2010

A NUMERICAL METHOD FOR ESTIMATING THE VARIANCE OF AGE AT MAXIMUM GROWTH RATE IN GROWTH MODELS

Semhar Ogbagaber
Virginia Commonwealth University

A NUMERICAL METHOD FOR ESTIMATING THE VARIANCE OF AGE AT

MAXIMUM GROWTH RATE IN GROWTH MODELS

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of

Science at Virginia Commonwealth University.

by

SEMHAR BEYENE OGBAGABER

M.A. Mathematical Statistics, Bowling Green State University, 2007

Co-Director: VISWANATHAN RAMAKRISHNAN, PH.D.

ASSOCIATE PROFESSOR, DEPARTMENT OF BIOSTATISTICS

Co-Director: SHUMEI S SUN, PH.D

PROFESSOR AND DEPARTMENT CHAIR, DEPARTMENT OF BIOSTATISTICS

Virginia Commonwealth University

Richmond, Virginia

May 2010

Acknowledgement

Above all I would like to thank my mother and father for all the sacrifices. Dr. Ramesh, thank you for pushing me to focus and put more effort in the thesis. Your mentorship helped me build character. The statistical knowledge that I gained while doing the thesis project is invaluable. Finally, I would like to thank the Department of Biostatistics and School of Graduate Studies for generously funding my education and teaching me statistics.

Table of Contents

Abstract

A NUMERICAL METHOD FOR ESTIMATING THE VARIANCE OF AGE AT

MAXIMUM GROWTH RATE IN GROWTH MODELS

By Semhar Beyene Ogbagaber, M.A.

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of

Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2010

Major Directors

Viswanathan Ramakrishnan

Associate Professor Department of Biostatistics

Shumei S Sun

Professor Department of Biostatistics

Most studies on maturation and body composition using the Fels Longitudinal data mention peak height velocity (PHV) as an important outcome measure. The PHV is often derived from growth models such as the triple logistic model fitted to the stature (height) data. The age at PHV is sometimes ordinalized to designate an individual as an early, average or late maturer. In theory, age at PHV is the age at which the rate of growth reaches the maximum. Theoretically, for a well behaved growth function, this could be obtained by setting the second derivative of the growth function to zero and solving for age. Such a solution would obviously depend on the parameters of the growth function. An estimate of the age at PHV would be a function of estimates of these parameters.

Since the estimates of age at PHV are ultimately used as a predictor variable for analyzing adulthood outcomes, the uncertainty in the estimation of the PHV inherent due to the uncertainty in the estimation of the growth model need to be accounted for. The asymptotic s.e. of the age at maximum velocity in simple growth models such as the logistic and the Gompertz models could be explicitly obtained because explicit formulas for the age at maximum velocity are available. In this thesis a numerical method is proposed for computing the s.e. of the age at PHV for those that do not lead to explicit solutions for the age at PHV. The accuracy of this method is demonstrated by computing the s.e. using the explicit method as well as the proposed numerical methods and by comparing them. Incorporating the estimates of the s.e. in regression models that use age at PHV as predictor is illustrated using the FELS data.

1   Introduction

The purpose of this chapter is to briefly introduce nonlinear regression models in the context of growth, describe methods for estimating the parameters of the model and list the scope of this thesis . Use of nonlinear models span across many disciplines such as agricultural research, biology, engineering and other applied disciplines. The study and development of methods that deal with nonlinear behaviors is crucial and has been extensively dealt with in the literature. In nonlinear regression a response, Y, is modeled as a function of an explanatory variable, X . Graphically, a nonlinear relationship between X and Y exists when data gathered display a significant departure from a linear relationship. For example, when Y is a measure of hormone level in rats at different dose levels (X) an S shaped dose response relationship is observed. In this instance, a nonlinear regression model is used to quantify the effectiveness of the drug that causes a response halfway between the minimum and maximum responses.

Most frequently, the interest lies in fitting nonlinear relationships of known mechanisms [Seber & Wild (2003)]. For example, for describing growth, models are developed based on mathematical functions that describe growth mechanisms of biological organisms over time. Examples of such models are Gompertz, logistic, double logistic and triple logistic, which will be further described in chapter 2. Fitting these models to growth data have been considered extensively in the statistical literature [Bock et al. (1973), Thissen et al. (1976), Kingsland (1982)].

To describe the general approach to fitting growth models, consider growth data y_t, observed at time x_t. Suppose there is a known nonlinear function describing the growth over time, $f(x_t; \boldsymbol{\theta})$. The nonlinear regression model may be written,

$$y_t = f(x_t; \boldsymbol{\theta}) + \varepsilon_t, \tag{1.1}$$

for $t = 1, 2, \ldots, n$, where $E(\varepsilon_t) = 0$, and $\boldsymbol{\theta}$ is a $p \times 1$ vector of parameters. In the simplest form of the model the errors are assumed to be independent over time. The estimate of $\boldsymbol{\theta}$, in general, is obtained by nonlinear least squares approach. That is, they are obtained by minimizing the sum of squares,

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left[y_t - f(x_t; \boldsymbol{\theta}) \right]^2 \tag{1.2}$$

over $\boldsymbol{\theta}$. The essence of least squares is that the most likely values of the vector $\boldsymbol{\theta}$ of parameters, say $\hat{\boldsymbol{\theta}}$, are attained such that the curve defined by $\hat{\boldsymbol{\theta}}$ is closest to the observed data. There is no assumption about the distribution of experimental errors in least squares estimation but if the residuals follow a Gaussian distribution, the least squares estimators are equivalent to the maximum likelihood estimators [Myers et al.. (2002)]. Suppose $f(x_t; \boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$. The minimum value of $S(\boldsymbol{\theta})$ occurs where its gradient defined by the first derivative is zero. To obtain this minimum value, p normal equations for the nonlinear model are set up such that $\hat{\boldsymbol{\theta}}$ is a solution of the set of simultaneous equations,

$$\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} = 0, \quad (r = 1, 2, \ldots, p). \tag{1.3}$$

That is the normal equations are,

$$\sum_{t=1}^{n} [y_t - f(x_t; \boldsymbol{\theta})] \left[\frac{\partial f(x_t; \boldsymbol{\theta})}{\partial \theta_r} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \qquad (r = 1, 2, \ldots, p). \qquad (1.4)$$

It is apparent from the above equations that the normal equations are also nonlinear in $\boldsymbol{\theta}$ and therefore often do not lead to explicit expression for $\hat{\boldsymbol{\theta}}$.

A simple example of a nonlinear model with three parameters is,

$$y_t = \beta_1 e^{(\beta_2 + \beta_3 x_t)} + \varepsilon. \qquad (1.5)$$

One of the characteristic of a nonlinear regression model is, at least one of the partial derivatives of y_t with respect to the unknown parameters in the model is a function of at least one of the parameters. Notice for the model in equation (1.5), the derivative of y_t with respect to β_1 is a function of β_1, β_2, β_3. Similarly, the derivative of y_t with respect to β_2 and β_3 is a function of β_1, β_2 and β_3. For this example, the least square estimation is illustrated as follows. For $\boldsymbol{\theta} = (\beta_1 \quad \beta_2 \quad \beta_3)'$, the least squares normal equations are,

$$\sum_{t=1}^{n} \left[y_t - b_1 e^{(b_2 + b_3 x_t)} \right] e^{(b_2 + b_3 x_t)} = 0,$$

$$\sum_{t=1}^{n} \left[y_t - b_1 e^{(b_2 + b_3)x_t} \right] b_1 e^{(b_2 + b_3 x_t)} = 0,$$

$$\sum_{t=1}^{n} \left[y_t - b_1 e^{(b_2 + b_3)x_t} \right] b_1 x_t e^{(b_2 + b_3 x_t)} = 0. \qquad (1.6)$$

Further simplification results in,

$$\sum_{t=1}^{n} y_t e^{(b_2+b_3 x_t)} - b_1 \sum_{t=1}^{n} e^{2(b_2+b_3 x_t)} = 0$$

$$\sum_{t=1}^{n} y_t b_1 e^{(b_2+b_3 x_t)} - b_1^2 \sum_{t=1}^{n} e^{2(b_2+b_3 x_t)} = 0 .$$

$$\sum_{t=1}^{n} y_t b_1 x_t e^{(b_2+b_3 x_t)} - b_1^2 \sum_{t=1}^{n} x_t e^{2(b_2+b_3 x_t)} = 0 \qquad (1.7)$$

From equation (1.7), one can deduce that explicit solutions for the parameters in terms of

y and x alone are not possible.

Therefore, nonlinear estimation requires iterative procedures. The iterative procedures

start from an initial value for $\hat{\theta}$ and make use of least square estimation by continuously

linearizing the nonlinear function at each iteration. The key process of linearization is

achieved through application of Taylor's series expansion. For any nonlinear estimation,

generally the steps are as follows:

1. Start with initial estimates for each parameter.

2. Compute the nonlinear function and the sum of squares using the current values

    of the parameters.

3. Update the values of the parameters in step 2 so that the sum of squares is

    reduced. There are several numerical methods to obtain this update. Some of

    these will be discussed further below.

4. Iterate between steps 2 and 3 until the sum of squares within a certain tolerance is reached.

5. When the tolerance is reached the values from the last iteration are the required estimates.

Step 1 is a very important step as all iteration procedures require "good" starting values for parameters to be estimated. Suitable starting values would facilitate relatively quick convergence to optimal parameter estimates. If one of the parameters could be guessed, there is apparent advantage to work with models that can be transformed to linear models For example, consider the Gompertz model of the form,

$$y = \gamma e^{-e^{-(\beta_0 + \beta_1 x)}}.$$

The initial guess for γ is obtained from the data as the maximum response. As $x \to \infty$, the value of y tends to γ which is the reason why the maximum response is an appropriate starting value for γ. Then, the above equation can be written in a linear form as,

$$-\log\left(-\log\left(\frac{y}{\gamma^{(0)}}\right)\right) = \beta_0 + \beta_1 x.$$

From the linearized form above, initial estimates for β_0 and β_1 can be obtained by

running a linear regression of the $-\log\left(-\log\left(\frac{y}{\gamma^{(0)}}\right)\right)$ on x. Finally, the initial guesses

for the three parameters would help speed up convergence to the final estimates in the nonlinear procedure.

[SAS 9.2 (2008)]The numerical methods that could be used in step 3 include Marquardt

and Levenberg, the method of steepest descent, the method of Gauss-Newton and

Newton Raphson. For example, in the Newton Raphson method, the first and second

derivatives are computed for updating the values of the parameters. The Newton

Raphson method was suggested as a modification to the Gauss-Newton method to help

deal with issues of ill-conditioning of the matrix of derivatives. This method is based on

quadratic approximation to the sum of squares function, $S(\boldsymbol{\theta})$. That is, near $\boldsymbol{\theta}^0$,

approximate,

$$S(\boldsymbol{\theta}) \approx S\left(\boldsymbol{\theta}^0\right) + \left(\frac{\partial S}{\partial \boldsymbol{\theta}}\right)^T \left(\boldsymbol{\theta} - \boldsymbol{\theta}^0\right) + \left(\boldsymbol{\theta} - \boldsymbol{\theta}^0\right)^T \left(\frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) \left(\boldsymbol{\theta} - \boldsymbol{\theta}^0\right) \tag{1.8}$$

where $\dfrac{\partial S}{\partial \boldsymbol{\theta}}$ is the derivative evaluated at $\boldsymbol{\theta}^0$ and $\dfrac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ is the Hessian of $S(\boldsymbol{\theta})$

evaluated at $\boldsymbol{\theta}^0$. To derive the Newton Raphson step, the approximating sum of squares

above (1.8) will have a stationary point when its gradient is zero. That is,

$$\frac{\partial S}{\partial \boldsymbol{\theta}} + \left(\frac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right) \left(\boldsymbol{\theta} - \boldsymbol{\theta}^0\right) = 0 \,.$$

The fixed (stationary) point will be a minimum if $\dfrac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ is positive definite [Seber &

Wild (2003)]. Thus, if $\dfrac{\partial^2 S}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ happens to be positive definite, Newton Raphson step is

given by,

$$\boldsymbol{\theta}_{new} = \boldsymbol{\theta}_{old} - \left(\frac{\partial^2 S}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} \right)^{-1} \frac{\partial S}{\partial\boldsymbol{\theta}} .$$

At each step of the iteration, the Newton Raphson step updates the previous parameter,

$\boldsymbol{\theta}_{old}$, estimate to give the current estimate, $\boldsymbol{\theta}_{new}$.

In the final step, the process stops when a convergence criteria is met. In other words, the

iteration would stop when there is no significant difference between successive parameter

estimates. The convergence criteria may be based on

$$\left| \frac{\theta_r^{m+1} - \theta_r^m}{\theta_r^m} \right| < Tolerance \qquad (r = 1, 2, \ldots, p)$$

where *Tolerance* is a very small number, say 10^{-8}. In essence, the convergence criteria

above tells the program that best-fit parameters have been achieved once the relative

improvement of successive parameter estimates is less than a fixed tolerance. Precision of

the final parameter estimates could be estimated from the inverse of the Hessian

calculated at $\hat{\boldsymbol{\theta}}$.

1.1. Age at Peak Height Velocity (PHV) and Age at Maximum Velocity (MV)

In modeling growth of children, physical maturation has been an important marker,

clinically, to describe the status of overall health. Typically, clinicians use stature to

detect and monitor breadth of health problems [Berkey (1993)]. Therefore, measuring

important characteristics of stature describing the growth would be useful. The peak

height velocity (PHV), defined as the maximum rate of growth in stature during a growth spurt, has been one of those characteristic used widely in the literature. The age at PHV has been used to identify children who mature early, moderate and late, which has been an important feature to understand the dynamics of normal growth in human beings. This in turn is useful for health professionals to understand growth disorders and to come up with remedies such as therapeutic intervention or psychological therapy.

The PHV curve is simply the first derivative of the growth curve computed over time. The age at which the velocity curve reaches the maximum, namely the PHV, is obtained by setting the derivative of the PHV curve (which is also the second derivative of the growth curve) to zero and solving. The corresponding age at PHV is where the rate of growth is considered maximum. The age at PHV is computed only for popular human growth models such as the double and triple logistic. Also, age at PHV is computed for Preece-Baines models [Preece & Heinrich(1981), Guo et al. (1992)]. However, the age at maximum velocity (MV) for simple logistic and Gompertz should be distinguished from the age at PHV for the double and triple logistic. This occurs at a relatively early age (0 to 3 years) but age at PHV usually occurs between the ages 10 to 13 years.

Age at PHV will be an important predictor for a weighted regression model in chapter 4. Its significance in the model emanates from the consensus among scientists as a biological indicator. Following estimation of age at PHV, equally important is its corresponding variance estimate. As it will be apparent in chapter 4, the standard error

estimate for age at PHV will be used as a weight variable when using weighted

regression to model levels of glucose and insulin.

1.2. Variance of Age at PHV

By assumption, the explanatory variables, x_t, from a regression model are considered to

be free of error. This happens when x_t are predetermined or fixed in an experimental

setting. But, often in real life problems, x_t, could be measured with error. The error

could arise from biological or physical variability. For instance, when measuring height,

age and other variables of interest in a longitudinal study, there is a joint variation in the

explanatory variables and the response variable. In this case, the relationship established

from the ordinary regression model could be in serious error. Including the variability in

the age at PHV used as an explanatory variable in subsequent modeling of adulthood

outcome measures could lead to improvements in the analysis [Ratkowsky (1983)].

In this thesis a method for estimating the variance and application of the variance in the

subsequent analyses is proposed.

1.3. Data Description

The data used in this study is obtained from the Fels Longitudinal Study [Roche (1992)].

The Fels Longitudinal Study is an ongoing study that involves measuring subjects from

childhood to adulthood. Originally, the Fels Longitudinal Study began with an interest in

studying the effects of The Great Depression on child development both in terms of physical and pyschological. The participants, called the "Fels Family," all have been committed to this research and have enabled researchers to gather a wealth of data on them, including physical growth, maturation and adulthood measures, for a number of years. Currently, Fels Longitudinal Study is the biggest serial study in the U.S.A. Since its inception in 1929, the study has enrolled and successively recorded data on 1923 participants at different intervals.

For this thesis, the Fels data collected over the 79 years from 1929 to 2007 will be used. The original sample contained a total of 1923 subjects, 930 of whom were boys and 993 were girls observed from birth to 18 years of age. Data on growth were, with some exceptions, collected at fairly regular intervals.   Measures of stature characteristics used for analysis included height, weight, sex, glucose level and insulin level. Since, the purpose of this thesis is to demonstrate the usefulness of the estimate of the variance of age at PHV in modeling adulthood outcomes, a subset of the above data for which the adulthood outcomes were available was considered. The subset data that contained information on glucose and insulin level included 632 subjects, of which 313 were males and 319 were females.

Over the duration of 79 years there were some new entrants and some others who withdrew from the study for a variety of reasons. Some subjects were missing in certain years contributing to missing data. Also, for some subjects there was no sufficient data

available for analysis. As a result, only subjects with complete and valid data were available for analysis. Further, the selection of legitimate subjects for analysis depends on the convergence criteria of the different growth models under consideration. When fitting the growth models, height is the response variable. It is an anthropometric measurement of interest in this work since age at PHV is derived using the functional relationship for height. Each patient is identified using a unique patient identifier. Full details of all anthropometric measurements in the whole Fels study can be obtained in Roche (1992).

The chapter 2 deals with the introduction of the common growth models, namely logistic, Gompertz, double logistic and triple logistic. Also, in this chapter, characteristics of these models as well as estimation of the variance of age at PHV using a Taylor series approximation (i.e. Delta method) are described. Using this method, for the logistic and Gompertz models explicit forms for the variances for age at maximum velocity (MV) are obtained. For the double and triple logistic models, a numerical method is proposed because age at PHV does not have a closed form in these models and subsequently explicit expressions for the variance are also not available. In chapter 3, first the accuracy of the proposed numerical method is demonstrated by applying it to the logistic and Gompertz models and by comparing the results with the explicit forms available in these two cases. Then, in the same chapter, the method is applied to the triple logistic model. In chapter 4, the results from the analysis in which the estimated standard errors for the age at PHV were used in a weighted regression to model the levels of glucose and

insulin are presented and discussed. In the last chapter, some recommendations and future work are outlined.

2   Methodology

2.1. Introduction

In this chapter, methods for estimating age at MV and age at PHV are described and a new method for estimating its variance is proposed. The methods are discussed for four typical growth models, namely, logistic, Gompertz, double logistic and triple logistic models. The estimates of age at MV and their variances are obtained explicitly for logistic and Gompertz models, while the methods to estimate age at PHV and variances for double and triple logistic are numerical.

The logistic and Gompertz models fit situations subject to S-shaped or sigmoidal growth. Sigmoidal curves start from an initial point and increase continuously with time until it reaches an inflection point where the growth starts to decrease, and eventually reaches an asymptote at maturity. The sigmoidal growth occurs in a variety of areas, such as economic, agriculture, and biological applications. In biology, growth of bacterial colonies, cellular growth of a plant, physical development in children and development of birds are some examples that are dictated by sigmoidal growth models [Kingsland (1982), Brisbin et al. (1987)]. They state, the methodology of growth curve fitting has contributed significantly to the understanding of the mechanisms of biological growth. Using a given sigmoid curve, it seems plausible to describe growth and development of many organisms mathematically [Bogin (1999)].

In Figure 1, examples of important features of growth are shown. Figure 1a) shows a logistic growth function. Figure 1b) shows the corresponding velocity curve, which is essentially the first derivative of the growth curve plotted against time. In Figure (c) the acceleration curve, which is the second derivative of the growth curve plotted against time, is shown. The age at maximum velocity (MV) is where the maximum rate of growth occurs. A single peak is indicative of an initial growth followed by a period of deceleration. For some growth functions there may be more than one peak and therefore there may be more than one estimate of the age at maximum velocity. The age at which the PHV occurs is of interest. In Figure 1 (c), where the acceleration is zero corresponds to the point of inflection on the velocity curve.
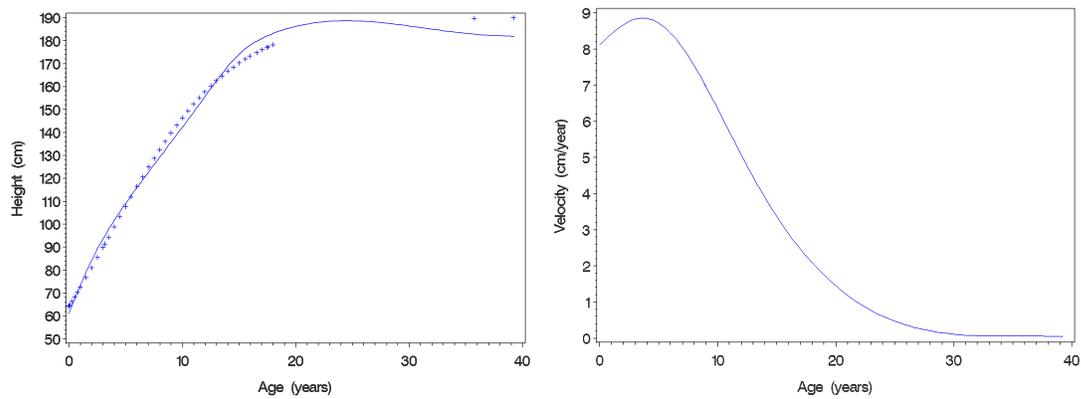
(a)

(b)
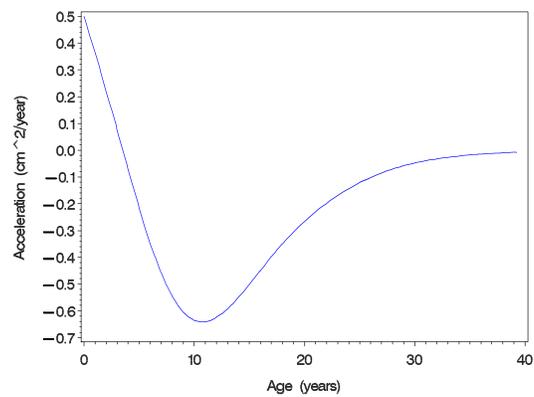
(c)

Figure 1 Logistic growth model. (a) Growth curve with $y = k$ final size and point of inflection. (b) Growth rate curve with maximum growth rate at t = age at MV. (c) Acceleration curve.

2.2. The Simple Logistic Model

The logistic growth model was suggested by the Belgian mathematician Pierre-Francois Verhulst in 1838 to describe growth of population or organisms empirically over time [Seber & Wild (2003), p.330)]. Subsequently, the logistic model was put forth as "the law of growth" by Professor Raymond Pearl in 1920 [Kingsland (1982)]. However, the logistic growth model has been criticized and has not been accepted as a law by later scientists, in different disciplines, mostly due to the lack of its prediction power. Unlike some physical laws, it was shown that an experiment involving the logistic could not be replicated under normal conditions [Kingsland (1982)]. Although it was initially rejected by statisticians, economists and biologists, the logistic model eventually became a central tool due to ecologists' gradual recognition of the intrinsic usefulness of the curve for describing growth [Kingsland (1982)].

The logistic function used to model individual growth could be written,

$$ y = \frac{k}{(1 + ae^{-\beta_1 t})}, \qquad (2.1) $$

where, y may be the stature in centimeters and t, age in years. The logistic function has asymptotes at $y = 0$ as $t \to -\infty$ and $y = k$ as $t \to \infty$. Thus, the initial growth is zero and the maturity in stature is attained at $y = k$.

As discussed in chapter 1, given data, the estimation of the parameters of the logistic model, which is nonlinear, has to be achieved numerically. This requires specification of

initial values for the parameters in order to ensure quick convergence of the numerical

method. An approach for providing good initial values, so that the numerical algorithm

converges quickly might be as follows. One could graph the data and approximately read

off the growth at maturity, k, denoted as $k^{(0)}$, by observing where the growth tapers.

However, this requires a visual examination and therefore does not lend itself to

automation. Therefore, if the visual method is not practical, a cruder initial estimate of k

might simply be the largest (maximum) of the observed data. Fixing this value, obtain

the initial values for parameters a and β, by linearizing the logistic function as,

$$z_0 = \log\left(\frac{k^{(0)}}{y} - 1\right) = \log(a) - \beta_1 t, \tag{2.2}$$

and fitting a linear model. Once the initial values are obtained in this fashion numerical

algorithms, such as the Marquardt and Levenberg nonlinear procedures, could be used to

obtain the least square estimates.

2.2.1. Computing the age at MV

To compute the age at maximum velocity (MV), the logistic function is differentiated

twice, equated to zero and solved for t. Solving the second derivative produces the

inflection point or the age which optimizes the velocity curve. Equation (2.3) displays

the first derivative of the simple logistic equation.

$$\frac{\partial y}{\partial t} = \frac{ka\beta_1 e^{-\beta_1 t}}{(1 + ae^{-\beta_1 t})^2}. \tag{2.3}$$

Therefore, the second derivative of the logistic function is,

$$\frac{\partial^2 y}{\partial t^2} = \frac{-(1+ae^{-\beta_1 t})^2(ka\beta_1^2 e^{-\beta_1 t}) + (1+ae^{-\beta_1 t})(2ka^2\beta_1^2 e^{-\beta_1 t})}{(1+ae^{-\beta_1 t})^4}. \qquad (2.4)$$

Simplifying equation (2.4) further results in,

$$\frac{\partial^2 y}{\partial t^2} = \frac{-ka\beta_1^2 e^{-\beta_1 t}(1+ae^{-\beta_1 t}) + 2ka^2\beta_1^2 e^{-2\beta_1 t}}{(1+ae^{-\beta_1 t})^3}. \qquad (2.5)$$

Setting the equation (2.5) to zero and solving for age at PHV leads to,

$$ka\beta_1^2 e^{-\beta_1 t}(1+ae^{-\beta_1 t}) = 2ka^2\beta_1^2 e^{-2\beta_1 t}.$$

Simple algebraic manipulation yields,

$$1+ae^{-\beta_1 t} = 2ae^{-\beta_1 t}.$$

Further simplification yields,

$$ae^{-\beta_1 t} = 1.$$

Subsequently, transforming the above equation using the natural logarithm gives,

$$-\beta_1 t = \ln\left(\frac{1}{a}\right).$$

Thus the age at MV is estimated by,

$$t = \frac{-\ln\left(\frac{1}{a}\right)}{\beta_1}.$$

2.3. The Gompertz Growth Curve

The Gompertz model was first introduced by Gompertz in 1825 for the hazard in life tables. Then Medawar (1940) used it as a growth model for modeling the growth of heart

in chicken [Seber & Wild (2003)]. Much like the logistic curve, the Gompertz has been used extensively in the literature to fit growth data of several kinds. Richards in 1959 derived equation (2.6) which became known as Richards function [Seber & Wild (2003)]. Both the logistic and the Gompertz are a form of Richards function [Preece & Heinrich (1981)]. The Richards function is,

$$y = A\left(1 \pm be^{-kt}\right)^{-1/n},$$
(2.6)

If $n = 1$, Richards function reduces to the logistic and as $n \to 0$, it is equivalent to the Gompertz. The Gompertz growth model also follows a sigmoidal growth pattern, where the growth rate decreases exponentially. Like most sigmoidal growth models, the Gompertz model initially starts with rapid growth before it gradually approaches a limit. Unlike the logistic model, the Gompertz model is asymmetrical about the point of inflection.

Following is a versatile and useful three-parameter formulation of the Gompertz function with a single inflection point,

$$y = \alpha e^{-e^{(\beta - \gamma t)}}.$$
(2.7)

The final maturity height is attained at α. The parameter $\gamma > 0$ describes the decay in the specific growth rate and the parameter $e^{\beta} > 0$ controls the difference between the initial and the final height. Below is a pictorial display of a Gompertz function modeling the patient height growth over time (from Fels data).
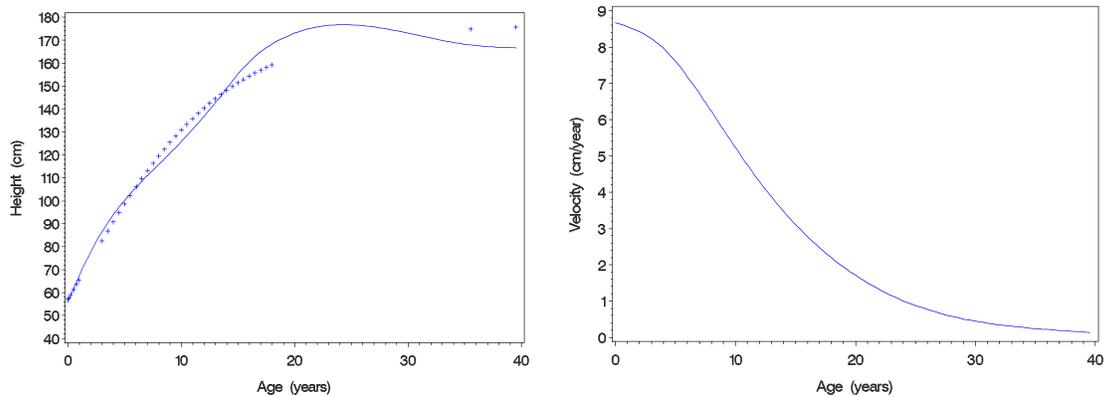
(a)

(b)
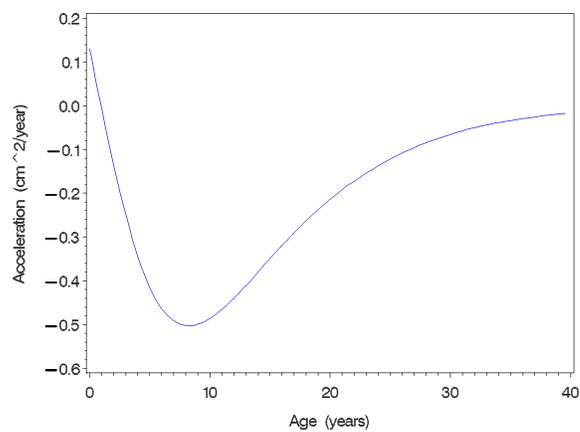
(c)

Figure 2 Gompertz growth model. (a) Growth curve with $y = \alpha$ final size and point of

inflection $t = \dfrac{\beta}{\gamma}$. Superimposed dotted plot is the predicted growth curve. (b) Growth

rate curve with maximum growth rate at t = age at MV. (c) Acceleration curve.

Once again estimation of the parameters in the model from a given data set needs to be done using numerical iterative methods. As in the logistic case, setting the initial values for the iterative algorithm judiciously is important for a speedy convergence. The steps to obtain initial parameters for the Gompertz model may be as follows. To facilitate estimation of initial values for β and γ, a raw estimate for α is vital in equation (2.8). From equation (2.7), maximum y is attained as $t \to \infty$. Therefore, the maximum observed value can be used as the estimate for $\alpha^{(0)}$. For a given $\alpha^{(0)}$, (2.7) could be rewritten as the following linear model;

$$z_0 = \log\left(-\log\left(\frac{y}{\alpha^{(0)}}\right)\right) = \beta - \gamma t. \tag{2.8}$$

Now fitting a linear regression model of z_0 versus t the initial values for parameters β and γ, denoted as $\beta^{(0)}$ and $\gamma^{(0)}$, respectively could be obtained.

2.3.1. Estimating the Age at MV

Differentiating equation (2.7) with respect to t yields,

$$\frac{\partial y}{\partial t} = \alpha e^{-e^{(\beta-\gamma t)}}\left[-e^{(\beta-\gamma t)}\right](-\gamma) = \alpha\gamma e^{(\beta-\gamma t)}e^{-e^{(\beta-\gamma t)}}. \tag{2.9}$$

To find the age at MV, the second derivative with respect to t is set to zero. It follows that the second derivative of expression (2.9) with respect to t is,

$$\frac{\partial^2 y}{\partial t^2} = \alpha\gamma e^{(\beta-\gamma t)}(-\gamma)e^{-e^{(\beta-\gamma t)}} + \alpha\gamma e^{(\beta-\gamma t)}\left[e^{-e^{(\beta-\gamma t)}}\left(-e^{(\beta-\gamma t)}\right)(-\gamma)\right]. \tag{2.10}$$

Simplifying (2.11) and equating to zero yields

$$-\gamma^2\alpha e^{(\beta-\gamma t)}e^{-e^{(\beta-\gamma t)}} + \gamma^2\alpha e^{2(\beta-\gamma t)}e^{-e^{(\beta-\gamma t)}} = 0$$

$$e^{-e^{(\beta-\gamma t)}} = e^{(\beta-\gamma t)}e^{-e^{(\beta-\gamma t)}}$$

$$1 = e^{(\beta-\gamma t)}.$$

Taking the natural logarithm on both sides and solving for t yields,

$$t = \frac{\beta}{\gamma}.$$

Thus the age at MV is $t = \dfrac{\beta}{\gamma}$.

In applications the age at MV in the two models described above is obtained by

substituting the corresponding nonlinear least squares estimates or the maximum

likelihood estimates (mles) of the parameters. Next an asymptotic approximation to the

variance of the estimate of age at MV using Taylor's series expansion (also called the

delta method), is described.

2.3.2. Asymptotic Variance of mle using Delta Method

Suppose for a scalar valued random variable the distribution depends on a parameter θ.

Suppose the maximum likelihood estimate of θ is $\hat{\theta}$, and the corresponding variance is

$V(\hat{\theta})$. Then the maximum likelihood estimate of a function $\psi = g(\theta)$, by invariance

property of the mle, is $\hat{\psi} = g(\hat{\theta})$. Then by the delta method [Bickel and Doksum (2001)]

estimate for the variance of $\hat{\psi}$ is given by,

$$V(\hat{\psi}) = \left(\frac{\partial \psi}{\partial \theta}\right)^2 V(\hat{\theta}).$$

An estimate of this variance could be obtained by substituting $\hat{\theta}$ for θ in the right hand side of the above equation. To extend to the multiparameter case, suppose $\boldsymbol{\theta}$, is a p x 1 vector valued parameter. Let the mle of $\boldsymbol{\theta}$ be $\hat{\boldsymbol{\theta}}$ and the corresponding p x p variance covariance matrix be denoted by $\mathbf{V}(\hat{\boldsymbol{\theta}})$.

Suppose $\mathbf{g}$ is a set of d (linear or nonlinear) functions of the elements of the parameter $\boldsymbol{\theta}$. Then as in the scalar case the mle of $\mathbf{g}$, $\hat{\mathbf{g}}$, could be obtained using the invariance property of the mle by substituting the mle $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$. The multiparameter version of the delta method for obtaining the d x d variance covariance matrix of the mle $\hat{\mathbf{g}}$ is given by the equation,

$$\mathbf{V}\left(\mathbf{g}(\hat{\boldsymbol{\theta}})\right) = \left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\right) \mathbf{V}(\hat{\boldsymbol{\theta}}) \left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\right)^T, \qquad (2.11)$$

where $\left(\dfrac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\right)$ is the d x p matrix of partial derivatives of the elements of $\mathbf{g}$ with respect to the elements of $\boldsymbol{\theta}$. Once again the estimates of the variance could be obtained by substituting the corresponding mle's.

In this thesis, the multivariate delta method becomes a useful tool to find the variance of the estimated age at PHV or age at MV for the growth models because the age at PHV or age at MV is a function of a vector of parameters.

2.4. The Delta Method for Simple Logistic

Consider the application of the delta method in estimating the variance of age at MV for logistic and Gompertz. Once the age at MV is computed, the delta method is applied directly. For logistic growth curve, the estimator of age at MV is $t = \dfrac{-\ln\left(\dfrac{1}{a}\right)}{b_1}$. Here, $\mathbf{g}$ is given by t, which is a nonlinear function of the model parameters k, a and b_1. The first term on the right hand side of the variance expression in (2.11) is a row vector containing partial derivatives of t with respect to k, a and b_1. Thus, to implement the delta method in the logistic case, $\left(\dfrac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}\right)$ is $\left(\dfrac{\partial g}{\partial k} \quad \dfrac{\partial g}{\partial a} \quad \dfrac{\partial g}{\partial b_1}\right)$, which is equal to

$$\left(0 \quad \frac{1}{ab_1} \quad \frac{\ln\left(\dfrac{1}{a}\right)}{b_1^2}\right).$$

Suppose the variance-covariance matrix of the estimates of the three model parameters, (which is produced by most software along with the estimates, such as in PROC NLIN in SAS) could be obtained as $\begin{pmatrix} \sigma_k^2 & \sigma_{ka} & \sigma_{kb_1} \\ \sigma_{ak} & \sigma_a^2 & \sigma_{ab_1} \\ \sigma_{b_1k} & \sigma_{b_1a} & \sigma_{b_1}^2 \end{pmatrix}$, where the diagonal elements are the variances and the off diagonal elements are the covariances. Plugging in the partial

derivatives and the variance covariance matrix in (2.11) and simplifying, the exact

variance estimate of the age at MV for the logistic model could be shown to equal,

$$V(t) = \frac{2}{b_1(2a-1)} \left(\frac{2\sigma_a^2}{b_1(2a-1)} + \frac{\ln\left(\frac{1}{2a-1}\right)\sigma_{b_1 a}}{b_1^2} \right)$$

$$+ \frac{\ln\left(\frac{1}{2a-1}\right)}{b_1^2} \left(\frac{2\sigma_{ab_1}}{b_1(2a-1)} + \frac{\ln\left(\frac{1}{2a-1}\right)\sigma_{b_1}^2}{b_1^2} \right).$$

2.5. The Delta Method for Gompertz

For Gompertz growth model, the estimator of age at MV is $t = \frac{\beta}{\gamma}$. Execution of the delta

method requires, $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}$ which can be expanded to $\left(\frac{\partial g}{\partial \alpha} \quad \frac{\partial g}{\partial \beta} \quad \frac{\partial g}{\partial \gamma} \right)$. This is equivalent to

$$\left(0 \quad \frac{1}{\gamma} \quad \frac{-\beta}{\gamma^2} \right).$$

As in the logistic case, $\mathbf{g}$ which is given by t is a nonlinear function of the model

parameters α, β and γ. Suppose the variance-covariance matrix of the estimates of the

three model parameters, could be obtained as $\begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} & \sigma_{\alpha\gamma} \\ \sigma_{\beta\alpha} & \sigma_\beta^2 & \sigma_{\beta\gamma} \\ \sigma_{\gamma\alpha} & \sigma_{\gamma\beta} & \sigma_\gamma^2 \end{pmatrix}$. Plugging in the partial

derivatives and the variance covariance matrix in (2.11) and simplifying, the exact

variance estimate of the age at MV for the Gompertz model could be shown to equal,

$$V(t) = \left(\frac{1}{\gamma}\left(\frac{\sigma_\beta^2}{\gamma} - \frac{\beta\sigma_{\gamma\beta}}{\gamma^2} \right) - \frac{\beta}{\gamma^2}\left(\frac{\sigma_{\beta\gamma}}{\gamma} - \frac{\beta\sigma_\gamma^2}{\gamma^2} \right) \right).$$

2.6. Estimating the Variance of Age at PHV When No Explicit Form for the Estimate Exists.

Since there are explicit solutions to the age at MV for the logistic and the Gompertz that are available, the corresponding estimation of the asymptotic variance is straight forward. However, when the solution to the age at MV or age at PHV itself requires numerical methods the corresponding estimation of the variances is not straight forward. This is mainly because the partial derivatives $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}$ are not readily available. In this section a method for estimating the age at PHV numerically, after obtaining the estimates of the model parameters (also numerically) and then estimating the derivatives numerically so that estimates of the variance of age at PHV could be obtained, is described.

Once the equation for the age at PHV is obtained, a nonlinear optimization procedures such as the PROC NLP in SAS, (which allows constrained optimization so that solutions in the reasonable range of ages say, between 0 to 18) could be used.
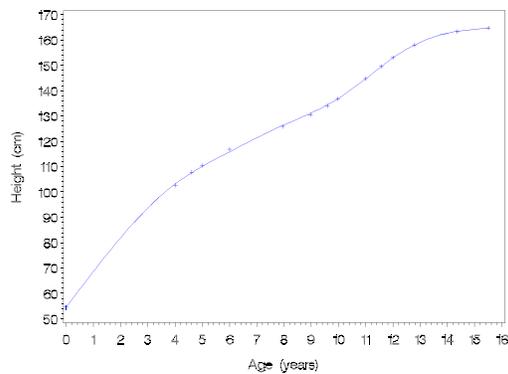
2.7. The Double Logistic Model

The double logistic model was first published by Bock et al.. in 1973 [Bock, Wainer,

Petersen, Thissen, Murray, and Roche (1973)]. Then it was used to analyze growth

models by Thissen et al. in 1976 [Thissen, Bock, Wainer, and Roche (1976)]. The model

is a summation of two simple logistic models. It describes human growth from the age of

one to adulthood. The model is,

$$y = f(t,\theta) = \left(\frac{D}{1+e^{-A_1(t-B_1)}} + \frac{f-D}{1+e^{-A_2(t-B_2)}} \right), \tag{2.12}$$

where $\theta = (D, f, A_1, B_1, A_2, B_2)$ are the model parameters. Here , y is the stature (cm);

t is age (year); D is the upper asymptote of the prepubertal component (year); A_1 shows

growth velocity for the prepubertal component (cm/year); B_1 reflects growth of the

prepubertal component (year); f is the observed mature stature (cm); $f - D$ is the

adolescent asymptote (cm); A_2 shows growth velocity for the adolescent component

(cm/year); B_2 reflects growth of the adolescent component (year).

Figure 3 below shows the double logistic model depicting the fit, the peak height velocity

and acceleration curves. The model as described above has biologically meaningful

parameters and encompasses a prepubertal and an adolescent component of the total

human growth. As in the previous growth models, the issue of starting values for

parameter estimates is essential. Here, a graphical approach is most useful to guess

starting values for parameter estimates as the growth model cannot be linearized. For this

study, an EXCEL spreadsheet was used to plot the data and manipulate the parameter

estimates to get good starting values for the double logistic model. One can tweak the

graph to further adjust the initial estimates to be used for the nonlinear procedure in SAS.
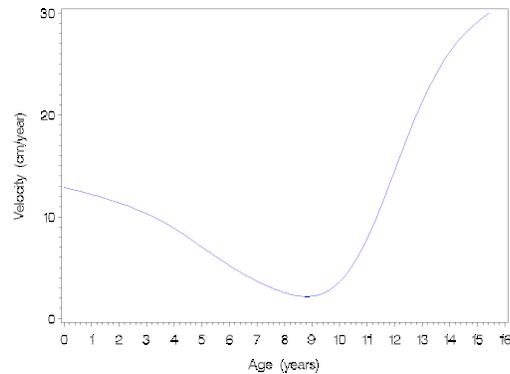
(a)   (b)
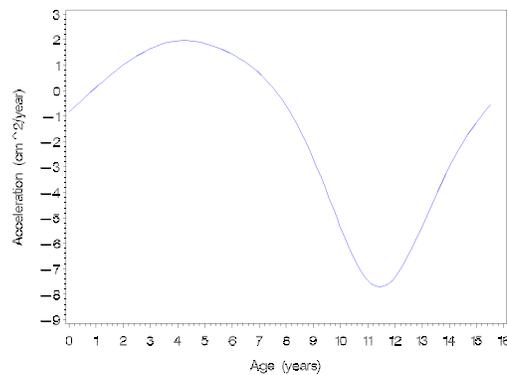
(c)

Figure 3 Double Logistic Model. (a) Growth curve. (b) Growth rate curve – the PHV

curve. (c) Acceleration curve.

To find the age at PHV, one has to resort to a numerical procedure as there is no closed form solution to the PHV equation. The age at PHV is obtained by differentiating twice and solving equation (2.12). The PHV equation for the double logistic is,

$$\frac{\partial y}{\partial t} = \frac{DA_1 e^{-A_1(t-B_1)}}{\left(1+e^{-A_1(t-B_1)}\right)^2} + \frac{(f-D)(A_2 e^{-A_2(t-B_2)})}{\left(1+e^{-A_2(t-B_2)}\right)^2}. \tag{2.13}$$

The derivative of the above equation used to solve for age at PHV is,

$$\frac{\partial^2 y}{\partial t^2} = \frac{DA_1^2 e^{-A_1(t-B_1)}(1+e^{-A_1(t-B_1)})+2DA_1^2 e^{-A_1(t-B_1)}}{\left(1+e^{-A_1(t-B_1)}\right)^3} +$$
$$\frac{A_2^2(f-D)e^{-A_2(t-B_2)}(1+e^{-A_2(t-B_2)})+2A_2^2(f-D)e^{-A_2(t-B_2)}}{\left(1+e^{-A_2(t-B_2)}\right)^3}. \tag{2.14}$$

However, the complexity of the second derivative is suggestive of the absence of explicit solution for age at PHV. As such, PROC NLP procedure in SAS based on Newton Raphson method would output an approximate age at PHV. Because of this the variance computations for double logistic will involve a numerical method.

2.8. The Triple Logistic Model

A modified version of the double logistic model called triple logistic model has been proposed by Bock and Thissen (1976). It is claimed that a much better fit of human height growth from birth to maturity can be modeled using triple logistic model [Lozy

(1978), Bock and Thissen (1976)]. Ever since the triple logistic has been successfully adopted to model growth data. The basic triple logistic model is,

$$y = f(t;\theta) = a_1 \left(\frac{1-D}{1+e^{-A_1(t-B_1)}} + \frac{D}{1+e^{-A_2(t-B_2)}} \right) + \frac{a_2}{1+e^{-A_3(t-B_3)}}$$

(2.15)

where $\theta = (D, a_1, A_1, B_1, A_2, B_2, a_2, A_3, B_3)$ is the function of parameters; D is contribution of the mid-childhood growth to the total growth in childhood growth. The parameters B_1, B_2, and B_3 are timing parameters which control the location of early-childhood, mid-childhood and adolescent component along the age axis. The parameters a_1 and a_2 decide the amount of growth contributed when an individual passes from pre-adolescent to adolescent stage respectively.

Commonly mentioned model parameters that help in identifying and quantifying growth landmarks are PHV and age at PHV. Also, considered growth indicators are minimum height velocity (MHV), age at early childhood MHV, pre-pubertal MHV, and age at MHV [Guo et al. (1992)].

There is the apparent advantage of a triple logistic model being able to capture the mid-growth spurt. The triple logistic has three additional parameters than the double logistic. Like the double logistic, the triple logistic model captures the two growth components – pre-pubertal and adolescent which are two distinct stages of human development. During pre-pubertal stage, some people show modest acceleration in growth velocity. But, in the adolescent stage, there is a considerable growth spurt or acceleration in growth velocity.

Corresponding growth peaks can be found by plotting individual acceleration curves against time. However, for the triple logistic model no closed form exists to solve the acceleration equation. This can be verified if one attempts to solve the age at PHV from following acceleration function,

$$\frac{\partial^2 y}{\partial t^2} = \frac{a_1(1-D)A_1^2 e^{(-A_1(t-B_1))}}{\left(1+e^{-A_1(t-B_1)}\right)^2} + \frac{2a_1(1-D)A_1^2 e^{(-2A_1(t-B_1))}}{\left(1+e^{-A_1(t-B_1)}\right)^3}$$
$$- \frac{a_1 D A_2^2 e^{(-A_2(t-B_2))}}{\left(1+e^{-A_2(t-B_2)}\right)^2} + \frac{2a_1 D A_2^2 e^{(-2A_2(t-B_2))}}{\left(1+e^{-A_2(t-B_2)}\right)^3} \qquad . \qquad (2.16)$$
$$- \frac{a_2 A_3^2 \exp(-A_3(t-B_3))}{\left(1+e^{-A_3(t-B_3)}\right)} + \frac{2a_2 A_3^2 \exp(-2A_3(t-B_3))}{\left(1+e^{-A_3(t-B_3)}\right)^2}$$

Recall that the equation above has been derived from the following PHV equation,

$$\frac{\partial y}{\partial t} = \frac{a_1(1-D)A_1 \exp(-A_1(t-B_1))}{\left(1+e^{-A_1(t-B_1)}\right)^2} + \frac{a_1 D A_2 \exp(-A_2(t-B_2))}{\left(1+e^{-A_2(t-B_2)}\right)^2} + \frac{a_2 A_3 \exp(-A_3(t-B_3))}{\left(1+e^{-A_3(t-B_3)}\right)^2} \qquad (2.17)$$

The complexity of the PHV equation (2.17) complicates model fitting and solving for age at PHV. We notice that t cannot be expressed explicitly in terms of all other model parameters. In such instances, PROC NLP is used to obtain t. The SAS procedure employs the Newton Raphson method to optimize the PHV equation.
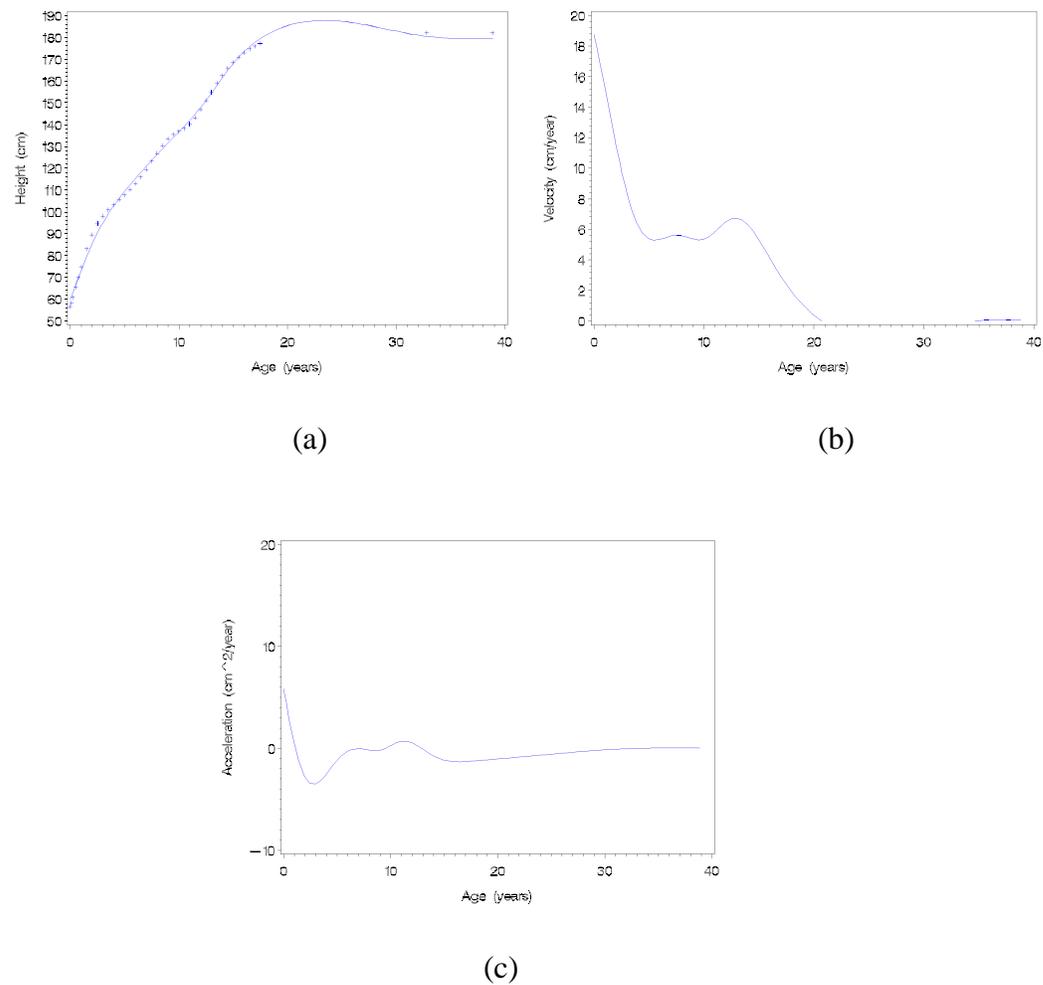
(a)

(b)

(c)

Figure 4 Triple Logistic Model. (a) Growth curve. (b) Growth rate curve – the PHV

curve. (c) Acceleration curve.

2.9. Proposed Variance Estimation Method

In this section, a method is proposed for estimating the variance of age at PHV when

there is no explicit solution for age at PHV as described in section 2.6. The proposed

method utilizes the delta method to find the variance of age at PHV but $\dfrac{\partial \mathbf{g}}{\partial \mathbf{\theta}}$ is

estimated numerically.

In calculus the derivative at any point corresponds to the slope of the tangent at that

point. Slope is constant for linear functions, but not for nonlinear functions. The slope of

a nonlinear function varies at different points along the curve.
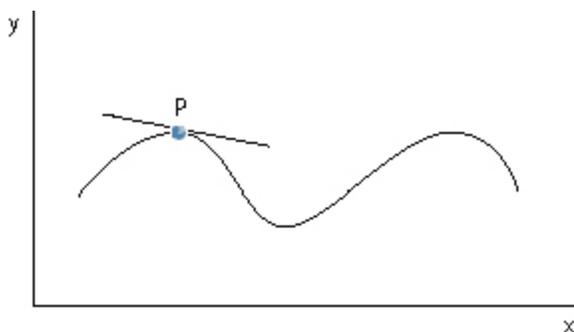
Figure 5 Slope of a nonlinear function at point P.

From figure 5, the slope of the function y at point p is defined as $\dfrac{\partial y}{\partial x}\bigg|x = x_p$. In this

thesis, the objective is to find the slope of the nonlinear relationship $\mathbf{g}$ with respect to

each of the parameter estimates in $\mathbf{\theta}$. The derivative of the age at PHV function, $\mathbf{g}$, with

respect to $\mathbf{\theta}$ is tackled numerically by computing the slope at $\hat{\mathbf{\theta}}$.

By definition, the slope of a scalar parameter function g or $\dfrac{\partial g}{\partial \theta}$ is,

$$\lim_{h \to 0} \frac{g(\hat{t}; \hat{\theta} + h) - g(\hat{t}; \hat{\theta})}{h}, \qquad (2.18)$$

where h is a very small number in the neighborhood of $\hat{\theta}$ and $\hat{t}$ is the estimated age at

PHV. (A convenient h in the order of 10^{-3} seems to work for the different models in

this study.) Here, the choice of h is crucial. The rule of thumb is to choose h where the

graph of $g(\hat{t}; \hat{\theta})$ is nearly linear between $\hat{\theta}$ and $\hat{\theta} + h$. Subsequently, the slope is

calculated by plugging in the estimates of t for $g(\hat{t}; \hat{\theta})$ and $g(\hat{t}; \hat{\theta} + h)$ respectively.

First, the age at PHV, $\hat{t}$, is obtained numerically by solving the equation $g(\hat{t}, \hat{\theta}) = 0$. This

is available in most statistical software packages. For example, PROC NLP in SAS could

be used.

Similarly, for a p dimensional multiparameter function $\mathbf{g}$, $\mathbf{g}(\hat{t}; \hat{\boldsymbol{\theta}})$ and $\mathbf{g}(\hat{t}; \hat{\boldsymbol{\theta}} + h)$ are

evaluated with respect to each parameter estimate. For instance, if $\hat{\boldsymbol{\theta}}$ is a $p \times 1$ vector of

parameter estimates, then $\mathbf{g}(\hat{t}; \hat{\boldsymbol{\theta}} + h)$ is evaluated p times. This could be achieved by

using the SAS subroutine PROC NLP to estimate the age at PHV at each element of $\hat{\boldsymbol{\theta}}$

and $\hat{\boldsymbol{\theta}} + h$. After the procedure is applied for each parameter estimate, the individual

slopes are computed using expression (2.18) by plugging in $\mathbf{g}(\hat{t}; \hat{\boldsymbol{\theta}})$ and $\mathbf{g}(\hat{t}; \hat{\boldsymbol{\theta}} + h)$. This

yields the numerical estimate of the $p \times 1$ vector $\dfrac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}$. Substituting this in expression

(2.11) will then yield the estimate for the variance of age at PHV.

In the next chapter (3), the proposed method of estimating the variance for age at PHV

is applied to the logistic and Gompertz models for which explicit solutions are available,

in order to compare its accuracy. Also, in the next chapter, estimates of the variance are

included as weights in models where age at PHV is used as a dependent variable to

illustrate the usefulness of estimating this variance.

3   Properties and Use of Numerical Estimates of Variance of Age at PHV

In the previous chapter a numerical method for estimating the variance for age at PHV was presented. The purpose here is to verify whether the proposed method of estimating variance yields accurate estimates of the variance. For this purpose, a subset of the Fels data was used. First, details of the fit of the logistic, Gompertz and triple logistic are described. For these growth models the age at MV and age at PHV are estimated and the corresponding variance of the age at MV or age PHV are estimated using the exact method and the numerical method. Then a comparison of the two methods of estimating the variance is made. These are summarized in tables 1 and 2, respectively. Second, for the estimates of age at PHV and the corresponding variance obtained from the triple logistic model are used in a data analysis of adulthood measurements of the individuals as outcome in which the age at PHV is used as a predictor.

The subjects in the Fels data were 1923 children (930 boys and 993 girls) who were born between 1928 and 2007. A subset of this data in which information on the variables height, weight, gender, adulthood glucose level and adulthood insulin level are available was obtained. Further, in order to demonstrate the method proposed in the thesis, only those individuals whose growth models could be fitted well were used. That is, only the individuals who were followed with sufficient height measurements from birth to maturity age of 18 years were included. Ultimately, the number of eligible subjects under these conditions turned out to be 632, of which 313 were males and 319 were females.

As with many longitudinal studies, there were subjects who were lost to follow-up and withdrew from the study for different reasons contributing to missing or unavailable data for analysis. Out of the 632 subjects, further selection of valid subjects with enough data depended on the convergence criteria of the nonlinear procedures employed for the different growth models fitted in the thesis. Therefore, the four growth models presented in this work produced varying number of eligible subjects for analysis. Thus, number of subjects that converged under logistic was 472 and under Gompertz was 528. However, for the more complicated double and triple logistic growth models, which require more observations in the first two years of growth, converged only for 75 and 125 cases, respectively.

For a preliminary investigation of fits, the simple logistic and Gompertz models were considered. The PROC NLIN in SAS was used to obtain the estimates of the model parameters ($\hat{\boldsymbol{\theta}}$). The initial estimates for the parameters were obtained as described in chapter 2. In addition to the exact estimates, The PROC NLP as shown below in Program 1, was used to obtain the estimates of the age at PHV. In the program, f_2 is equation (2.3) which is maximized to obtain the age at PHV.

The NLP is a popular technique to find the minimum or maximum of a continuous nonlinear function. The algorithm allows constraints for the solutions. For example, in the estimation of age at PHV, one could constrain the estimates to be within 0 and 18.

The NLP has several optimizers that could be chosen to suit a particular type of

optimization problem. To name few, the following are a class of optimization techniques

based on iterative process: Quadratic Active Set Technique, Trust-Region Method,

Newton-Raphson Method with Line Search, Newton-Raphson Method with Ridging,

Quasi-Newton Methods, etc. In particular, Newton-Raphson algorithm is used to find the

roots that maximize the PHV equation in this work [SAS 9.2]. The NLP is not to be

confused with the iterative procedure used to estimate parameters of the nonlinear model

using PROC NLIN. The NLIN procedure gives the best fit parameter estimates which in

turn are inserted in the NLP procedure to find the age at PHV.

```
***The data step sets the initial value for the age at PHV as 11.***;
data par2(type=est);
       keep _type_ t2;
       _type_='parms'; t2 = 11;

output;
run;

***The data deriv contains all parameter estimates and age at PHV, t2.
***The data par2 contains the intial parameter estimate for t2.***;


proc nlp data=deriv inest=par2 outest=outnlp2; by ptno;
max f2;
parms t2;
f2 = (k*(a)*b1*exp(-b1*t2))/(1+(a)*exp(-b1*t2))**2;
run;
```

Program 1. Estimating the age at PHV

The Program 2 below is the same SAS code as in Program 1 with $h = 0.0001$ added to

the parameter estimate a in the logistic model, and the estimate of t is recalculated.

Similarly, the same constant is added to the other two parameters and the corresponding estimates of *t* are obtained.

```
data par2(type=est);
      keep _type_ t2;
      _type_='parms'; t2 = 11;

output;
run;

proc nlp data=deriv inest=par2 outest=outnlp2; by ptno;
max f2;
parms t2;
f2 = (k*(a+0.0001)*b1*exp(-b1*t2))/(1+(a+0.0001)*exp(-b1*t2))**2;
run;
```

Program 2. Estimating the age at PHV

The Program 2 above, shows the code for estimating the age at PHV from the estimates after perturbing the model parameters by the constant *h*.

From the fitted plots (Figure 1(a) and Figure 2(a)), it could be noted that the logistic and Gompertz models do not fit the height data well. The fitted curve for both logistic and Gompertz is especially poor around the peak age where the horizontal asymptote lies. The age at maximum velocity is reached rather rapidly in both logistic and Gompertz models as mentioned in Marubini et al. (1971). The mean age at MV is 3.537 (SD = 1.635) for simple logistic and 1.046 (SD = 1.27) for Gompertz. Out of 472 subjects selected for the logistic, an additional 26 subjects were eliminated because of relatively large variance (of the order of 10) of age at MV. These observations were not used for the comparisons, resulting in a sample of 446 subjects. Figure 3.3 below summarizes the distribution of estimate of age at MV over the 446 subjects for the logistic model.
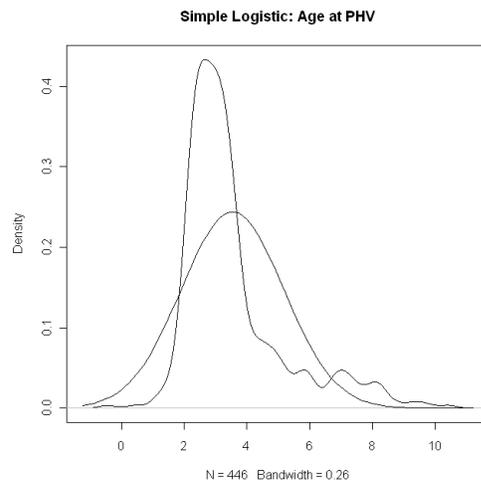
Figure 3.3 Distribution plot for Age at MV for Simple Logistic.

The best fitting normal distribution is also shown in Figure 3.3. The distribution of the

age at MV for the 446 subjects under the logistic model seems non-normal. Table 1

below shows the average age at MV, average variance computed using the exact method

as well as the numerical method proposed in the thesis. Clearly, the numerical method

has produced quite accurate estimates of the variance of age at MV.

Table 1: Means of Age at MV, Estimated and Exact Variances for Simple Logistic

	Simple Logistic		
Variables	N	Mean	SD
t (Age at MV)	446	3.537	1.635
Exact Variance of age at MV	446	0.430	1.175
Computed Variance of age at MV	446	0.431	1.176

The figure 3.4. shows the plot of the variances computed from the exact method (x-axis) versus the numerical method (y-axis) for the various subjects separated by gender. Clearly, the correlation is close to perfect. Therefore, the numerical method seems to work dramatically well.

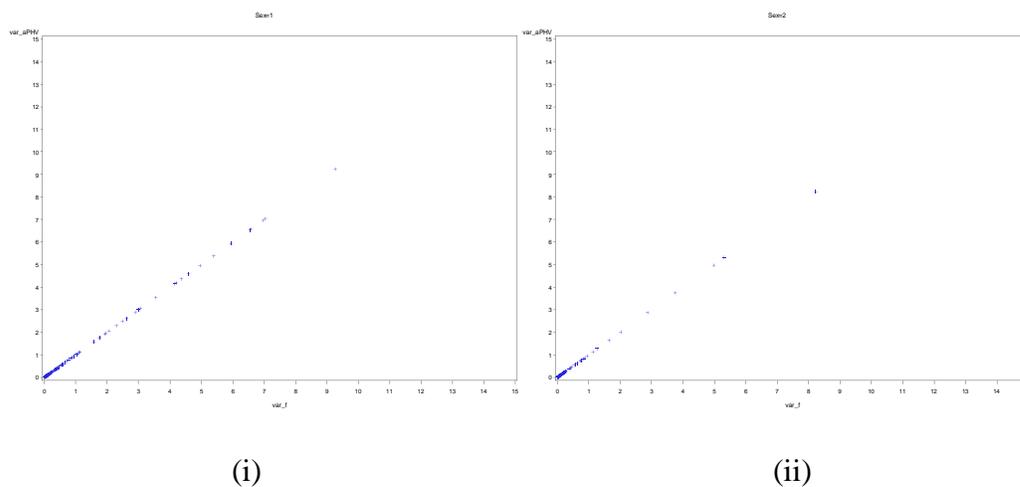(i)   (ii)

Figure 3.4 Simple Logistic: Variance Plots. (i) Male (ii) Female

Figures 3.5 and 3.6 and table 2 are shown below for the Gompertz model, and show similar results.
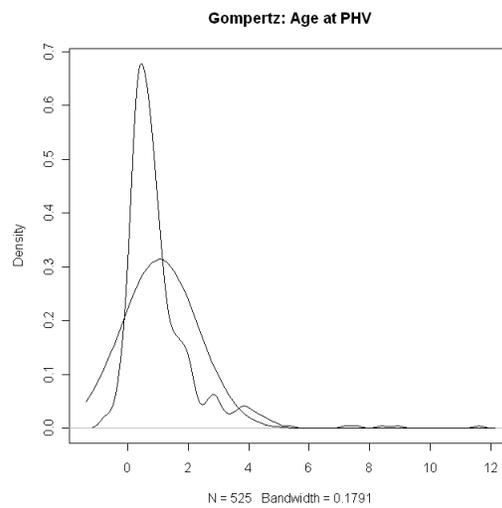
Figure 3.5 Distribution plots for Age at MV for Gompertz.

Table 2: Means of Age at MV, Estimated and Exact Variances for Gompertz

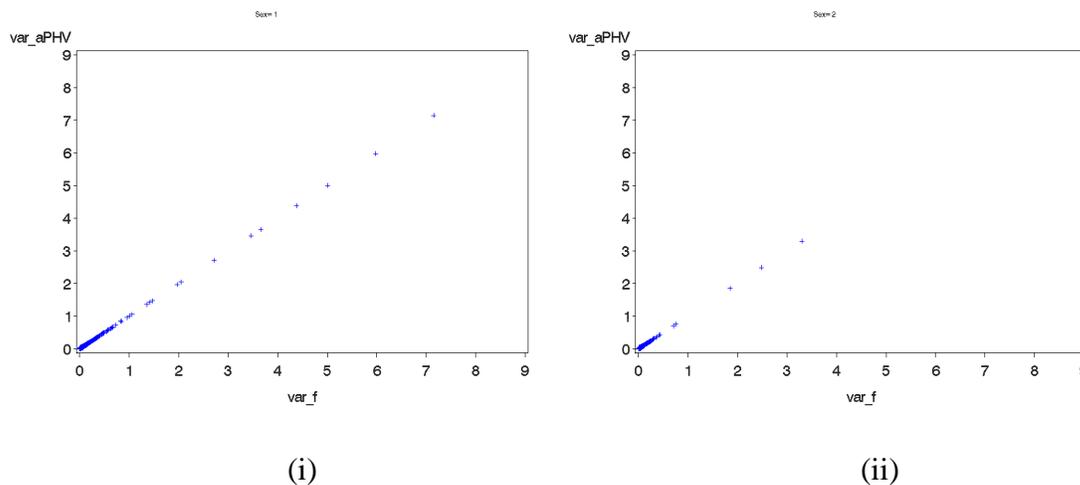|	Gompertz		
Variables	N	Mean	SD
T (Age at MV)	525	1.046	1.270
Exact Variance of age at MV	525	0.195	0.612
Calculated Variance of age at MV	525	0.195	0.612

Figure 3.6 Gompertz: Variance Plots. (i) Male (ii) Female

Having established the accuracy of the numerical methods through the comparisons made

for the logistic and Gompertz models next the numerical method is applied for the triple

logistic model. Recall that the logistic and the Gompertz models do not fit the growth

data well at all. The double and triple logistic models fit the data for the patients

appropriately, as previously shown by Bock and Thissen (1976) and Lozy (1978).

Because of the inherent human growth characteristics, the growth trajectories cannot be

easily explained using the simple logistic or Gompertz models. The triple logistic curve

captures the ages at early-childhood, mid-childhood and adolescent stages.

Mechanistically, the double and triple logistic curves display points of most interest such

as the maximum or minimum rates of growth associated with the pre-pubertal and

adolescent ages that appear naturally if a person sustains a normal growth. Furthermore,

the fit of the triple logistic for the height and velocity curve is often quite good but the

model lacks appropriate biological interpretation of the parameters, which is claimed

as a disadvantage of triple logistic by Lozy (1978) and Preece and Heinrich (1981).

The figure 3.7 and the Table 3 below show the results for the Triple logistic model. The

age at PHV for triple logistic is achieved at a comparatively reasonable age. Table 3 and

Figure 3.7 sum up the distribution of age at PHV for 54 subjects for whom the procedures

converged. Notice that the final number of subjects is a further reduction from the 125

eligible ones by including only subjects with age at PHV between 10 and 18. Due to lack

of sufficient data and initial age set for the NLP procedure, some subjects did not seem to

have their age at PHV fall in a reasonable range and therefore they were excluded. It is

also important to notice that the distribution of the age at PHV under the triple logistic

regression is closer to the normal distribution than the logistic or the Gompertz estimates.
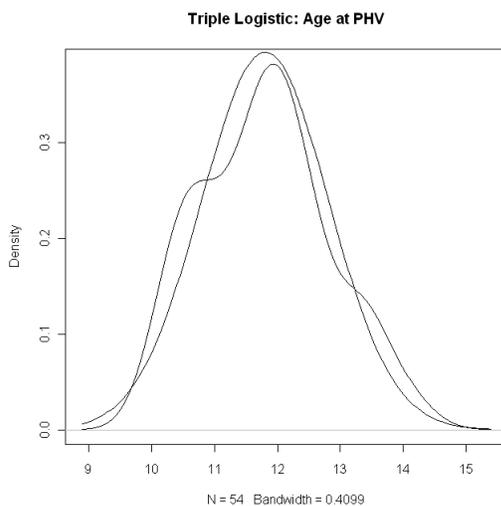

Figure 3.7 Distribution plots for Age at PHV for Triple logistic.

Table 3: Means of Age at PHV, Estimated and Exact Variances for Triple Logistic

Triple Logistic

Variables	N	Mean	SD
Age at PHV	54	11.804	1.011
Variance of age at PHV	54	1.111	5.194

Program 3 below is the sample SAS code used to estimate the age at PHV using PROC NLP for triple logistic where f is the objective function to be maximized. For triple logistic case, f is given by equation (2.17). The initial age at PHV is set at 11 years and the final output is stored in a dataset called outnlp.

```
***The data step sets the initial value for the age at PHV as 11.***;
data par(type=est);
       keep _type_ t;
       _type_='parms'; t = 11;

output;
run;


** PROC NLP;
proc nlp data=new_t inest=par outest=outnlp; by ptno;
max f;
parms t;
f =((a1*(1-g2)*H2)*(EXP(-H2*(age-b1))))/((1+EXP(-H2*(age-b1)))**2) +
((a1*g2*I2)*(EXP(-I2*(age-b2))))/((1+EXP(-I2*(age-b2)))**2)
                + ((a2*J2)*(EXP(-J2*(age-b3))))/((1+EXP(-J2*(age-
b3)))**2);

run;
```

Program 3. Estimating the age at PHV

As part of the last addendum to the thesis, the usefulness of estimating the variance of age at PHV is illustrated by applying the variance obtained from the triple logistic model as weights in the analysis of adulthood measurements.

In general, ordinary least squares (OLS) regression assumes that the variance of the residual errors is constant for all values of independent variables. Sometimes, as is the case when the age at PHV is estimated, it may happen that the independent variables have varying error variances. This could lead to unstable estimates of regression coefficients. These unstable estimates can be identified with small or large standard errors. As a result the power of significant tests may be reduced. Therefore, a weighted least square (WLS) regression should be considered to take care of the violation of constant error variances. The weighting scheme works as follows. Each data point is given proper amount of influence on the final parameter estimates. Observations whose value on the dependent variable correspond to large variances on the independent variables count less. Those with small variances count more in estimating the regression coefficients.

In the Fels data, applying the WLS regression to the final analysis data, observations whose value of glucose and insulin level (dependent variables) that correspond to large variances of the estimate of the age at PHV (independent variable) would count less. Apart from correcting for homoscedasticity, WLS regression adjusts fit by giving less weight to outliers and observations thought to be unreliable [Carroll & Ruppert (1988)].

Weighted regression is set up and implemented in SAS as follows. The 'weight' option is given by the standard error of age at PHV.

```
*** Weighted Regression***;
proc genmod data=time12; where sex=1; weight std_err;
model BCinsulin = t new_stature ANweight;
run;
```

The above weighted regression procedure models insulin level for males and takes the age at PHV, height and weight measurements as predictors. Accordingly, weighted and ordinary regression is applied to model glucose and insulin levels for each sex. In the sections that follow all the models are described and their summaries are presented in tables 4 to 12.

3.1. Results of the Statistical Analysis

In this section analysis using GLM was utilized to identify significant variables predictive of glucose level and insulin level for the triple logistic. It is of interest to measure levels of glucose and insulin in human beings as they could detect different types of conditions and disease if a person appears to have abnormal levels of glucose or insulin in the blood. Clinically, the normal levels of blood sugar is 70-90 mg/dl and the ideal insulin level should be less than 10 IU/ml. Anything over 10 IU/ml indicates that a person is eating too much food that are stimulating excess insulin release from the pancreas. The analysis was performed using multiple regression and weighted regression analysis where the standard errors are considered as weights. As described in section 2.6, the age at PHV

obtained from the triple logistic model is of particular interest for our regression model. Looking at Figure 3.5, the regression assumption of normal distribution of variables holds for the distribution of age at PHV obtained from triple logistic model.

The response variables are glucose level and insulin level. The predictors are age at PHV, stature and weight. There were a total of 54 observations. These observations were chosen so that all their glucose and insulin measurements were complete and there was no missing age at PHV which is between 10 and 18. Pertinent anthropometric measurements of the dependent and independent variables of the sample of 54 subjects are summarized in Table 4. The mean glucose and insulin levels for males in this sample were 86.43 and 8.41 units with standard deviations of 12.37 and 5.06, respectively. At the time of adulthood, the mean age at PHV of participants for males and females is 12.08 years (SD = 1.052) and 11.55 years (SD = 0.92), respectively. On average, the stature measurements for both males and females are 178.56 centimeters (SD = 8.01) and 166.71 centimeters (SD = 5.58). The mean weight measurements at adulthood for this samples are 81.60 kilograms (SD = 16.15) for males and 69.10 kilograms (SD = 18.80) for females. Analysis using GLM for both sexes is done separately. First, the simple regression analysis for males and females respectively is considered. Second, the weighted regression scheme is considered.

3.2. Regression Analysis For Males with Glucose and Insulin Levels as Response

Simple linear regression model (see Table 5 and Table 6) indicated that adulthood stature ($p = 0.0284$) was significant predictor of insulin level. But, age at PHV ($p = 0.6494$) and weight ($p = 0.0808$) were not significant predictors of insulin level. On the other hand, adulthood weight ($p = 0.0279$) was found to be significant predictor of glucose level while age at PHV ($p = 0.2046$) and adulthood height ($p = 0.1610$) were not significant univariate predictors of glucose level. In general, significant increases in insulin level were associated with decreases in adulthood stature. A 10 centimeters increase in stature was associated with 2.16 units decrease in total insulin level in males. Furthermore, significant increases in glucose level were associated with increases in adulthood weight. A 10 kilogram increase in weight was associated with 3.95 units increase in total glucose level in males.

3.3. Regression Analysis For Females with Glucose and Insulin Levels as Response

Simple linear regression model (see Table 7 and Table 8) indicated that none of the regressors was significant predictors of insulin level. On the other hand, adulthood weight ($p = 0.0434$) was found to be significant predictor of glucose level. Moreover, significant increases in glucose level were associated with increases in adulthood weight. A 10 kilogram increase in weight was associated with 1.94 units increase in total glucose level in females.

3.4. Weighted Regression For Males with Glucose and Insulin Levels as Response

In the weighted least squares regression of this problem, the weights used are the inverse of estimated standard error for age at PHV. Weighted least squares regression model (see Table 9 and Table 10) indicated that adulthood stature (p = 0.0481) was a significant predictor of insulin level. But, age at PHV (p = 0.9095) and weight (p = 0.5231) were not significant predictors of insulin level. On the other hand, adulthood stature (p = 0.0010) and weight (p = 0.0030) were significant predictors of glucose level while age at PHV (p = 0.5351) was not significant univariate predictor of glucose level. In general, significant decreases in insulin level were associated with increases in adulthood stature. A 10 centimeters increase in stature was associated with 1.66 units decrease in total insulin level in males. Also, significant decreases in glucose level were associated with increases in adulthood stature. A 10 centimeters increase in stature was associated with 12.04 units decrease in total insulin level. Furthermore, significant increases in glucose level were associated with increases in adulthood weight. A 10 kilogram increase in weight was associated with 5.77 units increase in total glucose level in males.

3.5. Weighted Regression For Females with Glucose and Insulin Levels as Response

Weighted least squares regression model (see Table 11 and Table 12) indicated that age at PHV (p = 0.0218) was significant predictor of insulin level. But, stature (p = 0.3264)

and weight (p = 0.1378) were not significant predictors of insulin level. On the other

hand, weight (p = 0.0172) was a significant predictor of glucose level while age at PHV

(p = 0.7144) and stature (p = 0.9424) were not significant univariate predictors of glucose

level. In general, significant increases in insulin level were associated with decreases in

age at PHV in females. A 1 year increase in age at PHV was associated with 1.53 units

decrease in total insulin level in females. Furthermore, significant increases in glucose

level were associated with increases in adulthood weight. A 10 kilogram increase in

weight was associated with 2.09 units increase in total glucose level in females.

Table 4: Means of Predictors and Responses

(Continuous) Variables	N	Mean	SD
t (age at PHV)	54	11.803	1.01
Stature	54	172.412	9.04
Weight	54	75.1194	18.52
Glucose	54	86.43	12.37
Insulin	54	8.41	5.06

Table 5: Means of Predictors of Insulin: Males

(Continuous) Variables	p-value	Slope	(95% CI)
t (age at PHV)	p = 0.6494	-0.2976	(-1.581, 0.986)
Stature †	p = 0.0284	-0.2161	(-0.410, -0.023)

	p-value	Slope	(95% CI)
Weight	p = 0.0808	0.0817	(-0.010, 0.173)

CI = confidence interval, † = statistically significant

Table 6: Means of Predictors of Glucose: Males

(Continuous) Variables	p-value	Slope	(95% CI)
t (age at PHV)	p = 0.2046	-3.1931	(-8.126, 1.740)
Stature	p = 0.1610	-0.5313	(-1.274, 0.212)
Weight †	p = 0.0279	0.3954	(0.043, 0.748)

CI = confidence interval, † = statistically significant

Table 7: Means of Predictors of Insulin: Females

(Continuous) Variables	p-value	Slope	(95% CI)
t (age at PHV)	p = 0.6946	0.4703	(-1.878, 2.818)
Stature	p = 0.4523	0.1536	(-0.247, 0.554)
Weight	p = 0.1013	0.1008	(-0.020, 0.221)

CI = confidence interval, † = statistically significant

Table 8: Means of Predictors of Glucose: Females

(Continuous) Variables	p-value	Slope	(95% CI)
t (age at PHV)	p = 0.5023	-1.2568	(-4.929, 2.415)
Stature	p = 0.7266	0.1118	(-0.515, 0.738)
Weight	p = 0.0434	0.1943	(0.0057, 0.383)

CI = confidence interval, † = statistically significant

Table 9: Weighted Regression: Means of Predictors of Insulin: Males

(Continuous) Variables	p-value	Slope	(95% CI)
t (age at PHV)	p = 0.9095	-0.0587	(-1.070, 0.953)
Stature †	p = 0.0481	-0.1660	(-0.331, -0.0014)
Weight	p = 0.5231	0.05	(-0.059, 0.116)

CI = confidence interval, † = statistically significant

Table 10: Weighted Regression: Means of Predictors of Glucose: Males

(Continuous) Variables	p-value	Slope	(95% CI)
t (age at PHV)	p = 0.5351	-1.3933	(-5.796, 3.010)
Stature †	p = 0.0010	-1.2040	(-1.921, -0.488)
Weight †	p = 0.0030	0.5768	(0.196, 0.957)

CI = confidence interval, † = statistically significant

Table 11: Weighted Regression: Means of Predictors of Insulin: Females

(Continuous) Variables	p-value	Slope	(95% CI)
t (age at PHV) †	p = 0.0218	-1.5300	(-2.837, -0.223)
Stature	p = 0.3264	-0.1844	(-0.553, 0.184)
Weight	p = 0.1378	0.1150	(-0.037, 0.267)

CI = confidence interval, † = statistically significant

Table 12: Weighted Regression: Means of Predictors of Glucose: Females

(Continuous) Variables	p-value	Slope	(95% CI)
t (age at PHV)	P =	-0.2756	
	0.7144		(-1.751, 1.200)
Stature	P =	-0.0153	
	0.9424		(-0.431, 0.401)
Weight †	P =	0.2085	
	0.0172		(0.037, 0.380)

CI = confidence interval, † = statistically significant

3.6. Weighted Regression with Sex, Age at PHV and BMI as Predictors

It is suggested that body mass index (BMI; in kg/cm^2) is used to predict adult

overweight or obesity [Guo, Wu, Chumlea and Roche (2002)]. Therefore, the last

weighted regression analysis includes sex as categorical, age at PHV and BMI as

predictors to model adulthood levels of glucose and insulin. In the analysis, adulthood

BMI (p = 0.0127) was found to be a significant predictor of glucose level. On the other

hand, age at PHV (p = 0.1533) and sex (p = 0.1084) were not significant predictors.

Similarly, adulthood BMI (p = 0.0271) was found to be a significant predictor of insulin

level while age at PHV (p = 0.8964) and sex (p = 0.5987) were not significant. For both

adulthood outcomes, BMI is significant predictor which is the case in real life as higher

BMI level is associated with overweight and hence higher levels of glucose and insulin.

In the final chapter of the thesis, some time is spent summarizing the idea communicated

through the work and opens doors for amendments in the future.

4   Conclusions and Future Research

In Chapter 1 we introduced nonlinear regression and specifically discussed about the common growth models, namely simple logistic, Gompertz, double logistic and triple logistic. Uncertainties introduced when measuring certain characteristics such as height or weight ought to be taken seriously. Chapter 1 highlighted the importance of not ignoring vagueness introduced due to human or mechanical errors in statistical analysis. The nature of scientific research is that some variables of interest inevitably introduce measurement error. As such, uncertainties introduced in the estimation of age at PHV when longitudinal measurements are modeled against height could be captured by the variance.

Chapter 2 focused on estimation of age at MV and age at PHV along with their corresponding variances for different growth models. For the simple logistic and Gompertz, the variances were computed using the delta method. But, computing the variance of age at PHV is not always straightforward for any given growth model. Typically, the double and triple logistic do not have closed form solutions for age at PHV and hence a new method for estimating the variance numerically was proposed.

The key part of our proposed method centered on estimation of the vector of derivatives of the function of age at PHV with respect to each parameter numerically. This facilitated in producing variance in the triple logistic models. Having obtained the

variances, usefulness of it was established using a weighted regression modeling of adulthood outcomes. It used the standard errors (square root of the variance produced by the proposed method) for age at PHV as weights to model glucose and insulin levels. Finally, results from weighted and ordinary regression analysis were compared shown that the inclusion of variance could alter the results.

4.1. Limitations and Suggestions

In this section we touch on the limitations of our proposed method and suggest ways to improve it in the future. Further, we recommend how the demonstrated model can be analyzed by considering interactions; multivariate analysis by treating the two response variables as a response vector.

Most frequently, to construct confidence intervals for nonlinear functions of the parameters we look into the standard errors produced by the asymptotic results using the delta method. Recall the asymptotic result in equation (2.11). In short, the asymptotic result states that for a given nonlinear function of parameters, $g(t,\theta)$, the variance of $g(t,\theta)$ is estimated using first order Taylor series expansion. From theory, it is well known that for a consistent estimator $\hat{\theta}$ of θ, $E(\hat{\theta}) = \theta$ and $V(\hat{\theta}) = \Sigma$. Now let the function $f = g(\hat{t}, \hat{\theta})$ be differentiable at every point $\hat{\theta}$ that defines it. Then, Taylor series at θ (close neighbor to $\hat{\theta}$) is

$$f = g(\hat{t},\hat{\theta}) = g(\theta) + \frac{\partial g(\hat{t},\hat{\theta}-\theta)}{\partial \hat{\theta}}|\hat{\theta} = \theta \quad + \quad \ldots \tag{3.1}$$

For large n or as $n \to \infty$, $E\left(g\left(\hat{\theta}\right)\right) = g\left(\theta\right)$ and $V\left(g\left(\hat{\theta}\right)\right) = g'\Sigma g$ where

$g' = \frac{\partial g}{\partial \hat{\theta}}|\hat{\theta} = \theta$. Thus, inference about $f = g(\hat{t},\hat{\theta})$ can be tackled using the delta method

approximation. The delta method can be useful if estimation of the second term in the

Taylor series can be easily computed.

For some nonlinear functions $g(\hat{t},\hat{\theta})$, it is not always possible to express an estimator of

interest (such as age at PHV, $\hat{t}$, in this case) in terms of all the model parameters. In

such instances, one has to approximate the term $g' = \frac{\partial g}{\partial \hat{\theta}}|\hat{\theta} = \theta$ to estimate variance

using the delta method.

In section 3.1, the proposed method showed how one can use the slope to approximate

the derivatives. The slope is determined by $\frac{g(\hat{t},\hat{\theta}+h) - g(\hat{t},\hat{\theta})}{h}$ for a small value h.

Essentially, the derivative is the slope of the tangent line joining point estimates

$\left(\hat{\theta}, g(\hat{t},\hat{\theta})\right)$ and $\left(\hat{\theta}+h, g(\hat{t},\hat{\theta}+h)\right)$. The new method works efficiently since the first

order Taylor series around $\hat{\theta}$ is linear over the range of values likely to be encountered,

hence meeting one of the delta method assumptions.

First, the parameters were estimated using the Marquardt nonlinear procedure (based on maximum likelihood) in SAS, then age at PHV, $\hat{t}$, is solved from $g(t,\hat{\theta}) = 0$ and $g(t,\hat{\theta} + h) = 0$ in the NLP procedure in SAS. The NLP procedure picks the t that maximizes the PHV function $g(t,\hat{\theta})$ or $g(t,\hat{\theta} + h)$. The method achieves good estimates of $\dfrac{\partial g}{\partial \theta}$ and appears to work well in estimation of the variance for $g(\hat{t},\hat{\theta})$. There are, however, issues that could lead to undesired variance estimates.

If a model is not specified well or has insufficient data, then the linearization process employed in the nonlinear procedure to come up with parameter estimates could fail to converge to a good estimate. In turn, inadequacy of the linear approximation can be questioned and hence questionable inference about the statistic of interest. In addition, the optimization procedure used to obtain $\hat{t}$ would not be guaranteed to give the global maximum $g(\hat{t},\hat{\theta})$. Instead, the optimization techniques can report $\hat{t}'s$ that produce only local maximum values of $g(\hat{t},\hat{\theta})$. This would lead to unsatisfactory variance estimates. Thus, asymptotic results may not apply [Li & Maddala (1999)].

To overcome the above stated inferential problems, a resampling technique known as *bootstrap* methods can be implemented [Myers et al.. (2002)]. Bootstrap distribution of age at PHV can be generated and used to make inferences about $g(\hat{t},\hat{\theta})$ by approximating its standard errors and confidence intervals. The bootstrap quantities are then compared to the asymptotic results. An agreement between bootstrap and

asymptotic estimates justifies validity of the inference results. Besides inferential issues, modeling issues that arose in this thesis are discussed.

After obtaining the standard errors for age at PHV, they were used as weights to model levels of glucose and insulin. The model that was fitted to the Fels data for each sex is assumed additive. That is, the effect of sex is the same for all values of age at PHV, stature and weight and vice versa. However, additivity may not be appropriate all the time. In such case, interactions should be considered. Therefore, instead of doing the analysis for separate sexes, the independent variable sex could enter the GLM model as a categorical variable. Then the interaction between age at PHV and sex can be looked at to see if there are significant differences between males and females in insulin or glucose level at each age at PHV values. Basically, interaction would enable us to formally test whether the insulin or glucose level variation in relation to age at PHV depends on sex. Finally, it appears that multivariate analysis can be considered given the two simultaneous response variables: glucose and insulin levels.

Let y_1 and y_2 be the glucose and insulin level responses. Putting them in a multivariate form, we create a vector of responses, Y where $Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Also, stacking the regressor variables in appropriate format, it is possible to analyze the data using multivariate analysis.

Throughout the thesis the assumption of independence of data holds. Realistically speaking, longitudinal data could be correlated as a result of taking repeated measurements on subjects. If this correlation happens to be significant and is not acknowledged in the analysis, the variance estimates could be underestimated. Therefore, if there is significant dependence among observations per subject in a longitudinal data, one should consider a correlation structure between observations from the same subject. In this thesis, an autoregressive (1) or AR(1) can be considered as a reasonable correlation structure since the correlation between measurements that are close in time is greater than the correlation of measurements that are widely separated in time.

In conclusion, the new method implemented in the thesis produces satisfactory results in common growth models. Computation of variances for age at PHV using the delta method has not been mentioned much in the literature (Gou et al. (1992), Demerath et al. (2004)) and the compendium of results in this work would be a great addition for further research.

.

Guo, S., Siergovel, R.M., Roche, A. F., and Chumlea, W.C.(1992). Mathematical Modelling of Human Growth: A Comparative Study. *American Journal of Human Biology* 4, 93-104.

Kingsland, S. (1982). The Refractory Model: The Logistic Curve And The History Of Population Ecology. *The Quarterly Review of Biology* 57(1), 29-52.

Li, H., and Maddala, G. S. (1999). Generalized Linear Models. *The Review of Economics and Statistics* 81 (4), 728-733, pp. 63-92.

Lozy, M. E. (1978). A Critical Analysis of the Double and Triple Logistic Growth Curves. *Annals of Human Biology* 5(4), 389-394.

Marubini, E., Resele, L.F., Barghini, G. (1971). A comparative fitting of the Gompertz and logistic functions to longitudinal height data during adolescence in girls. *Human Biology* 43, 237-252.

Myers, R.H., Montgomery, D.C., and Vining, G.G. (2002). *Generalized Linear Models With Applications in Engineering and the Sciences*, NY: John Wiley and Sons, Inc.

Preece, M. A., Heinrich, I. (1981). Mathematical Modeling Of Individual Growth Curves. *British Medical Bulletin* 37(3), 247-252.

Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling: a unified practical approach*, New York: Marcel Dekker.

Roche, A.F. (1992). *Growth, Maturation and Body Composition: The Fels Longitudinal Study 1929-1991*, Cambridge: Cambridge University Press.

SAS 9.2, (2008). SAS Institute Inc., Cary, North Carolina, USA.

Seber, G.A.F., and Wild, C.J. (2003). *Nonlinear Regression*, New Jersey: John Wiley and Sons, Inc.

Guo, S. S., Wu, W., Chumlea, W. C., and Roche, A. F. (2002). Predicting Overweight and obesity in adulthood from body mass index values in childhood and adolescence. *American Journal of Clinical Nutrition* 76 (3), 653-658.

Thissen, D., Bock, R.D., Wainer, H., and Roche, A. (1976). Individual Growth in Stature: A Comparison of Four Growth Studies in the U.S.A.. *Annals of Human Biology* 3 (6), 529-542.