD,MAXP,ESP,DETD)
```

```
****calculates likelihood ratios associated with original normal subjects ****

      CALL LIKR(P,XBARN,XBARD,INVN,INVD,DETN,DETD,NTOTAL,XN,LRN,C)

****classifies original normal subjects according to assumed age distribution ****
****stores in OAN ****

      CALL AGEBIAS(NTOTAL,LRN,RISK,OAN,C,1)

****calculates likelihood ratios associated with original Down's subjects ****

      CALL LIKR(P,XBARN,XBARD,INVN,INVD,DETN,DETD,DTOTAL,XD,LRD,C)

****classifies original Down's subjects according to assumed age distribution ****
****stores in OAD ****

      CALL AGEBIAS(DTOTAL,LRD,RISK,OAD,C,2)

****initializes the bias in each group to zero ****

      DO 2 M=AGEMIN,AGEMAX
      DO 1 I=1,G
      DS(I,M)=0
      dsdr(i,m)=0
 1    CONTINUE
 2    CONTINUE

****resamples ****

      DO 35 K=1,R

****resamples from normal data, stores in XNB ****

      CALL BOOT(NTOTAL,INDEX)
      DO 70 J=1,NTOTAL
      DO 80 D=1,P
      XNB(J,D)=XN(INDEX(J),D)
 80   CONTINUE
 70   CONTINUE


****resamples from Down's data, stores in XDB ****

      CALL BOOT(DTOTAL,INDEX)
      DO 90 J=1,DTOTAL
      DO 100 D=1,P
      XDB(J,D)=XD(INDEX(J),D)
100   CONTINUE
 90   CONTINUE


      CALL SETCOV(NTOTAL,P,XBARN,SN,XNB)
      CALL INVERT(SN,MAXP,P,INVN,MAXP,ESP,DETN)
      CALL SETCOV(DTOTAL,P,XBARD,SD,XDB)
      CALL INVERT(SD,MAXP,P,INVD,MAXP,ESP,DETD)

****calculates likelihood ratios of bootstrap normal samples ****
```

```
      CALL LIKR(P,XBARN,XBARD,INVN,INVD,DETN,DETD,NTOTAL,XNB,LRN,C)

****classifies XNB for each m stores in BAN ****

      CALL AGEBIAS(NTOTAL,LRN,RISK,BAN,C,1)

****calculates likelihood ratios of bootstrap downs sample****

      CALL LIKR(P,XBARN,XBARD,INVN,INVD,DETN,DETD,DTOTAL,XDB,LRD,C)


****classifies XDB for each m, stores in BAD ****

      CALL AGEBIAS(DTOTAL,LRD,RISK,BAD,C,2)

****calculates lrs of original normal data using bootstrap rule ****

      CALL LIKR(P,XBARN,XBARD,INVN,INVD,DETN,DETD,NTOTAL,XN,LRN,C)

****classifies normal data lrs computed by bootstrap rule, OABN ****

      CALL AGEBIAS(NTOTAL,LRN,RISK,OABN,C,1)

****calculates lrs of original downs data using bootstrap rule ****

      CALL LIKR(P,XBARN,XBARD,INVN,INVD,DETN,DETD,DTOTAL,XD,LRD,C)

****classifies downs data lrs computed by bootstrap rule, OABD ****

      CALL AGEBIAS(DTOTAL,LRD,RISK,OABD,C,2)

****calculates bootstrap bias of apparent error rate of normals, (1), ****
****by averaging over the age specific bias in each bootstrap sample ****

      DO 51 M=AGEMIN,AGEMAX
      OABERRORN=ERROR(NTOTAL,OABN,M)
      BIAS(1,M)=ERROR(NTOTAL,BAN,M)-OABERRORN
      DS(1,M)=DS(1,M)+BIAS(1,M)

****calculates bootstrap bias of apparent error rate of Down's, (2), ****
****by averaging over the age specific bias in each bootstrap sample ****

      OABERRORD=ERROR(DTOTAL,OABD,M)
      BIAS(2,M)=ERROR(DTOTAL,BAD,M)-OABERRORD
      DS(2,M)=DS(2,M)+BIAS(2,M)
      OABRIGHTD=1-OABERRORD
      BADRIGHTD=1-ERROR(DTOTAL,BAD,M)
      BIASDR(2,M)=BADRIGHTD-OABRIGHTD
      DSDR(2,M)=DSDR(2,M)+BIASDR(2,M)
 51   CONTINUE
 35   CONTINUE


****calculates the age specific group conditional error rates ****
****of the original data sets ****

      DO 52 M=AGEMIN,AGEMAX
      AP(1,M)=ERROR(NTOTAL,OAN,M)
      AP(2,M)=ERROR(DTOTAL,OAD,M)
      ODR(2,M)=1-AP(2,M)
```

```
****corrects the age specific group conditional apparent error rates for bias ****
****to provide unbiased estimates of the actual group conditional****
****error rates ****

        BIAS(1,M)=DS(1,M)/R
        BIAS(2,M)=DS(2,M)/R
        BIASDR(2,M)=DSDR(2,M)/R

        BIASC(1,M)=AP(1,M)-BIAS(1,M)
        BIASC(2,M)=AP(2,M)-BIAS(2,M)
        BIASCDR(2,M)=ODR(2,M)-BIASDR(2,M)


****writes results to file ****

        WRITE(4,*)'AP(1 ',M,')=',AP(1,M)
        WRITE(4,*)'AP(2 ',M,')=',AP(2,M)
        WRITE(4,*)'BIAS(1 ',M,')=',BIAS(1,M)
        WRITE(4,*)'BIAS(2 ',M,')=',BIAS(2,M)
        WRITE(4,*)'BIASC(1 ',M,')=',BIASC(1,M)
        WRITE(4,*)'BIASC(2 ',M,')=',BIASC(2,M)
        WRITE(4,*)'BIASDR(2 ',M,')=',BIASDR(2,M)
        WRITE(4,*)'BIASCDR(2 ',M,')=',BIASCDR(2,M)

  52    CONTINUE


****calculates the age related probabilities for normals ****
****and Down's outcomes, CONDPMN and CONDPMD ****

    CALL AGESET(PM,MRISKN,CONDPMN)
    CALL AGESET(PM,MRISKD,CONDPMD)


****calculates weighted averages of the age specific error rates ****
****according to the proportions of the maternal age distribution ****
****for normal outcomes ****


      CALL RATES(SUMAP,CONDPMN,AP,1)
      CALL RATES(SUMBIAS,CONDPMN,BIAS,1)
      CALL RATES(SUMAP,CONDPMN,AP,1)
      CALL RATES(SUMBIASC,CONDPMN,BIASC,1)


****calculates weighted averages of the age specific error rates ****
****according to the proportions of the maternal age distribution ****
****for Down's outcomes ****

      CALL RATES(SUMODR,CONDPMD,ODR,2)
      CALL RATES(SUMBIAS,CONDPMD,BIAS,2)
      CALL RATES(SUMBIASDR,CONDPMD,BIASDR,2)
      CALL RATES(SUMBIASCDR,CONDPMD,BIASCDR,2)

****writes results to file ****

        WRITE(4,*)'SUMMED BIAS(1 )=',SUMBIAS(1)
```

```
      WRITE(4,*)'SUMMED BIASDR(2)=',SUMBIASDR(2)
      WRITE(4,*)'SUMMED BIASC(1)=',SUMBIASC(1)
      WRITE(4,*)'SUMMED BIASCDR(2)=',SUMBIASCDR(2)
      WRITE(4,*)'SUMMED AP(1)=',SUMAP(1)
      WRITE(4,*)'SUMMED ODR(2)=',SUMODR(2)


      CLOSE(3)
      CLOSE(4)
      CLOSE(5)
      CLOSE(6)
      CLOSE(7)

      PRINT* ,'PROGRAM HAS TERMINATED'
      STOP
      END

  SUBROUTINE SEED

****Reads seeds for random no. generator

  INTEGER IX,IY,IZ,I,ISEED(3)
  COMMON /RAND/IX,IY,IZ
  LOGICAL OK
 DO 5 I=1,3
  PRINT*,'SEED='
  READ(*,*)ISEED(I)
5 CONTINUE
  OK=.TRUE.
  DO 3 I=1,3
  IF((ISEED(I).LT.1) .OR. (ISEED(I).GT.30000))OK=.FALSE.
3 CONTINUE
  IF(.NOT. OK) THEN

  PRINT*,' *** SEED OUTSIDE RANGE 1 - 30000 ***'
                    STOP
                ENDIF
  IX=ISEED(1)
  IY=ISEED(2)
  IZ=ISEED(3)
  RETURN
  END


  REAL FUNCTION RANDOM(DUM)

****Algorithm as 183 appl. statist. (1982) vol.31, no.2
****returns a pseudo-random number rectangularly distributed
****between 0 and 1.
****IX, IYand IZ should be set to integer values between
****1 and 30000 before first entry
****integer arithmetic up to 30323 is required

  INTEGER IX,IY,IZ
  REAL DUM
  COMMON /RAND/IX,IY,IZ
  REAL FLOAT
  INTEGER MOD
  REAL DMOD
```

```
      IX=171* MOD(IX,177)-2* (IX/177)
      IY=172* MOD(IY,176)-35* (IY/176)
      IZ=170* MOD(IZ,178)-63* (IZ/178)
      IF(IX.LT.0) IX=IX+30269
      IF(IY.LT.0) IY=IY+30307
      IF(IZ.LT.0) IZ=IZ+30323
      RANDOM=MOD(FLOAT(IX)/30269.0D0 +FLOAT(IY)/30307.0D0
     +         +FLOAT(IZ)/30323.0D0,1.0D0)
      RETURN
      END


      SUBROUTINE BOOT(M,INDEX)

****This subroutine returns a bootstrap sample of a set of M integers ****
****and stores the sample in INDEX ****

      PARAMETER (MAXN=300)
      INTEGER INDEX(MAXN),J
      REAL DUM,RANDOM,Y
      COMMON /RAND/IX,IY,IZ

      DUM = 0
      DO 100 J=1,M
      INDEX(J)=INT( RANDOM(DUM)*M)+1
  100 CONTINUE
      RETURN
      END




      REAL FUNCTION MEAN(T,D,X,P)

****This function returns the arithmetic mean of T numbers read
****from a data matrix , with D indicating the data column
****to be accessed

      PARAMETER (MAXP=3,MAXN=300)
      REAL X(MAXN,MAXP)
      INTEGER T,J,D,P


      MEAN=0
      DO 4 J=1,T
      MEAN=X(J,D)+MEAN
    4 CONTINUE
      MEAN=MEAN/T
      RETURN
      END


      REAL FUNCTION COV(MU,T,D,Q,X,P)

****This function computes the covariance of two variables consisting
****of 1,2,..t observations read from a data matrix X. D and Q
****index the variables used. the array MU stores the mean of the
****variables
```

```
      PARAMETER (MAXP=3,MAXN=300)
      REAL X(MAXN,MAXP),MU(MAXP)
      INTEGER T,D,Q,P



      COV=0
      DO 6 J=1,T
      COV=COV+((X(J,D)-MU(D))*(X(J,Q)-MU(Q)))
    6 CONTINUE
      COV=COV/(T-1)
      RETURN
      END

      SUBROUTINE SETCOV(T,P,XBAR,S,X)

****Returns the covariance, S(P,P),for each variable P=1,2,..P
****from COV. XBAR stores the mean vector of the variables.

      PARAMETER (MAXP=3,MAXN=300)
      REAL XBAR(MAXP),S(MAXP,MAXP),COV,MEAN,X(MAXN,MAXP)
      INTEGER T,SUM,P,Q,D,J


      DO 3 D=1,P
      XBAR(D)=MEAN(T,D,X,P)
    3 CONTINUE
      DO 40 SUM=1,P
      D=SUM
      DO 10 Q=SUM,P
      S(D,Q)=COV(XBAR,T,D,Q,X,P)
      S(Q,D)=S(D,Q)
   10 CONTINUE
   40 CONTINUE
      RETURN
      END



      SUBROUTINE INVERT(A,IA,N,B,IB,EPS,DEL)

****Matrix inversion by elimination with partial pivoting
****original matrix A inverse matrix =B
****note that A is changed on exit
****EPS is a small quantity used to see if the matrix is singular
****DEL is value of determinant on exit

      PARAMETER (MAXP=3,MAXN=300)
      INTEGER IA,IB,N
      REAL A(IA,N),B(IB,N),EPS
      REAL DEL,AMAX,AMULT,ATMP,BTMP,DIV
      INTEGER I,J,K,IMAX,KP1



****construct identity matrix B(I,J)= I
```

```
      EPS=0.1E-11
      DO 6 I=1,N
      DO 5 J=1,N
      B(I,J)=0.0
    5 CONTINUE
      B(I,I)=1.0
    6 CONTINUE

****locate maximum magnitude A(I,K) on or below main diagonal

      DEL=1.0
      DO 45 K=1,N
      IF (K-N) 12,30,30
   12 IMAX=K
      AMAX=ABS(A(K,K))
      KP1=K+1
      DO 20 I=KP1,N
      IF(AMAX - ABS(A(I,K)))15,20,20
   15 IMAX=I
      AMAX=ABS(A(I,K))
   20 CONTINUE

****interchange rows IMAX and K if IMAX not equal to K

      IF(IMAX-K)25,30,25
   25 DO 29 J=1,N
      ATMP=A(IMAX,J)
      A(IMAX,J)=A(K,J)
      A(K,J)=ATMP
      BTMP=B(IMAX,J)
      B(IMAX,J)=B(K,J)
   29 B(K,J)=BTMP
      DEL=-DEL
   30 CONTINUE

****test for singular matrix

      IF(ABS(A(K,K))- EPS)93,93,35
   35 DEL=A(K,K)* DEL

****divide pivot row by its main diagonal element

      DIV=A(K,K)
      DO 38 J=1,N
      A(K,J)=A(K,J)/DIV
   38 B(K,J)=B(K,J)/DIV

****replace each row by linear combination with pivot row

      DO 43 I=1,N
      AMULT=A(I,K)
      IF (I-K) 39,43,39
   39 DO 42 J=1,N
      A(I,J)=A(I,J) - AMULT* A(K,J)
   42 B(I,J)=B(I,J) - AMULT* B(K,J)
   43 CONTINUE
   45 CONTINUE
```

```
      RETURN
   93 WRITE(* ,113)K
  113 FORMAT(' SINGULAR MATRIX FOR K=',I3)
      STOP
      END



      REAL FUNCTION PDF(P,XBAR,INV,X,J)

****returns mvn pdf (not including pi parts)
****set values for censoring

      PARAMETER (MAXP=3,MAXN=300)
      INTEGER P,I,J,A
      REAL X(MAXN,MAXP),XBAR(MAXP),INV(MAXP,MAXP)
      REAL XMIN(MAXP),XMAX(MAXP)
      REAL XMMU(MAXP),CXMMU(MAXP),QUAD



      DO 1 I=1,P

C           IF (X(J,I).LT.XMIN(I))   X(J,I) = XMIN(I)
C           IF (X(J,I).GT.XMAX(I))   X(J,I) = XMAX(I)
            XMMU(I)=X(J,I) - XBAR(I)

    1 CONTINUE

****evaluate PDF

       QUAD=0.0D0
      DO 2 I= 1,P
            CXMMU(I)=0.0D0
            DO 3 A=1,P
             CXMMU(I)=CXMMU(I)+INV(I,A)*XMMU(A)
    3 CONTINUE
            QUAD=QUAD+(XMMU(I)*CXMMU(I))
    2 CONTINUE
      PDF= (-0.5*QUAD)

      RETURN
      END


      SUBROUTINE LIKR(P,XBARN,XBARD,INVN,INVD,DETN,DETD,T,X,LR,C)

****This subroutine calculates the positive likelihood ratio of****
****a normal outcome to a Down's, storing the result in LR****
****The routine employs PDF****

      PARAMETER (MAXP=3,MAXN=300)
      INTEGER P,T,J,I,C
      REAL XBARN(MAXP),XBARD(MAXP),INVN(MAXP,MAXP),MAX,MIN
      REAL INVD(MAXP,MAXP),DETN,DETD,X(MAXN,MAXP),PDF
      DOUBLE PRECISION LR(MAXN)
```

```
      MAX=LOG((SQRT(DETN)*C)/(SQRT(DETD)*28))
      MIN=LOG((SQRT(DETN)*C)/(SQRT(DETD)*1578))

      DO 3 J=1,T
      LR(J)=PDF(P,XBARN,INVN,X,J)
      LR(J)=LR(J)-PDF(P,XBARD,INVD,X,J)
      IF (LR(J).GT.MAX) THEN
       LR(J)=MAX
      ENDIF
      IF (LR(J).LT.MIN) THEN
       LR(J)=MIN
      ENDIF
      LR(J)=EXP(LR(J))
      LR(J)=LR(J)*(SQRT(DETD)/SQRT(DETN))

    3 CONTINUE
      RETURN
      END



      SUBROUTINE AGERISK(MRISKN,MRISKD,RISK)

****This subroutine calculates the maternal age related risk of an age ****
****distribution from AGEMIN to AGEMAX, according to Cuckle(1987),****
****and stores them in MRISKN and MRISKD. RISK stores the ratio of the two probablilties.****

      PARAMETER (NCATS=60,AGEMIN=11,AGEMAX=55)
      REAL MRISKN(NCATS),MRISKD(NCATS),RISK(NCATS)



      DO 61 J=AGEMIN,AGEMAX
      MRISKN(J)=0.999373-EXP(-16.2395+(0.286*J))
      MRISKD(J)=(1-MRISKN(J))
      RISK(J)=MRISKN(J)/MRISKD(J)

   61 CONTINUE
      RETURN
      END



      SUBROUTINE AGEBIAS(T,LR,RISK,ALL,C,I)
      PARAMETER (MAXN=300,NCATS=60, AGEMIN=11, AGEMAX=55)
      DOUBLE PRECISION LR(MAXN)
      REAL RISK(NCATS)
      INTEGER J,C,T,I,ALL(MAXN,NCATS)

****AGEBIAS calls ALLOCATE for each subject in order to allocate the subject to ****
****either Down's or normals , for an assumed  maternal age distribution ****



      DO 100 J=1,T
      CALL ALLOCATE(J,RISK,LR,ALL,C,I)
  100  CONTINUE
```

```
      RETURN
      END


      SUBROUTINE ALLOCATE(J,RISK,LR,ALL,C,I)
      PARAMETER (AGEMIN=11, AGEMAX=55, MAXN=300, NCATS=60)
      DOUBLE PRECISION LR(MAXN)
      REAL  RISK(NCATS)
      INTEGER C,S,M,J,TEMP,I,ALL(MAXN,NCATS)


****This routine allocates each subject to the classification Down's or normals by allocating****
****a 0 or 1, with 1 indicating membership of group i. Subjects are ****
****allocated over an assumed age distribution from AGEMIN to AGEMAX. ****
****The criteria of allocation is based on a risk cut off value c. ****
****if  LR > (c/RISK) then the subject is classified as normal.***


            TEMP=AGEMIN

      IF (I.EQ.2) THEN
       DO 1 M=TEMP,AGEMAX,1

                  IF (LR(J).GT.(float(C)*(1/RISK(M)))) THEN
                    ALL(J,M)=0
                  ELSE
                    DO 2 S=M,AGEMAX
                         ALL(J,S)=1
 2                  CONTINUE
                    goto 999
                  ENDIF
 1          CONTINUE

      ELSE

            TEMP=AGEMAX
        DO 3 A=TEMP,AGEMIN,-1

                  IF (LR(J).LE.(float(C)*(1/RISK(A)))) THEN
                    ALL(J,A)=0
                  ELSE
                    DO 4 U=A,AGEMIN,-1
                         ALL(J,U)=1
 4                  CONTINUE
                    goto 999
                  ENDIF
 3          CONTINUE
      ENDIF
999     RETURN
      END



      INTEGER FUNCTION SET(U)

****This function returns the error of an allocation integer U****

      INTEGER U
```

```
      IF (U.EQ.1) THEN
      SET=0
      ELSE
      SET=1
      ENDIF
      RETURN
      END

      REAL FUNCTION ERROR(T,ALL,M)

****This function returns the apparent error rate of an allocation****
****vector C, based on T subjects, employing the function SET****

      PARAMETER (MAXN=300,NCATS=60)
      INTEGER T,J,SET,ALL(MAXN,NCATS),M

      ERROR=0
      DO 10 J=1,T
      ERROR=ERROR+SET(ALL(J,M))
   10 CONTINUE
      ERROR=ERROR/T
      RETURN
      END



      SUBROUTINE AGESET(PM,MRISK,CONDPM)
      PARAMETER (AGEMIN=11,AGEMAX=55,NCATS=60,MAXG=4)
      REAL SUM,CONDPM(NCATS)
      REAL MRISK(NCATS),PM(NCATS)
      INTEGER J


****This subroutine provides a weighted average of the maternal age related
*** probabilities by summing the age related probabilities of risk ****
****over the maternal age distribution***

      SUM=0
      DO 601 J=AGEMIN,AGEMAX
      SUM=SUM+PM(J)*MRISK(J)
   601 CONTINUE
      DO 701 J=AGEMIN,AGEMAX
      CONDPM(J)=(PM(J)*MRISK(J))/SUM
   701 CONTINUE
      RETURN
      END



      SUBROUTINE RATES(TOTAL,CONDPM,AMOUNT,I)
      PARAMETER (AGEMIN=11,AGEMAX=55,NCATS=60,MAXG=4)
      REAL TOTAL(maxg),CONDPM(NCATS)
      REAL AMOUNT(MAXG,NCATS)
      INTEGER J,I

****This subroutine sums any received amount by the ****
****maternal age distribution ***
```

```
      TOTAL=0
      DO 656 J=AGEMIN,AGEMAX
      TOTAL(i)=TOTAL(i)+AMOUNT(I,J)*CONDPM(J)

  656 CONTINUE

      RETURN
      END
```

Appendix D

```
{ATYPICALITY INDEX.PAS}
{AUTHOR: CHRISTINE DONOVAN}
{DATE: 10/04/92}
{PURPOSE: This program is primarily designed to produce sets of atypicality}
{indices from a random sample of recorded analyte concentrations or MoM}
{values taken from either Down's syndrome or Normal pregnancies, to provide}
{an assessment of how typical each observation is of each catagorization.}
{The risk algorithm given by Wald et al (1988) is used to assign a risk of abnormality to}
{each pregnancy.}

program atypicality(input,output,filein,fileout);

type vectora=1..5;
    vectorb=1..5;
    vector1=array[vectora] of real;
    vector2=array[vectora,vectorb] of real;
    vector3=array[1..2000] of real;

var p,n,i,j,nsamp,count5,count2,count1,countrisk:integer;
   eps,detn,detd,ain,aid,aintrunc,aidtrunc,likr,agerisk,risk:real;
   meann,meand,xminn,xmind,xmaxn,xmaxd,nmin,dmin,nmax,dmax,x:vector1;
   cvinvn,cvinvd,covn,covd:vector2;
   age:vector3;
   nonsingular:boolean;
   filein,fileout,logout:text;
   namein,nameout,xout:string[20];

procedure invert(var v,inv:vector2;var p1,p2,n1:integer;
            var nonsingular:boolean;var eps,det:real);

{calculates v inverse by elimination with partial pivoting, original }
{matrix = v, inverse matrix = inv, v changes on exit; eps is a small }
{value used to see if matrix is singular, det is the determinant on exit}

var amax,amult,atmp,btmp,divn:real;
   i,j,k,imax,kp1:integer;

begin
 for i:=1 to p1 do
  begin
   for j:=1 to p1 do
    inv[i,j]:=0;
   inv[i,i]:=1
  end;

 det:=1;
 nonsingular:=true;
 k:=1;
 eps:=0.1e-11;

 while ((k<=p1) and nonsingular) do
 begin
  if (k<p1) then
   begin
    imax:=k;
```

```
amax:=abs(v[k,k]);              {locate maximum magnitude a(i,k) on or}
kp1:=k+1;                       {below the main diagonal}
 for i:=kp1 to p1 do
  begin
   if (amax<abs(v[i,k])) then
    begin
     imax:=i;
     amax:=abs(v[i,k])
    end;
   end;

  if (k<>imax) then
   begin                        {interchange rows imax and k if imax}
    for j:=1 to p1 do           {not equal to k}
     begin
      atmp:=v[imax,j];
      v[imax,j]:=v[k,j];
      v[k,j]:=atmp;
      btmp:=inv[imax,j];
      inv[imax,j]:=inv[k,j];
      inv[k,j]:=btmp
     end;
     det:=-det;
    end;
   end;

  if (abs(v[k,k])<=eps) then        {test for singular matrix}
   nonsingular:=false
   else
    begin
     det:=v[k,k]*det;
     divn:=v[k,k];
      for j:=1 to p1 do         {divide pivot row by its main}
       begin                    {diagonal element}
        v[k,j]:=v[k,j]/divn;
        inv[k,j]:=inv[k,j]/divn
       end;
     for i:=1 to p1 do
      begin
       amult:=v[i,k];

      if (i<>k) then
       begin
        for j:=1 to p1 do
         begin
          v[i,j]:=v[i,j]-amult*v[k,j];
          inv[i,j]:=inv[i,j]-amult*inv[k,j]
         end;
       end;
      end;
    end;
    k:=k+1;
end;
end;



function pdf(var n1,p1,p2:integer;var det,quad:real;
```

```
        var mean,xmin,xmax,x:vector1;var cvinv:vector2):real;

{this function returns mvn pdf, not including pi parts, and produces an}
{atypicality index}

type lista=1..5;
    list1=array[lista] of real;

var i,j:integer;
   cxmmu,xmmu:list1;

begin
 for i:=1 to p1 do
  begin
   if (x[i]<xmin[i]) then
    x[i]:=xmin[i];              {censoring x values by comparison with}
   if (x[i]>xmax[i]) then       {cut-off values read from file}
    x[i]:=xmax[i];
   xmmu[i]:=(x[i]-mean[i]);
  end;
 quad:=0.0e-10;                 {evaluate pdf}

 for i:=1 to p1 do
  begin
   cxmmu[i]:=0.0e-10;
   for j:=1 to p1 do
    cxmmu[i]:=cxmmu[i]+cvinv[i,j]*xmmu[j];
   quad:=quad+xmmu[i]*cxmmu[i]              {quad=atypicality index}
  end;
pdf:=(exp(-0.5*quad)/sqrt(det));
end;



procedure lr(var n1,p1,p2:integer;var detn,detd,ain,aid,likr:real;
        var meann,meand,xminn,xmind,xmaxn,xmaxd,x:vector1;
        var cvinvn,cvivnd:vector2);

{this procedure calculates the likelihood ratio, or odds modifier}

var v,u:real;
   i,j:integer;

begin


 u:=pdf(n1,p1,p2,detn,ain,meann,xminn,xmaxn,x,cvinvn);
 v:=pdf(n1,p1,p2,detd,aid,meand,xmind,xmaxd,x,cvinvd);
 likr:=u/v
end;

begin
 writeln('input file?');
 readln(namein);
 writeln('output file?');
 readln(nameout);
 writeln('output log file?');
 readln(xout);
 assign(filein,namein); assign(fileout,nameout); assign(logout,xout);
```

```
 reset(filein);
 rewrite(fileout);rewrite(logout);

 write('matrix dimensions n,p=');
 readln(n,p);

for i:=1 to p do
 readln(filein,meann[i]);
for i:=1 to p do                    {reads from file mean vectors}
 readln(filein,meand[i]);              {x cut-offs and covariance matrices}
for i:=1 to p do
 readln(filein,xminn[i],nmin[i]);
for i:=1 to p do
 readln(filein,xmind[i],dmin[i]);
for i:=1 to p do
 readln(filein,xmaxn[i],nmax[i]);
for i:=1 to p do
 readln(filein,xmaxd[i],dmax[i]);
for i:=1 to p do
 begin
 for j:=1 to n do
 readln(filein,covn[i,j]);
 end;

for i:=1 to p do
 begin
 for j:=1 to n do
  readln(filein,covd[i,j]);
 end;

 invert(covn,cvinvn,p,p,n,nonsingular,eps,detn);     {calculates cov inverse}
 invert(covd,cvinvd,p,p,n,nonsingular,eps,detd);     {for downs and normals}
 readln(filein,nsamp);

 write(fileout,'      AIN          AID          LIKR ');
 writeln(fileout,'      AGE RISK          RISK');

 count5:=0;
 count2:=0;
 count1:=0;
 countrisk:=0;
for i:=1 to nsamp do
 begin
 read(filein,age[i]);
  for j:=1 to p do
  begin
    read(filein,x[j]);
    x[j]:=(ln(x[j])/2.3025851);
    write(logout,x[j],' ');
  end;
  readln(filein); writeln(logout);
 lr(n,p,p,detn,detd,ain,aid,likr,meann,meand,xminn,xmind,
    xmaxn,xmaxd,x,cvinvn,cvinvd);

 agerisk:=((0.999373-exp(-16.2395+(0.286*age[i])))/
        (0.000627+exp(-16.2395+(0.286*age[i]))));
 lr(n,p,p,detn,detd,aintrunc,aidtrunc,likr,meann,meand,nmin,dmin,
    nmax,dmax,x,cvinvn,cvinvn);
  risk:=likr*agerisk;
 writeln(fileout,ain,aid,likr,agerisk,risk);
```

```
 {fileout contains the atypicality}
 {indices and}
 {the likelihood ratios for each subject}

if (risk > 300) then
 begin
  if (ain >= 5.991) then
   count5:=count5+1;
  if (ain >= 7.824) then
   count2:=count2+1;
  if (ain >= 9.210) then
   count1:=count1+1;
 end
 else
  countrisk:=countrisk+1;
end;
 writeln(fileout,'No. screened neg. but atypical of normal at');
 writeln(fileout,' 5%   2%   1%');
 writeln(fileout,count5:3,count2:5,count1:5);
 writeln(fileout,'No. screened pos.=',countrisk);

 close(fileout); close(logout);
end.
□
```

Appendix E

Software written in S-Plus designed to construct univariate and bivariate kernel density estimates

```
> de
function(x, data, h)
{
        pts <- length(x)
        de <- vector(length = pts)
        n <- length(data)
        for(i in 1:pts) {
                de[i] <- sum(dnorm(x[i], data, h)/n)
        }
        as.double(de)
}
> de2
function(x, data, h1, h2)
{
#
#
#  x is an mx2 vactor of points
#
        pts <- length(x[, 1])
        n <- length(data[, 1])
        de <- vector(length = pts)
        for(i in 1:pts) {
                print(i)
                de[i] <- sum(dnorm(x[i, 1], data[, 1], h1) * dnorm(x[i, 2],
                        data[, 2], h2))/n
        }
        as.double(de)
}
>
```

Appendix F

Software written in S-Plus designed to construct univariate and bivariate parametric densities

```
gauss_function(p,x,mu,inv,j,det,xmin,xmax)
{
        pts <- length(x)
        x <- matrix(nrow=pts,ncol=2)
        mu <- vector(length=2)
        xmmu <- vector(length=2)
        cxmmu <- vector(length=2)
        inv <- matrix(nrow=2,ncol=2)


for ( i in 1:p){
ifelse((x[j,i]> xmin[i]),x[j,i],xmin[i])
ifelse((x[j,i]< xmax[i]),x[j,i],xmax[i])
}
for ( i in 1:p){
xmmu[i]<-x[j,i]-mu[i]
}

quad<-0
for (i in 1:p){
cxmmu[i]<-0
for (k in 1:p) {
cxmmu[i]<- cxmmu[i]+inv[i,k]*xmmu[k]
}
quad<-quad+(xmmu[i]*cxmmu[i])
}
temp<-(-0.5*quad)
prod<-(6.283*sqrt(det))^-1
prod*exp(temp)
}
```

Appendix G

Software written in FORTRAN designed to calculate detction rates and false positive rates.

```
      PARAMETER (maxn=1500,maxg=2,agemin=11,agemax=55)
      PARAMETER (ncats=60)
      REAL DR,FPr
      REAL CONDPMN(NCATS),CONDPMD(NCATS),PM(NCATS),FREQM(NCATS),TALLY
      REAL lrn(maxn),lrd(maxn),risk(ncats),mriskn(ncats)
      REAL mriskd(ncats),pos(maxg,ncats)
      INTEGER n(maxg),C



      OPEN(1,FILE='a:lrn')
      OPEN(2,FILE='a:lrd')
      OPEN(3,FILE='a:AGE.DAT')
      open(5,file='a:results')

      DO 121 J=AGEMIN,AGEMAX
      READ(3,*)FREQM(J)
      TALLY=TALLY+FREQM(J)
  121 CONTINUE



      DO 232 J=AGEMIN,AGEMAX
      PM(J)=FREQM(J)/TALLY
  232 CONTINUE

      WRITE(*,*)'SAMPLE SIZE NORMALS='
      READ(*,*)N(1)
      WRITE(*,*)'SAMPLE SIZE DOWNS='
      READ(*,*)N(2)
      WRITE(*,*)'CUTOFF VALUE='
      READ(*,*)C


      DO 12 J=1,n(1)
       READ(1,*)lrn(j)
   12 CONTINUE

      DO 13 J=1,n(2)
       READ(2,*)lrd(j)
   13 CONTINUE


         call agerisk(mriskn,mriskd,risk)

      CALL SCREENPOS(N(1),LRN,pos,C,1,risk)
      CALL SCREENPOS(N(2),LRD,pos,C,2,risk)
      CALL AGESET(PM,MRISKN,CONDPMN)
      CALL AGESET(PM,MRISKD,CONDPMD)
      CALL RATES(FPR,CONDPMN,POS,1)
      CALL RATES(DR,CONDPMD,POS,2)
      write(5,*)'dr=',dr
```

```
      write(5,*)'fpr=',fpr

         close(1)
         close(2)
         close(3)



         STOP
         END


      SUBROUTINE AGERISK(MRISKN,MRISKD,RISK)

****This subroutine calculates the maternal age related risk of an age ****
****distribution from AGEMIN to AGEMAX, according to Cuckle(1987),****
****and stores them in MRISKN and MRISKD. RISK stores the ratio of the two probablilties.****

      PARAMETER (ncats=60,agemin=11,agemax=55)
      REAL MRISKN(NCATS),MRISKD(NCATS),RISK(NCATS)


      DO 61 J=AGEMIN,AGEMAX
      MRISKN(J)=0.999373-EXP(-16.2395+(0.286*J))
      MRISKD(J)=(1-MRISKN(J))
      RISK(J)=MRISKN(J)/MRISKD(J)

   61 CONTINUE
      RETURN
      END



      SUBROUTINE SCREENPOS(N,LR,POS,C,T,risk)
      PARAMETER (AGEMIN=11,AGEMAX=55,MAXN=1500,maxg=2,ncats=60)
      real  pos(maxg,NCATS)
      INTEGER N,I,J,C,T
      REAL LR(MAXN),RISK(NCATS),no1(maxg,NCATS)


DO 1 J=AGEMIN,AGEMAX
       no1(T,J)=0.d0
       DO 2 I=1,N
            IF (LR(I).LT. (C/RISK(J))) THEN
          NO1(T,J)=NO1(T,J)+1.d0
         ENDIF
    2  CONTINUE

      pos(T,J)=no1(t,j)/N
      write(5,*)'screen pos(',T,',',J,')=',pos(T,J)
    1 CONTINUE
      RETURN
      END

      SUBROUTINE AGESET(PM,MRISK,CONDPM)
      PARAMETER (AGEMIN=11,AGEMAX=55,NCATS=60,MAXG=2)
      REAL SUM,CONDPM(NCATS)
      REAL MRISK(NCATS),PM(NCATS)
      INTEGER J
```

```
****This subroutine provides a weighted average of the maternal age related
*** probabilities by summing the age related probabilities of risk ****
****over the maternal age distribution***

      SUM=0
      DO 601 J=AGEMIN,AGEMAX
      SUM=SUM+PM(J)*MRISK(J)
  601 CONTINUE
      DO 701 J=AGEMIN,AGEMAX
      CONDPM(J)=(PM(J)*MRISK(J))/SUM
  701 CONTINUE
      RETURN
      END



      SUBROUTINE RATES(TOTAL,CONDPM,POS,I)
      PARAMETER (AGEMIN=11,AGEMAX=55,NCATS=60,MAXG=2)
      REAL CONDPM(NCATS)
      REAL POS(MAXG,NCATS)
      INTEGER J,I

****This subroutine sums any received amount by the conditional ****
****maternal age distribution ***


      TOTAL=0
      DO 656 J=AGEMIN,AGEMAX
      TOTAL=TOTAL+POS(I,J)*CONDPM(J)
  656 CONTINUE
      RETURN
      END
```

REFERENCES

Abramson, I., S. (1982). On bandwidth variation in kernel estimates - a square root law. Ann. Statist., 10, 1217-1233.

Aitchison, J., Dunsmore., I., R.(1975) Statistical Prediction Analysis. Cambridge University Press.

Altman, D.,G. (1993) Construction of age-related reference centiles using absolute residuals. Statistics in Medicine. 12, 917-924.

Altman D.,G. and Chitty L., S. (1994) Charts of fetal size. Br. J. Obstet. and Gynaecol. 101, 29-34.

Bishop, J., C. (1994). Statistical techniques involved in establishing reference ranges.

Bogart, M., H., Pandian, M., R., Jones, O., W. (1987). Abnormal maternal serum choronic gonadotrophin levels in pregnancies with fetal chromosome abnormalites. Prenat. Diagn. 7, 623 - 630.

Bowman, A., W. (1984) An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71, 2, 353-360.

Box, G., E., P., and Cox, D., R. (1964). An analysis of transformations (with discussion). Journal of the Royal Statistical Society, B, 26, 211-252.

Boys, R., J.(1992). On a kernel approach to a screening problem. J. R. Statist. Soc. B, 54, No. 1, 157-169.

Brambati, B. and Simoni, G. (1983). Diagnosis of fetal trisomy 21 in the first trimester of pregnancy. Lancet, i, 586.

Breiman, L., Meisel, W., Purcell, E. (1977). Variable kernel estimates of multivariate densities. Technometrics, 19, 135-144.

Burger, P., C., Vogel, S., F. (1973) The development of the pathologic changes of Alzheimer's disease and senile dementia in patients with Down's syndrome. Am .J. Pathol. 73, 457-476.

Cacoullos, T. (1966). Estimation of a multivariate density. Ann. Inst. Statist. Math., 18, 179-189.

Campbell S., Warsof, S., L., Little, D., Cooper, D., J. (1985). Routine ultrasound screening for the prediction of gestational age. Obstet. Gynecol, 65, 613-620.

Canick J., A., Knight G., J., Palomaki, G., E., Haddow, J., E., Cuckle, H., S., Wald, N., J. (1988) Low second trimester maternal serum unconjugated oestriol in pregnancies with Down's syndrome. Br. J. Obstet. Gynaecol, 95, 330-333.

Chow, Y., S., Geman, S., Wu, L., D. (1983). Consistent cross-validation density estimation. Ann. Statist., 11, 25-38.

Crossley, J., A., Aitken D., A., and Conner J., M. (1993). Second trimester unconjugated oestriol levels in maternal serum from chromosomally abnormal pregnancies using an optimised assay. Prenat. Diagn., 13, 271-280.

Cuckle, H., S., Wald, N., J., Thompson, S., G. (1987) Estimating a woman's risk of having a pregnancy associated with Down's syndrome using her age and serum alpha-fetoprotein level. Br. J. Obstet. Gynaecol. 94 : 387 - 402.

Davies, C.,J., Spencer, K., Selby, C., Aiken, D., Bartels, I., Reinsburg, J., Wright, D., E., and Lewis, G. (1991) Pre-natal screening for Down's syndrome: interim results of a mulitcentre retrospective trial, Proceedings of the ACB National Meeting. Glasgow, B67.

Deheuvels, P. (1977). Estimation nonparametrique de la densite par histogrammes generalises. Rev. Statist. Appl., 35, 5-42.

DiPietro, J., A., Allen, M., C. (1991) Estimation of gestational age: Implications for developmental research, Child Development, 62, 1184-1199.

Down, J., L., H. (1866) Observations on an ethnic classification of idiots. London Hospital Clinical Lectures and Reports, 3:259-262.

Efron, B. (1979) Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1-26.

Efron, B. (1982) The Jackknife, the Bootstrap and other Resampling Plans. Philadelphia:SIAM.

Efron, B. (1983) Estimating the error rate of a prediction rule: improvements on cross-validation. J. Amer. Statist. Assoc. 78, 316-331.

Efron, B., and Tibshirani, R. (1993). An introduction to the bootstrap.Chapman and Hall.

Ellis, A. (1993). Antenatal screening for Down's syndrome- Can we do better ? Ann.Clin.Biochem. 30, 421-424.

Epanechnikov, V., A. (1969). Nonparametric estimation of a multidimensional probability density. Theory Probab. Appl., 14, 153-158.

Feluch, W. and Koronacki, J. (1992). A note on modified cross-validation in density estimation. Computational Statistics & Data Analysis, 13, 143-151.

Fix, E. and Hodges, J., L. (1951). Discriminatory analysis, nonparametric estimation: consistency properties. Report No. 4, Project No. 21-49-004, USAF, School of Aviation Medicine, Randolph Field, Texas.

Fryer, M., J. (1977). A review of some nonparametric methods of density estimation. J. Inst. Maths. Applics., 20, 335-354.

Fukunaga, K. (1972). Introduction to statistical pattern recognition. New York: Academic Press.

Fuller, W., A. (1980). Properties of some estimators for the errors-in-variables model. Ann. Statist. 8, 407 - 422.

Gleser, .L., J. (1981). Estimation in a multivariate "Errors in variables ". Regression model Large sample results. The Ann. Statist., 9, 24-44.

Habbema, J., D., F., Hermans, J., van der Broek. (1974). A stepwise discriminant analysis program using density estimation. Compstat., Proc. Computational Statistics. Vienna: Physica-Verlag, 101-110.

Hall, P. (1983). Large sample optimality of least squares cross-validation in density estimation, Ann. Statist., 11, 1156-1174.

Hand, D., J. (1981). Discrimination and Classification. Wiley series in probability and mathematical statistics.

Härdle, W. (1991). Smoothing techniques with implementation in S. New York: Springer-Verlag-Verlag.

Hecht, C., A., and Hook, E., B. (1994). The imprecision in rates of Down syndrome by 1-year maternal age intervals: A critical analysis of rates used in biochemical screening,. Prenat. Diagn., 14, 729-738.

Heyl, S.,P., Miller, W., Canick, J., A (1990). Maternal serum screening for aneuploid pregnancy by alpha-fetoprotein, hCG and unconjugated estriol. Obstet Gynecol, 76, 1025-31.

Jacobson, C., B. and Barter, R., H. (1967). Intrauterine diagnosis and management of genetic defects. Am. J. Obstet. Gynecol., 99, 796-807.

Johnson, A., Cowchock, F., Darby, M., Wapner, R., Jackson, L. (1991). First trimester maternal serum alpha-fetoprotein and chorionic gonadotriphin in aneuploid pregnancies. Prenat. Diagn. 11, 443-450.

Knight G., J. (1991) Maternal Serum α - Fetoprotein Screening. Techniques in Diagnostic Human Biochemical Genetics. A Laboratory Manual. Ed. Hommes, A., H..

Kratzer, P., G., Golbus, M., S., Schonberg, S., A., Heilbron, D., C., Taylor, R., N. (1992). Cytogenetic evidence for enhanced selective miscarriage of trisomy 21 pregnancies with advancing maternal age. Am. J. Med. Genet, 44, 657-663.

Lachenbruch, P., A. (1975) Discriminant Analysis. New York: Hafner Press.

Lachenbruch, P., A., and Mickey, M., R. (1968). Estimation of error rates in discriminant analysis. Technometrics, 10, 1-11.

Lamson, S., H., and Hook, E., B. (1981). Comparison of mathematical models for maternal age dependence of Down's syndrome rates. Hum. Genet. 59, 232-234.

Lejeune, J. (1979) Investigations biochemiques et trisomie 21. Ann. Genet. (Paris), 22, 67-75.

Loftsgaarden, D., O. and Quesenberry, C., P.(1965). A nonparametric estimate of a multivariate density function. Ann. Math. Statist.., 36, 1049-1051.

Lynch, L. and Berkowitz, R., L. (1992). Amniocentesis, skin biopsy, and umbilical cord blood sampling in the prenatal diagnosis of genetic diseases. In: Reece, E.A., Hobbins, J., C., Mahoney, M., J., Petrie, R., H.(Eds). Medicine of the fetus and mother. Philadelphia: J., B. Lippincott Company, 641-652.

Macri, J., N., Kasturi, R., V., Krantz, D., A., Cook, E., J., Moore, N., D., Young, J., A., Romero, K., Larsen, J., W. (1990). Maternal Serum Down's Syndrome screening : free β-protein is a more effective marker than human chorionic gonadotrophin : Am. J. Obstet. Gynecol, 163, 1248-1253.

MacDonald, M., L., Wagner, R., M., Slotnick, R., N. (1991). Sensitivity and specificity of screening for Downs Syndrome with Alpha-fetoprotein, HCG, Unconjugated Estriol, and maternal Age.Obstet. Gynecol., 77, 63-68.

Marron, J.,S. and Wand, P., M. (1992). Exact mean integrated squared error. Ann. Statist., 20, No. 2, 712-736.

Martin, G.,M. (1978) Genetic syndromes in man with potential relevance to the pathobiology of aging. Birth Defects 14/1,5-39.

McLachlan, G., J. (1992) Discriminant analysis and statistical pattern recognition. Wiley series in probability and mathematical statistics.

Merkatz, I., R., Nitowsky, H., M., Macri, J., N., Johnson, W., E. (1984). An association between low maternal serum alpha-fetoprotein and fetal chromosome abnormalities. Am. J. Obstet. Gynecol. 148, 886-891.

Nyquist, H. (1987). Robust Estimation of the structural errors-in-variables model. Metrika, 34, 177-183.

Parvin, C., A. (1991) Estimating the performance characteristics of quality control procedures when the error persists until detection. Clin. Chem.. 37, 1720-1724.

Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist., 33, 1065-1076.

Penrose, L., S. (1934). The relative aetological importance of birth order and maternal age in mongolism. Proceedings of the Royal Society of Biology, 115, 431-450.

Reynolds, T., M., Penney, M., D., Hughes, H., John, R. (1991). The effect of weight correction on the risk calculations for Down's syndrome screening. Ann. Clin. Biochem., 28, 245-249.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. Ann. Math. Statist., 27, 832-837.

Rossiter, J., E. (1991). Calculating centile curves using kernel density estimation methods with application to infant kidney lengths. Statistics in Medicine, 10, 1693-1701.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. Scand. J. Statist., 9, 65-78.

Schuster, E., F. and Gregory, C., G. (1981). On the nonconsistency of maximum likelihood nonparametric density estimators. In Eddy, W. F., computer science and statistics: Proceedings of the thirteenth Symposium on the interface. New York, Springer-Verlag, 295-298.

Scott, D., W. and Factor., L., E. (1981). Monte Carlo study of three data-based nonparametric density estimators. J. Am. Statist. Assoc., 76, 9-15.

Scott, D., W. and Terrell, G., R. (1987). Biased and unbiased cross-validation in density estimation. J. Am. Statist. Assoc., 82, No.400, 1131-1146.

Séguin, E.(1846) Le traitement moral, l'hygiene et l'education des idiots. Bailliére, Paris.

Sheather, S., J. and Jones, M., C. (1991). A reliable data-based band width selection method for kernel density estimation. J. R. Statist. Soc. B, 53, No. 3, 638-690.

Silverman, B., W. (1986) Density estimation for statistics and data analysis. Monographs on Statistics and applied probability. Chapman and Hall.

Snijders, R.,J.,M. (1993) Screening by ultrasound for fetal chromosomal abnormalities.

Spencer, K. (1991). Evaluation of an assay of the free β-subunit of choriogonadotropin and its potential value in screening for Down's Syndrome. Clin. Chem. 37, 809-814.

Staples, A., J., Robertson, E., F., Ranieri, E., Ryall, R., G., Haan, E., A. (1991). A maternal screen for Trisomy 18 : An extension of maternal serum screening for Down Syndrome. Am. J. Hum. Genet. 49, 1025-1033.

Stone, C., J., (1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., 12, 1285-1297.

Stute, W. (1992). Modified cross-validation in density estimation. J. Statist. Plann. Inf., 30, 293-305.

Tabor, A., Philip, J., Madsen, M., Bang, J., Obel, E., B., Norgaard-Pederson, B. (1986). Randomized control trial of genetic amniocentesis in 4606 low-risk women. Lancet, 1, 1287 - 93.

Terrell, G., R., and Scott, D., W,. (1992). Variable kernel density estimation. Ann. .Statist., 20, No. 3, 1236-1265.

Titterington, D.,M., Murray, G.,D., Murray, L., S., Spiegelhalter, D., J., Skene, A., M., Habbema, J., D., F., Gelpke, G., J. (1981). Comparison of discrimination techniques applied to a complex data set of head injured patients. J. R. Statist. Soc. A. 144, 145-174.

Van Lith, J., M., M. (1994) First trimester screening for Down's syndrome.

Wald, N., J., Cuckle, H., S., Densem, J.W., Nanchahal, K., Royston, P., Chard, T., Haddow, J., E., Knight, G., J., Palomaki, G., E., Canick, J., A. (1988). Maternal serum screening for Down's syndrome in early pregnancy, Br. Med. J., 297, 883-887.

Wald, N., J., Cuckle, H., S., Densem, J.,W., Kennard, A., Smith, D. (1992). Maternal serum screening for Down's syndrome : the effect of routine ultrasound scan determination of gestational age and adjustment for maternal weight. Br. J. of Obstet. and Gynaecol, 99, 144 - 149.

Wald, N.J., Kennard, A., Densem, J., W., Cuckle, H., S., Chard, T., Butler, L. (1992a). Antenatal maternal serum screening for Down's syndrome: results of a demonstration project Br.Med. J. 305, 391-394.

Wald, N.,J., Densem, J.,W. (1993). Letter to the editor. Prenatal diagnosis, 13,000-000.

Wand, M., P. and Jones, M., C. (1993). Comparision of smoothing parameterizations in bivariate kernel density estimation. J. Am. Statist. Assoc., 88, No. 422, 520-528.

Wright, D., E., Reynolds, T., M., Donovan, C., M. (1993). Assessment of atypicality : an adjunct to prenatal screening for Down's Syndrome that facilitates detection of other abnormalities. Ann. Clin. Biochem., 30. 578-583.

Wright D., Donovan C., and Davies C. (1993a). Uncertainty in reported detection rates for Down's Syndrome screening: Mathematical modelling techniques to obtain confidence intervals. In Proceedings of the 4^{th} conference on Endocrinology and Metabolism in Human Reproduction: Screening for Down's Syndrome, Eds. Grudzinskas, J., Chard, T., Chapman, M., and Cuckle, H.). Cambridge University Press, Cambridge.

Wright, D.,E., Davies, C. (1995) Improved methodology for establishing and maintaining normal median values in Prenatal Screening programs. In print.

Wright, D., E. (1995). Parametric distortions of a nonparametric density estimate. Technical report No. 6, School of Mathematics and Statistics, University of Plymouth.

Youings, S., Gregson, N., Jacobs, P. (1991). The efficacy of maternal age screening for Down's syndrome in Wessex, Prenat. Diagn., 11, 419-425.

Zeitune, M., Aitken, D., A., Crossley, J., A, Yates, J., R., W., Cooke, A., Ferguson-Smith, M., A. (1991). Estimating the risk of a fetal autosomal trisomy at mid-trimester using maternal serum alpha-fetoprotein and age: a retrospective study of 142 pregnancies. Prenat. Diagn. 11, 847-857.