Constructing internally consistent BMI Z-scores for adults and children to examine intra-household health effects

Felix Naschold
Department of Economics
University of Wyoming, Laramie, WY 82071
fnaschol@uwyo.edu

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

This work was supported by NHLBI with NIGMS at the National Institutes of Health with grant no. R01 HL126666-01: "*Growing Resilience* on the Wind River Indian Reservation – an RCT to measure the impact of gardening on health."

Abstract:

The ultimate goal of development is to improve health and well-being at the individual level. However, much of the data that is routinely collected measures development outcomes at the household level. This dichotomy would not be a concern if households' collective decision making was accurately described by the unitary household model. However, since collective decision making is the norm intra-household allocations are important, though these have proven difficult to measure in the direct and indirect methods proposed in the existing literature.

The purpose of this paper is to propose a new, direct method to measure individual-level well-being and intra-household allocation. Using NHANES I-III as a reference population age-gender specific Body Mass Index (BMI) distributions are estimated for all ages 2-90 years old. These growth curve distributions form the basis for calculating BMI Z-scores that are consistent for children and adults and that can be used to examine intra-household allocation patterns.

BMI Z-scores constitute an improvement over existing methodologies to measuring intra-household allocations. BMI Z-scores are an indicator that is commonly available for all household members and are a health outcome measure rather than a health input measure such as consumption. Moreover, collecting height and weight information necessary to calculate BMI is much cheaper and less prone to measurement error than eliciting consumption data.

The statistical method to estimate BMI Z-scores is proposed here is grounded in best practice and uses LMS estimation in a GAMLSS framework which allows completely flexible modelling of all necessary parameters. This paper's estimates of age-gender specific BMI distributions provide the necessary information to calculating BMI Z-scores for any population. In turn these Z-scores allow a wide range of previously unmeasurable within-household comparisons, for example, comparing any two individuals or comparing relative movement of any household member versus the household average. Moreover, the use of BMI Z-scores opens up additional sources of survey data to answer new research questions in health and development economics.

1 Introduction

Who benefits from changes in public policy? Who gets adversely affected by shocks? These are core questions that have motivated the large theoretical and empirical program evaluation literatures. While these literatures have become increasingly methodologically sophisticated in addressing these questions at the household level they often struggle with evaluating impacts at the level of the individual. Though, arguably, it is changes at the individual level that we are even more interested in. Did everyone in the household suffer equally from, say, a period of drought or were the impacts primarily on the most vulnerable household members? Were the gains from a new public policy shared equally within the household?

Empirically examining these intra-household effects of policies and other shocks has often proved problematic primarily because of the difficulty and expense of collecting individual-level information. In most household surveys monetary, caloric and nutrient consumption data is only collected at the household-level. Thus, by necessity the existing intra-household allocation literature has relied on indirect methods to tease out individual consumption from total household aggregates. A select few surveys did collect separate consumption data for men and women but it is even rarer for surveys to also include individual level data for children and other household members.

This paper proposes a new method to evaluate individual and intra-household effects of policies and shocks based on Body Mass Index (BMI) data. The BMI has two main advantages compared to consumption data: it is often available in existing data (and simple to collect in new data) and it provides a direct indicator for health and welfare outcomes, in contrast to consumption which measures inputs. The method proposed in this paper constructs BMI Z-scores that are consistent for people of all ages. It employs state-of-the-art growth curves estimation techniques and applies them to the same three rounds of National Health and Nutrition Examination Survey (NHANES) reference population data used by the 2000 Center for Disease Control (CDC) growth charts.

The next section reviews methods and findings from the literature on intra-household allocation. Sections 3 discusses the advantages of using BMI Z-scores as a consistent and accurate measure of nutritional status and well-being. This paper's core statistical methodology of constructing BMI Z-scores using Generalized Additive Models of Location, Scale and Shape (GAMLSS) is laid out in section 4. Sections 5 and 6 introduce the NHANES data and present the estimation results. Conclusions round out the paper.

2 Intra-household allocation – methods and findings from the existing literature

The unitary household assumes that demand and consumption is proportionally shared among household members and that this demand only depends on total household income, independent on who earned this income. This theoretical model has been framed as a model of a single benevolent decision maker, as perfect altruism, or as common preferences. Empirically, this model has been rejected time and time in again in favor of collective household models which improve upon the unitary household model in that intra-household allocation concerns are no longer ignored. In these collective household models each HH member has their own rational preferences, the source of income matters and the final demand and consumption functions are the outcome of non-cooperative, within household bargaining and subsequent sharing rules (see, for instance, Becker (1973, 1981) Chiappori (1988, 1992)). Typically, in survey data we only observe household-level income and consumption. Thus, to get to the individual level consumption it is necessary to recover the sharing rule implicit in the household. The existing literature has attempted to do this in several innovative but ultimately not fully satisfactory ways.

One strand of the intra-household allocation literature examines the effects of income earned by different household members on consumption and on anthropometric outcomes.

(Thomas 1990) examines male and female non-labor income and treats them as exogenous, in contrast to endogenous male and female labor income. He then separately estimates the effect of male and female non-labor income on a) budget shares for food, housing, health, leisure, and adult and household consumption goods, b) nutrient and calorie demand per capita, and c) children's WFH and HFA anthropometric outcomes. Compared to men, women in Brazil are found to spend more on education, health, household goods, nutrients and calories, and on children, in particular on girls. Extra income for women increase WFH and HFA outcomes 4-8 times as much than if extra income is earned by men. (Thomas 1997) extends this analysis and comes to similar conclusions. In Cote d'Ivoire (Haddad and Hoddinott 1994) uncover similar patterns for girls' vs boys' anthropometric outcomes as a function of women's income share in the household. They also detect a bias towards biological and against adopted children.

(Hoddinott and Haddad 1995) use Ulph's (1988) non-cooperative Nash household model and find that bargaining occurs between members, rather than the household acting as a unitary utility maximizer. Using data from Cote d'Ivoire they find that an increase women's income leads to an increase in food budget shares and to a decrease in the budget shares of alcohol and cigarettes. These studies clearly reject the common-preference neoclassical unitary household model but are limited to only testing for the effect of gendered income sources on child outcomes and *total* household budget shares.

A second strand of the literature concentrates on intra-household consumption differences. In the absence of direct data on individual consumption these studies estimate individual consumption indirectly and thus make inference on allocation within the household. One variant of this literature strand exploits knowledge about male vs female crops and their attendant incomes. Households in Cote d'Ivoire tend to use income from jointly produced crops such as yams to finance household public goods that are consumed jointly. Income from female and male crops instead is used for buying adult private goods. Female crop income also raises food consumption (Duflo and Udry 2004). (Haddad and Hoddinott 1994) find similar patterns in Cote d'Ivoire.

A second approach in this strand of the literature tries to overcome the lack of data on intra-household allocation by distinguishing between adult and other consumption goods and use that to identify allocations of resources towards boys vs girls. (Deaton 1989) proposes such a method to make inference based on household-level expenditure data. The key identification strategy is to determine an 'adult' good and then assume that for a given level of income, households with more children spend less on adult goods. If the household had a preference towards boys, then a household that has more boys would spend less on adult goods than if it had the same number of girls. Deaton finds no evidence of child gender bias in Cote d'Ivoire and a small, statistically insignificant bias towards boys in Thailand. This weak result could reflect the actual situation in these two settings or also reflect that too much is asked of the modeling assumptions.

A third approach is to distinguish between male and female consumption goods. (Wang 2014) examines the effect of individual-level transfer of property rights from housing reform in China on individual consumption and time allocation. Tenants could buy property from their state employers with property rights granted at the individual level. When property rights went to men household consumption of some male goods and women's time spent on chores increased. Conversely, when women received property rights consumption of some male goods fell.

A fourth approach is based on the Rothbart (1943) method that was originally developed to model resource allocation between adults and children by estimating how much consumption of adult goods falls in the presence of children. (Bargain, Donni et al. 2014) use an extension of the Rothbart method that allows for scale economies for household public goods that are jointly consumed. They infer total individual level consumption from observations of total household consumption and that of a single, individual adult consumption good. Their findings for Cote d'Ivoire suggest that differences in consumption are small between men and women but that children are disadvantaged within the household.

Using a different extension of the Rothbart method (Dunbar, Lewbel et al. 2013) examine children's resources in collective households in Malawi and Cote d'Ivoire. They identify an individual's consumption from household consumption through semiparametric restrictions on individual preferences within a collective model by observing how consumption of a single private consumption good varies with income and family size. Children receive a share of household resources of between 5 and 20%. Men receive more than women and boys get more than girls, particularly in Malawi and to a lesser extent in Cote d'Ivoire. This suggests substantial intra-household allocation differentials which, when accounted for, substantially raise poverty estimates compared to using simple per adult equivalent adjustments of total household consumption.

The key to any Rothbart-based methodology is the strong assumption that we know the shape of the function that links a single adult good to total household resources. Only then is observing a single adult consumption good sufficient to recover overall household allocation patterns. Nevertheless, this approach probably represents best practice in the indirect identification of individual consumption from total household consumption. However, it still focuses on input indicators such as consumption rather than outcomes.

A third, smaller strand of the literature uses outcomes and nutritional inputs to measure intra-household allocations. (Sahn and Younger 2009) use the BMI to compare outcomes for children and adults. However, BMI levels cannot be directly compared across all ages, which is the prime motivation for this paper's new methodology. The related health sciences literature falls into two broad categories. The first has used anthropometric outcomes in the form Z-scores for WFA, HFA, WFH and BMI. These studies have only studied children as Z-score calculations have historically been limited to children. Or it has resorted to using different measures for household members of different ages as. For example, (Thomas, Lavy et al. 1996)'s evaluation of structural adjustment on health outcomes in Côte d'Ivoire uses WFH Z-scores for children and BMI for adults. However, using different outcome indicators does not allow inference on the relative, intra-household effects of the policy.

The second category has studied dietary diversity in terms of nutrients and calories (see (Anuradha 1998) for an overview). However, constructing such health input measures, rather than anthropometric outcome measures introduces two major sources of measurement error: survey recall bias and incomplete food to nutrient conversion factors. Not every food has a conversion factor, and not every green bean has the same nutrient content. Nutritional input indicators are problematic for three other reasons. First, different activity patterns require different amount of inputs to produce a healthy 'output'. For instance, manual agricultural laborers require more calories than their fellow household members who take on more domestic work. Various calorie thresholds exist by age, gender and activity, but there is no unified method within and across countries, and even if there was agreement on calorie thresholds, it is unlikely that measurement could be accurate enough. Second, individuals vary with respect to their metabolism. Third, Recommended Daily Allowances are undergoing constant change and there are no universal standards for calorie consumption.

In sum, existing approaches to identifying individual level consumption and health outcomes from household survey data have been innovative but suffer from either methodological shortcomings, strong untestable assumptions, or conceptual problems. The BMI Z-score measure introduced below suggests an improved method that focuses on outcomes rather than inputs and is consistent across all ages and make use of commonly available height and weight survey data.

3 The BMI as an indicator nutrition & the advantages of BMI Z-scores

Weight-stature indices are commonly used to measure to measure nutritional status. The most common among these indices is the Body Mass Index (BMI) defined as

$$BMI_i = \frac{weight_i}{height_i^2}$$

with weight and height for person ***i*** measured in kilograms and meters. The BMI is a good indicator of nutritional status as it is only weakly correlated with height (ideally it should be uncorrelated) and strongly correlated with body fat (Cole 2014). It is considered the most suitable, objective anthropometric indicator of nutritional status of individuals and populations of all ages (FAO, 1994). A BMI smaller than 18.5 indicates chronic undernutrition and values greater than 25 suggests overweight.

In addition, the BMI possesses a number of practical advantages. It is accurate as the probability of misclassifying nutritional status on the basis of the BMI is very small. It is closely related to individual food consumption as well as to inadequacy of food in the community. The height and weight data needed are relatively easy to collect in surveys and inexpensive to analyze. Further, the BMI can be used for the purpose of nutritional surveillance and for monitoring the effectiveness of intervention programs, and it also allows for interregional and intercountry comparisons over seasons, years or decades (FAO, 1994).

Anthropometric Z-scores are common statistical tools to assess child growth and nutritional outcomes. Traditionally, absolute BMI numbers were used for adults while BMI Z-scores defined as

$$Z - BMI_i = \frac{BMI_i - \overline{BMI_{RP}}}{\sigma_{BMI,RP}} \qquad \textit{Equation 1}$$

were used for children, where $\overline{BMI_{RP}}$ and $\sigma_{BMI,RP}$ represent the mean and the standard deviation of the BMI in the reference population. While this common practice is sufficient for studying either children or adults separately it does not allow for comparison across ages and within a household. Yet, there is no statistical or anthropometric reason why adults' BMI value cannot be transformed into Z-scores as well. Statistically, Z-scores are just a transformation of a level variable. Indeed, Z-scores are likely to be preferable, even when not attempting to make intra-household comparisons.

The BMI Z-score methodology developed in this paper improves upon the methods used in the existing program evaluation and intra-household allocation literatures in six different ways. First, BMI Z-scores allow analyzing individual and intra-household effects of policy changes and shocks on health and welfare status. Second, they do this using an easy-to-collect anthropometric outcome indicator (BMI) rather than a difficult-to-collect input indicator, such as calorie or nutrient consumption or consumption expenditure that has commonly been used in health and development economics. Third, measuring BMI is less prone to measurement error than individual-level consumption or expenditure data. Fourth, in most circumstances our interest lies in comparing individual as well as intra-household health outcomes. Then, using BMI Z-score outcomes rather than caloric or monetary inputs into a health production function circumvents variations related to individual activity patterns and metabolism and is unaffected by periodic revisions of RDAs for nutrients and the absent RDAs for calories.

Fifth, BMI varies greatly for young children and adolescents. This is the reason why children's BMI values are routinely transformed into Z-scores. However, even for adults the BMI undergoes life-cycle changes. The expected BMI first increases before decreasing again at an old age. This pattern is visible and statistically significant in the NHANES reference population data that are used in the CDC growth charts and in this paper. The BMI for adults (age 20 and older) varies systematically with age for women, men and for the adult population as a whole (see results in Appendix 1). Thus, even for adults BMI Z-scores are preferable when comparing changes over time or when making population-level comparisons across different ages of adults. Sixth, using BMI Z-scores opens up a large number of existing datasets for new kinds of health and development economic analysis, for example, DHS surveys, which contain height and weight information but little information individual-level health inputs and expenditures; or other economic household surveys that provide consumption and expenditure information for the entire household but not for individual members.

4 The methodology to construct BMI growth curves

Until the 1990s ago the statistics behind growth curves was rather simple: distributions for anthropometric variables were estimated for each gender and age group. These estimates were then smoothed across age groups either by hand or by simple polynomial regression[1] (Cole 2014). The former involves obvious subjectivity, whereas the latter is too simplistic given the distributions of anthropometric indicators, including the BMI, do not resemble normal distributions, particularly in terms of skewness.

[1] For a full history of the anthropometrics and statistics behind growth chart estimation see Cole, T. J. (2012). "The development of growth references and growth charts." Annals of human biology **39**(5): 382-394.:

Two key methodologies have since been proposed to estimate growth centiles for individuals from a reference population that can appropriately manage skewed distributions. First, using non-parametric quantile regression (Koenker, 2005; Koenker and Ng (2005), He and Ng (1999) and Ng and Maechler (2007). And second, using the parametric LMS method initially developed by Cole (1988) and Cole and Green (1992). This paper uses an extension of the LMS method to calculate BMI growth centile reference distributions using the NHANES I-III data as the reference population.

Z-scores assume an underlying normal distribution. If the reference distribution used to calculate Z-scores is skewed then Z-scores calculated from the raw, non-transformed distribution are incorrect. Hence, the reference distribution needs to be corrected for skewness. Healey et al 1988, Cole 1988 and (Cole and Green 1992) introduced the LMS method of estimating growth curves that are consistent in the presence of skewness. The original LMS method assumes the Box- Cox Cole and Green distribution (BCCG) for the response variable, BMI. 'LMS' refers to the three parameters of the BMI distributions to be estimated for each age and gender subgroup where μ is the median (M), σ is the coefficient of variation (S), and ν is the skewness measure (L).

Let BMI be a random variable with range BMI > 0 defined through the transformed variable Z_{BMI} given by:

$$Z_{BMI} = \begin{cases} \frac{1}{\sigma\nu}\left[\left(\frac{\mathrm{BMI}}{\mu}\right)^{\nu} - 1\right] \text{ if } \nu \neq 0 \\ = \frac{1}{\sigma} log\left(\frac{BMI}{\mu}\right)^{\nu} \quad \text{if } \nu = 0 \end{cases} \qquad \text{Equation 1}$$

Then the LMS method assumes $Z_{BMI} \sim N(0,1)$ and $BMI \sim BCCG(\mu, \sigma, \nu)$. The L, M and S parameters define the expected distribution of the BMI outcome variable at any given age and for each gender. The power transformation parameter **ν** corrects for skewness. For instance, if the data are right skewed then **ν** is chosen to be smaller than 1. If the data are not skewed and, thus, **ν**=1 it is easy to show that the equation 2 reduces to the common Z-score in equation 1. Dividing by **μ** centers the distribution around 1 and dividing by **σ** controls for the spread of the distribution.

Each of these three LMS parameters is allowed to vary by age, t. Thus, L(t), M(t) and S(t) are estimated for each age (and gender) before being smoothed across t. The original LMS method uses natural cubic smoothing splines with knots at each t (Cole and Green 1992) .

In the presence of kurtosis in the BMI distribution the LMS method gives incorrect Z-score estimates as the LMS transformations are insufficient to reduce a kurtotic BMI distribution to a standard Normal. Rigby and Stasinopoulos extend the LMS method to the class of Generalized Additive Models for Location, Scale and Shape (GAMLSS) (Rigby and Stasinopoulos 2005) in two ways. First, GAMLSS additionally allows for modelling kurtosis in the data by introducing for a fourth curve fitting parameter, τ, and by replacing the BCCG distribution with the Box-Cox power exponential (BCPE) (Rigby and Stasinopoulos 2004) or the Box-Cox t (BCT) distributions (Rigby and Stasinopoulos 2006). Respectively, they termed these extended LMS models LMSP and LMST. Second, GAMLSS allows each of the four first moments of the distribution to be modeled separately.

The LMSP assumes that the transformed random variable Z has a truncated)exponential power distribution with $Z_{BMI} \sim PE(0,1,\tau)$ and $BMI \sim BCPE(\mu,\sigma,\nu,\tau)$. LMST assumes that Z has a truncated t-distribution with $Z_{BMI} \sim t_\tau$ and $BMI \sim BCT(\mu,\sigma,\nu,\tau)$. The LMSP and the LMST models can be expressed as subcases of GAMLSS. Note that in the absence of kurtosis, that is when τ=2, the LMSP and LMST models reduce to the simpler LMS model.

In the case of centile estimation for BMI distributions given the explanatory variable ***x***=age$^{\xi}$, the GAMLSS model is

$$BMI \sim D(\mu,\sigma,\nu,\tau)$$

$$\log(\mu) = s_1\left(x, df_{\mu}\right)$$

$$\log(\sigma) = s_2(x, df_\sigma)$$

$$\nu = s_3(x, df_\nu)$$

$$\log(\tau) = s_4(x, df_\tau)$$

The monotonic link functions on the left-hand side above could take more general forms but in empirical applications to the BMI the link function is typically the identity function for ν and log for the other three parameters. The four *s* functions on the right-hand side are non-parametric smoothing functions, for example the penalized B-splines used in the estimations below.

GAMLSS centile estimation, hence, requires the estimation of five hyper-parameters: the degrees of freedom for μ, σ, ν and τ and the age transformation, ξ, which stretches the time scale for BMI and makes smooth curve fitting more accurate in the presence of BMI growth spurs at different ages.

Having estimated these five hyper-parameters we can now transform individual BMI observations, both from the reference population and any other population, and calculate BMI Z-scores using equation 2.

The methodology for constructing BMI Z-scores outlined above compares favorably with existing best-practice growth curves that have been constructed at the international, the national and the local level. At the international and national level the 1977 National Center for Health Statistics (NCHS) growth reference charts have been replaced by the 2006 WHO Growth Standards and 2007 Growth Reference, and the 2000 Center for Disease Control growth charts, respectively.

The more recent efforts by the World Health Organization (WHO) (Borghi, de Onis et al. 2006) represents the current the gold standard of growth curve and Z-score estimation. This group of statisticians and child growth experts conducted a thorough survey of 30 different methods for growth curve estimation, including quantile regressions and the variants of the LMS method described above. Their recommendation led the WHO to adopt the BCPE/LMSP method to estimate age-gender specific BMI distributions and to use cubic splines to smooth across age as these splines performed better than fractional polynomials.

The WHO technical report on the WHO Multicenter Growth Reference Study (MGRS) project tested the goodness of fit for these using MGRS data for children under the age of 6 from six countries: Oman, Norway, Brazil, Ghana, India and the US (Organization" 2006). The general approach to estimation was to use GAMLSS with the BCPE distribution(Rigby and Stasinopoulos 2004). However, the final selected model estimated τ as a constant equal to2. Thus, the model simplified to the standard LMS model (Cole and Green 1992) since the empirical BMI distribution did not require adjustment for kurtosis. Therefore, the computation of BMI z-scores in the 2006 WHO Child Growth standards (http://www.who.int/childgrowth/en/) are based on the LMS method.

The WHO (de Onis, Onyango et al. 2007) subsequently used the same BCPE methodology applied to the 1977 NCHS data to calculate the WHO Reference 2007 growth curves (http://www.who.int/growthref/en/) for children aged 5-19 as a methodologically and statically consistent complement to its Child Growth Standards.

Conceptually, the methodology in this paper mirrors that used by the 2006 WHO Growth Standards and 2007 WHO Reference though differs in the reference population used (see next section) which in turn, for the male subsample, leads to preferring the BCT method to the very similar BCPE method used by the WHO.

At the US national level the CDC 2000 growth charts pre-date LMSP and LSMT methodologies and were estimated using the simple LMS method using both simpler smoothers for the age-gender specific BMI distributions as well as simpler parametric and non-parametric smoothers across age groups (Kuczmarski, Ogden et al. 2002). For a detailed comparison of the WHO child growth standards and the CDC 2000 growth charts see (De Onis, Garza et al. 2007)

Two recent studies on small, rural subsistence populations in the Amazon in Bolivia (Blackwell, Urlacher et al. 2017) and Peru (Urlacher, Blackwell et al. 2016) construct ethnically specific, local population Z-scores. Their methodologies closely follow the WHO's approach of using GAMLSS to estimate Z-scores.

In sum, all existing best-practice approaches to estimating anthropometric Z-scores at the international, national and local level use the LMSP variant of the GAMLSS framework that is also used in this paper. The next two sections describe the NHANES data and how using these data to construct BMI Z-score ultimately leads to selecting the LMST and LMSP variants of GAMLSS for men and women, respectively.

5 The NHANES reference population data

The choice of reference population is clearly important but there is no clear 'best choice'. Two key characteristics need to be weighed for the particular measurement objective (Cole 2014).

First, should the reference population be representative of the target population (a "growth reference") such as the NHANES is for the US population, or of a healthy population (a "growth standard") such as the 2006 WHO Growth Standards whose reference population was not representative of the populations in the six countries but was purposefully selected to only include non-malnourished and non-overweight children.

Second, is the growth reference to be used clinically, in which case it should reflect the particular local population or is its main use for public health assessment when it needs to be nationally (and internationally) comparable? For the purposes of the BMI Z-scores developed in this paper we need a standard that is internationally comparable. Therefore, the analysis below draws on NHANES I, II and III (1960-1990s) as the reference population. These data pre-date the steep increase in obesity in the US and therefore represent a better fixed standard to measure against.[2] Further, the NHANES data have historically been used to construct anthropometric growth curves, including the 2000 CDC and the 1977 NCHS growth curves that were used extensively both within the US as well as for other populations around the world.

The 2006 WHO Growth Standards followed a similar approach to construct its international growth charts for children but chose a broader reference population based on 6 country datasets from around the world. This makes it a more appropriate growth reference for children (under 6) but since it does not contain BMI for anyone older than 6 these data cannot be used to construct consistent BMI Z-scores for all ages.

The reference population data was constructed from the first three rounds of the NHANES (United States Department of 1975, United States Department of 1980, United States Department of, Human Services. Centers for Disease et al. 1998). Sample sizes of the number of people sampled and the number of people observed are given in Table 1.

Table 1 NHANES I-III sample sizes

	NHANES I	NHANES II	NHANES III
Number sampled	32,000	27,801	31,311
Number observed	23,308	20,322	31,071
Age range	1-74 years	6 months – 74 years	2 months - 90 years
Number of complete observations, age>2 years	23,226	19,254	27,361

BMI Z-scores were constructed for people 2 years and older as only recumbent length, which is not directly comparable to height measures, is available for ages less than 2. The variable measuring age at date of examination was constructed by month for children and youth younger 20 years and by year for adults older than 20. After discarding observations for anyone below the age of 2 and dropping observations for which either height, weight or age were not available the final merged dataset comprises 69,841 observations with the breakdown across NHANES I-III shown in Table 1.

6 Empirical Implementation and Results

The methodology for constructing internally consistent BMI Z-scores for children and adults follows the WHO's best practice methodology. The data used are the pooled NHANES I-III. Observations are weighted by NHANES sample weights. The actual LMS parameter and BMI centile curve estimation is implemented using the function lms() in the R package GAMLSS (Rigby and Stasinopoulos 2005). Estimations are run separately for the subsamples of 37,497 women and 32,344 men.

The first step is to let the data decide which particular variant of the GAMLSS methodology provides the best fit: LMS, LMSP or LMST. For the male subsample the LMST model using the BCT distribution was a

[2] For the same reasons the UK has frozen its BMI growth curve standards at the 1990 level Hall, D. M. B. and T. J. Cole (2006). "What use is the BMI?" Archives of Disease in Childhood **91**(4): 283-286.

better fit compared to the LMS and LMSP method as it minimized both the Global Deviance and the Akaike Information Criterion (see Table 2). For the female subsample the LMSP model using the BCPE distribution fit best. The appropriate link functions for both male and female subsamples were identified as the identity function for ν and the log function for μ, σ and τ.

Second, having chosen the LMST and LSMP models for men and women, respectively, we need to determine the five hyper-parameters that are used to model the non-parametric fitting of the age-gender specific BMI curves: the four effective degrees of freedom for the non-parametric smoothing functions dof_μ , dof_σ , dof_ν and dof_τ , and the power transformation function ξ for age. The fast growth in BMI for young children required a power transformation for the age variable.

The five hyper-parameters were chosen by minimizing the GAIC[3] with k=3 (Rigby and Stasinopoulos 2006) and the Validation Global Deviance (VGD) from within all LMST models. The validation by global deviance is the preferred fitting method and is faster (Rigby and Stasinopoulos 2005). It splits the data into 60% and 40% components and then uses the former as training for fitting, before using the second component to estimate the 5 parameters. However, for these data the AIC parameter choices were the same. The LMST and LMSP models estimated the 5 hyper-parameters using penalized B-splines using restricted maximum likelihood estimation. These results are summarized in Table 2.

Table 2 – Effective degrees of freedom of smoothing parameters & power transformation of age

	Male	Female
Effective dof_μ	17.46	18.64
Effective dof_σ	18.38	16.67
Effective dof_ν	15.77	2.00
Effective dof_τ	7.81	9.81
Age power transform ξ	0.80	0.95
dof for the fit	59.43	47.13
Best fitting distribution	Box-Cox t	Box-Cox Power Exponential
Global Deviance	165696.2	210071.9
Akaike Information Criterion, k=3	165815.1	210166.1
# of observations	32344	37497

These first two steps produce the percentile values of the smoothed BMI curve for each age and gender bracket as well as the LMS parameters of the median (μ), generalized coefficient of variation (σ), the power of the Box-Cox transformation (ν), and the kurtosis-correction parameter (τ) that can be used to construct BMI Z-scores.

Third, since these GAMLSS models are fitted non-parametrically we need to check whether the fitted model chosen also produces a good local and global fit. To measure the goodness of fit of GAMLSS model we cannot use raw residuals as these hold only for normal distributions. Similarly, deviance residuals and Pearson residuals are inadequate as they don't allow for multiple parameter distributions of BMI and for highly skewed and kurtotic data, respectively. Thus, the assessing the fit of a GAMLSS

[3] Other possible values for k include k=2 which represents the Akaike Information Criterion and k=log(n) which is the Bayesian Information Criterion. K = 3 is good compromise

requires the use of normalized quantile residuals (Rigby and Stasinopoulos 2005). If the model fits, the BMI measurements follow a standard normal distribution on all ages after a suitably chosen Box-Cox transformation. This assessment of fit was done using three different types of diagnostics on the residuals.

First, by examining the moments of the normalized quantile residuals. These are summarized in Table 3 and show that the residual distributions are very close to normal with means and skewnesses close to zero, variance close to one, and kurtosis close to 3. Graphical tests of the normality assumptions are included in the appendix and confirm the results in Table 3.

Table 3 Summary of the Quantile Residuals

	Women	Men
Mean	-0.003013808	-0. 0002686209
Variance	0. 999273	0. 9971081
Skewness	0. 0004258214	-0. 0215842
kurtosis	3.062806	2.964691

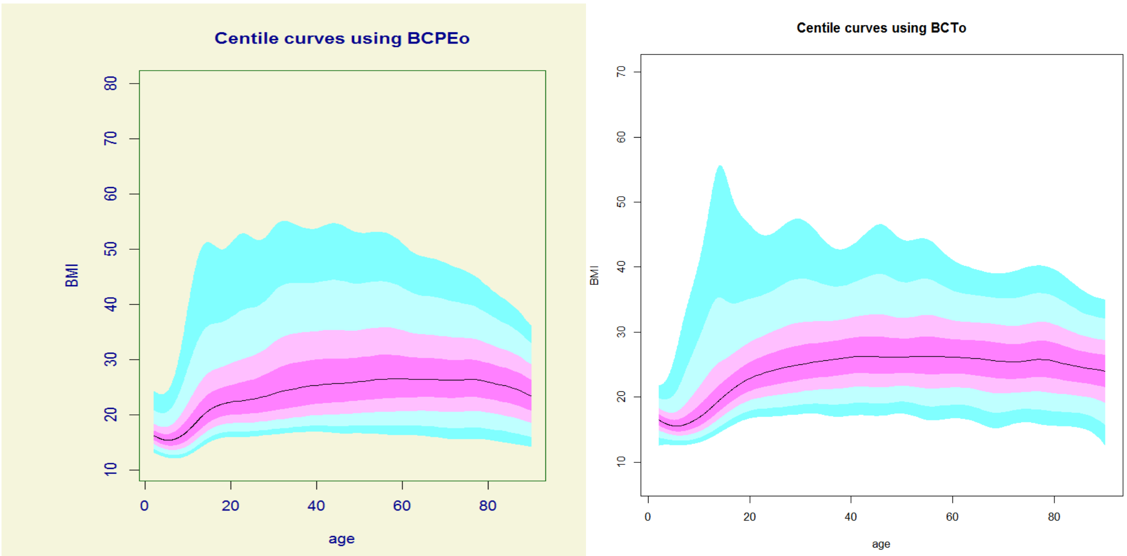
Second, we use worm plots which identify whether the fitted distribution of normalized quantile residuals is adequate. The worm plot is a de-trended QQ plot and shows, by age group, whether the transformed BMI observations are normally distributed, as is required for constructing Z-scores (van Buuren and Fredriks 2001). The worm plots' offset, slope and curvature show to what extent the transformation of the BMI observations (see equation 2) have centered the distribution, reduced the standardized variance to 1, and eliminated skewness and kurtosis.

The level of the worm plot compared to the horizontal zero degree line shows the transformed location parameter, μ, has mean zero. The slope of the worm plot indicates residual variance with a positive slope indicating too high residual variance. Quadratic and cubic shapes of the worm plot suggest residual presence of skewness and kurtosis in the transformed data.

9 worm plots each for men and women are given in the appendix representing equal-sized splits in the age distribution. In all 18 worm plots mean, variance and skewness appear adequately controlled for. There may be some degree of platy-kurtosis in the quantile residuals, suggesting the transformed BMI variable still has some remaining high kurtosis. To investigate whether this platy-kurtic pattern in the worm plot is an actual model violation we can examine the coefficient parameters of the cubic polynomial line fitted through the residuals shown in the worm plot. None of these coefficients (show in the appendix below the worm plots) are significant suggesting that any remaining kurtosis does not violate normality assumptions necessary for constructing valid BMI Z-scores.

Third, we use Q-statistics (Royston and Wright 2000) to assess whether the mean, variance, skewness and kurtosis of the residuals differ from a normal distribution. Visual representations and exact numerical values of the four Q-statistics, Z1, Z2, Z3 and Z4 are given in the appendix. Again, these are provided for different age groups. Some of these statistics suggest some skewness and kurtosis, particularly for ages between 10 and 13. However, for example in the case of men, only 4 out of 84 Q-statistics suggest departure from normality which is on par with being random at a 95% confidence level.

Based on these three different normality checks for the quantile residuals the transformed age-gender specific BMI distributions are approximately normal. Figures 1 below show these BMI distributions as BMI growth curves smoothed across ages for women (on the left) and men (on the right).



Using these estimated BMI distributions for all ages between 2 and 90 years old we can transform any individual-level BMI observations into BMI Z-scores that are consistent for children and adults. Mathematically, this is done by inserting BMI survey observations and the estimated parameters μ, σ, ν, and τ into equation 2.

In principle, this can be done to create internal BMI Z-scores, that is, for BMI observations from within the sample used to construct the BMI growth curves. (Urlacher, Blackwell et al. 2016) and (Blackwell, Urlacher et al. 2017) do this for particular local Amazonians populations. Though, unless the objective is to study historical BMI patterns in the US this is not particularly meaningful application for these NHANES-based estimated BMI distributions.

Instead, the purpose of this paper is to provide a methodology to calculate external BMI Z-scores, that is use the estimates of the LMST and LMSP parameter from the NHANES data to construct BMI Z-scores for a different populations. Thus constructed external BMI Z-scores based on NHANES can be used to make global comparisons, in the same way as is common for other anthropometric Z-scores where populations in developed and developing countries are compared to healthy US populations. Note also that when the objective is to measure changes in health over time rather than health status at a particular point in time, then the choice of reference population is less important.

We can now use any other dataset and calculate consistent, external BMI Z-scores for people of all ages. These Z-scores can then be used as a left-hand side outcome variable to analyze intra-household allocations and changes in these allocations. This represents an improvement over existing papers which had to limit their comparisons to women vs men, adults vs children or girls vs boys and/or have typically relied on diet adequacy or consumption expenditure data, with the former being a rather crude indicator of health outcomes and the latter typically only available directly for at most two household members. Using BMI Z-scores instead provides a better indicator of health outcomes as well as allowing comparisons beyond two household members. For example, these BMI Z-scores can be used to test the level of – and the change in- status of any individual household member vs the household average or vs any other household member. This could include studying birth order effects (with and without gender dimension), the status of the most disadvantaged, which depending on the setting could be girl children or daughters-in-law residing in their husband's parents household, and others. Moreover, using BMI Z-scores as the outcome indicators of well-being allows new analyses using existing datasets, such as the Demographic and Health Surveys, that contain much detailed information, but lack measures of well-being that economists typically like to put on the left-hand side of their regressions.

7 Conclusion

Households are not single units that make joint, fair decisions in the best interest of all household members. Gains and losses in total household well-being are not equally shared. We, therefore, cannot use household aggregates of income and consumption collected by economic surveys to directly identify the well-being of individual household members or has this may have been affected by shocks and policies. The existing literatures on intra-household allocation has introduced innovative approaches to get at this indirectly but these indirect approaches by design have to rely on strong assumptions or focus on narrow definitions of intra-household allocation or both.

The purpose of this paper is to propose a new, direct method to measure individual-level well-being and intra-household allocation. Using NHANES I-III as a reference population age-gender specific BMI distributions are estimated for all ages 2-90 years old. These distributions form the basis for calculating BMI Z-scores that are consistent for children and adults.

BMI Z-scores constitute an improvement over existing methodologies to measuring intra-household allocations, primarily as BMI Z-scores are an indicator that is (or can easily be made) available for all household members. Another key advantage is that BMI Z-scores measure an outcome of health rather than consumption or income which are inputs into health production which are positively but perfectly correlated with health outcomes. Moreover, collecting height and weight information necessary to calculate BMI is much cheaper and less prone to measurement error than eliciting consumption data.

The statistical method to estimate BMI Z-scores is proposed here is grounded in best practice and uses LMS estimation in a GAMLSS framework which allows completely flexible modelling of all necessary parameters. This paper's estimates of age-gender specific BMI distributions provide the necessary information to calculating BMI Z-scores for any population. In turn these Z-scores allow a wide range of previously unmeasurable within-household comparisons, for example, comparing any two individuals or comparing relative movement of any household member versus the household average. Moreover, the use of BMI Z-scores opens up additional sources of survey data to answer new research questions in health and development economics.

8 Bibliography

Anuradha, S. (1998). Intra-household consumption patterns - Issues, evidence and implications for human development. Background paper for 1998 Human Development Report.
Bargain, O., O. Donni and P. Kwenda (2014). "Intrahousehold distribution and poverty: Evidence from Côte d'Ivoire." Journal of Development Economics **107**: 262-276.
Blackwell, A. D., S. S. Urlacher, B. Beheim, C. von Rueden, A. Jaeggi, J. Stieglitz, B. C. Trumble, M. Gurven and H. Kaplan (2017). "Growth references for Tsimane forager-horticulturalists of the Bolivian Amazon." American Journal of Physical Anthropology **162**(3): 441-461.
Borghi, E., M. de Onis, C. Garza, J. Van den Broeck, E. A. Frongillo, L. Grummer-Strawn, S. Van Buuren, H. Pan, L. Molinari, R. Martorell, A. W. Onyango and J. C. Martines (2006). "Construction of the World Health Organization child growth standards: selection of methods for attained growth curves." Stat Med **25**(2): 247-265.
Buuren, S. v. and M. Fredriks (2001). "Worm plot: a simple diagnostic device for modelling growth reference curves." Statistics in Medicine **20**(8): 1259-1277.
Cole, T. J. (2012). "The development of growth references and growth charts." Annals of human biology **39**(5): 382-394.
Cole, T. J. (2014). Growth and Development. Wiley StatsRef: Statistics Reference Online, John Wiley & Sons, Ltd.
Cole, T. J. and P. J. Green (1992). "Smoothing reference centile curves: the LMS method and penalized likelihood." Stat Med **11**(10): 1305-1319.
De Onis, M., C. Garza, A. W. Onyango and E. Borghi (2007). "Comparison of the WHO child growth standards and the CDC 2000 growth charts." Journal of Nutrition **137**(1): 144-148.
de Onis, M., A. W. Onyango, E. Borghi, A. Siyam, C. Nishida and J. Siekmann (2007). "Development of a WHO growth reference for school-aged children and adolescents." Bull World Health Organ **85**(9): 660-667.
Deaton, A. (1989). "Looking for Boy-Girl Discrimination in Household Expenditure Data." The World Bank Economic Review **3**(1): 1-15.
Duflo, E. and C. Udry (2004). Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices. NBER Working Paper 10498. NBER.
Dunbar, G. R., A. Lewbel and K. Pendakur (2013). "Children's Resources in Collective Households: Identification, Estimation, and an Application to Child Poverty in Malawi." American Economic Review **103**(1): 438-471.
Haddad, L. and J. Hoddinott (1994). Household Resource Allocation in the Côte d'Ivoire: Inferences from Expenditure Data. Poverty, Inequality and Rural Development: Case-Studies in Economic Development, Volume 3. T. Lloyd and O. Morrissey. London, Palgrave Macmillan UK: 67-88.

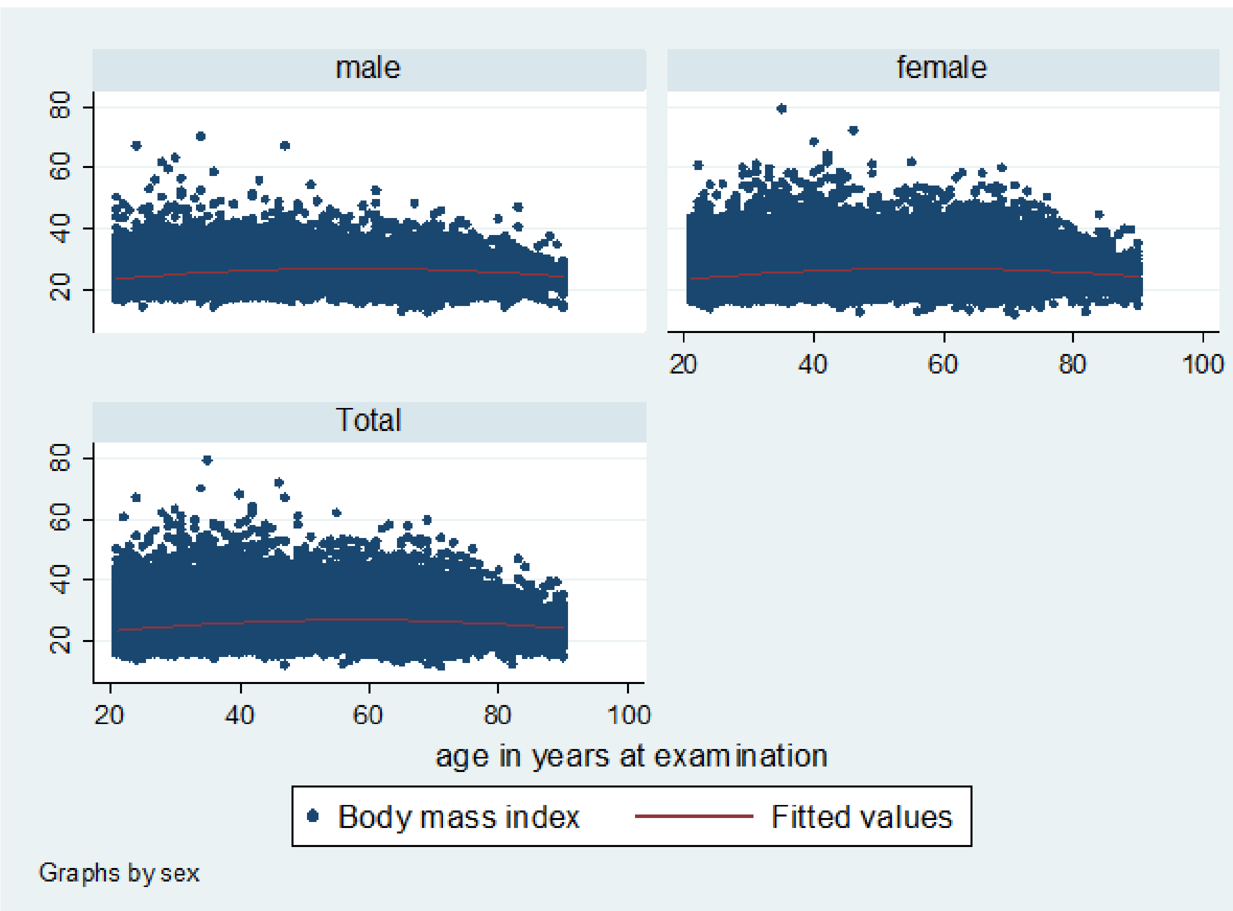
Haddad, L. and J. Hoddinott (1994). "Women's income and boy-girl anthropometric status in the Côte d'Ivoire." World Development **22**(4): 543-553.
Hall, D. M. B. and T. J. Cole (2006). "What use is the BMI?" Archives of Disease in Childhood **91**(4): 283-286.
Hoddinott, J. and L. Haddad (1995). "DOES FEMALE INCOME SHARE INFLUENCE HOUSEHOLD EXPENDITURES? EVIDENCE FROM CÔTE D'IVOIRE." Oxford Bulletin of Economics and Statistics **57**(1): 77-96.
Kuczmarski, R. J., C. L. Ogden, S. S. Guo, L. M. Grummer-Strawn, K. M. Flegal, Z. Mei, R. Wei, L. R. Curtin, A. F. Roche and C. L. Johnson (2002). "2000 CDC Growth Charts for the United States: methods and development." Vital Health Stat 11(246): 1-190.
Organization", W. H. (2006). WHO Child Growth Standards: Methods and Development. Geneva, WHO Press.
Rigby, R. A. and D. M. Stasinopoulos (2004). "Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution." Statistics in Medicine **23**(19): 3053-3076.
Rigby, R. A. and D. M. Stasinopoulos (2005). "Generalized additive models for location, scale and shape." Journal of the Royal Statistical Society: Series C (Applied Statistics) **54**(3): 507-554.
Rigby, R. A. and D. M. Stasinopoulos (2006). "Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis." Statistical Modelling **6**(3): 209-229.
Royston, P. and E. M. Wright (2000). "Goodness-of-fit statistics for age-specific reference intervals." Statistics in Medicine **19**(21): 2943-2962.
Sahn, D. E. and S. D. Younger (2009). "Measuring intra-household health inequality: explorations using the body mass index." Health Economics **18**(S1): S13-S36.
Thomas, D. (1990). "Intra-Household Resource Allocation: An Inferential Approach." The Journal of Human Resources **25**(4): 635-664.
Thomas, D. (1997). "Incomes, expenditures, and health outcomes: Evidence on intrahousehold resource allocation." Intrahousehold resource allocation in developing countries: 142-164.
Thomas, D., V. Lavy and J. Strauss (1996). "Public policy and anthropometric outcomes in the Côte d'Ivoire." Journal of Public Economics **61**(2): 155-192.
United States Department of, H., C. Human Services. Centers for Disease and S. Prevention. National Center for Health (1998). National Health and Nutrition Examination Survey III, 1988-1994, Inter-university Consortium for Political and Social Research (ICPSR) [distributor].
United States Department of, H. a. H. S. (1975). National Health and Nutrition Examination Survey (NHANES I). Hyattsville, MD, Centers for Disease Control and Prevention.

United States Department of, H. a. H. S. (1980). Second National Health and Nutrition Examination Survey (NHANES II). Hyattsville, MD, Center for Disease Control and Prevention.
Urlacher, S. S., A. D. Blackwell, M. A. Liebert, F. C. Madimenos, T. J. Cepon-Robins, T. E. Gildner, J. J. Snodgrass and L. S. Sugiyama (2016). "Physical growth of the shuar: Height, Weight, and BMI references for an indigenous amazonian population." American Journal of Human Biology 28(1): 16-30.
van Buuren, S. and M. Fredriks (2001). "Worm plot: a simple diagnostic device for modelling growth reference curves." Stat Med 20(8): 1259-1277.
Wang, S.-Y. (2014). "Property rights and intra-household bargaining." Journal of Development Economics 107: 192-201.

9 Appendix 1 Adult BMI varies systematically with age

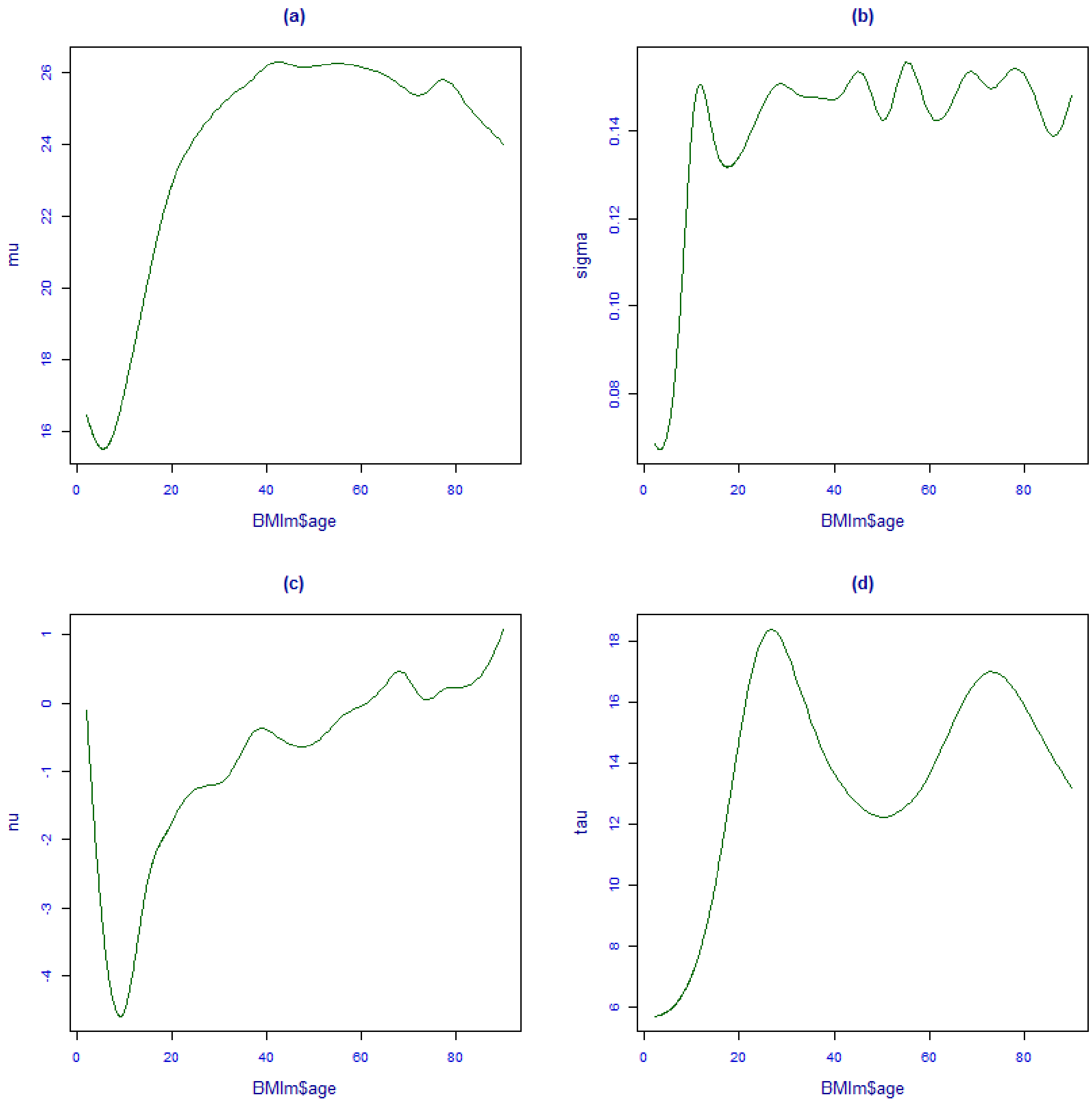
	All adults
age	0.281
	(24.56)**
age_2	-0.002
	(21.12)**
_cons	18.732
	(73.93)**
R^2	0.04
N	44,163

* $p<0.05$; ** $p<0.01$



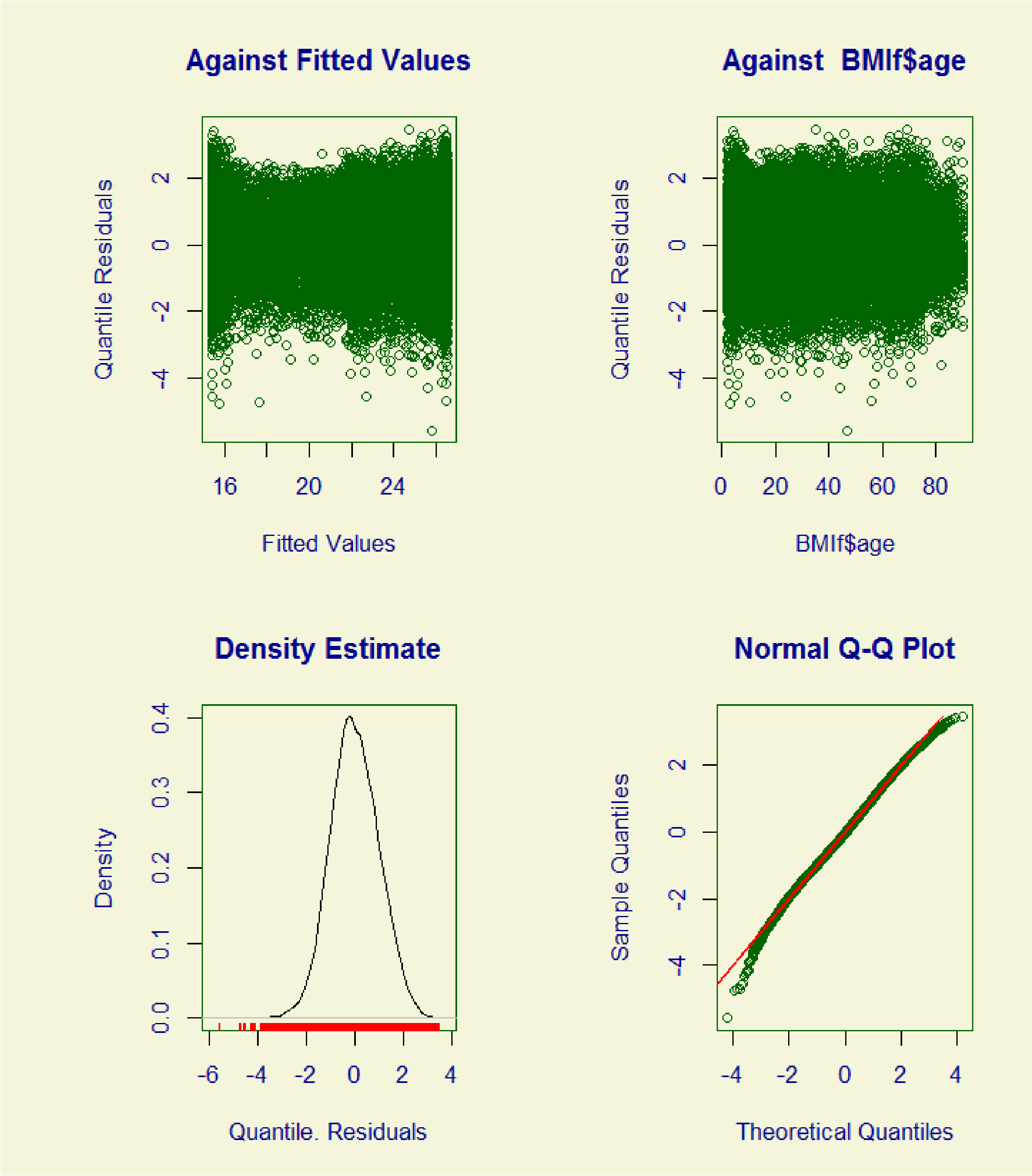
10 Appendix 2 Model selection details

10.1 Estimated LMST parameters smoothed across age

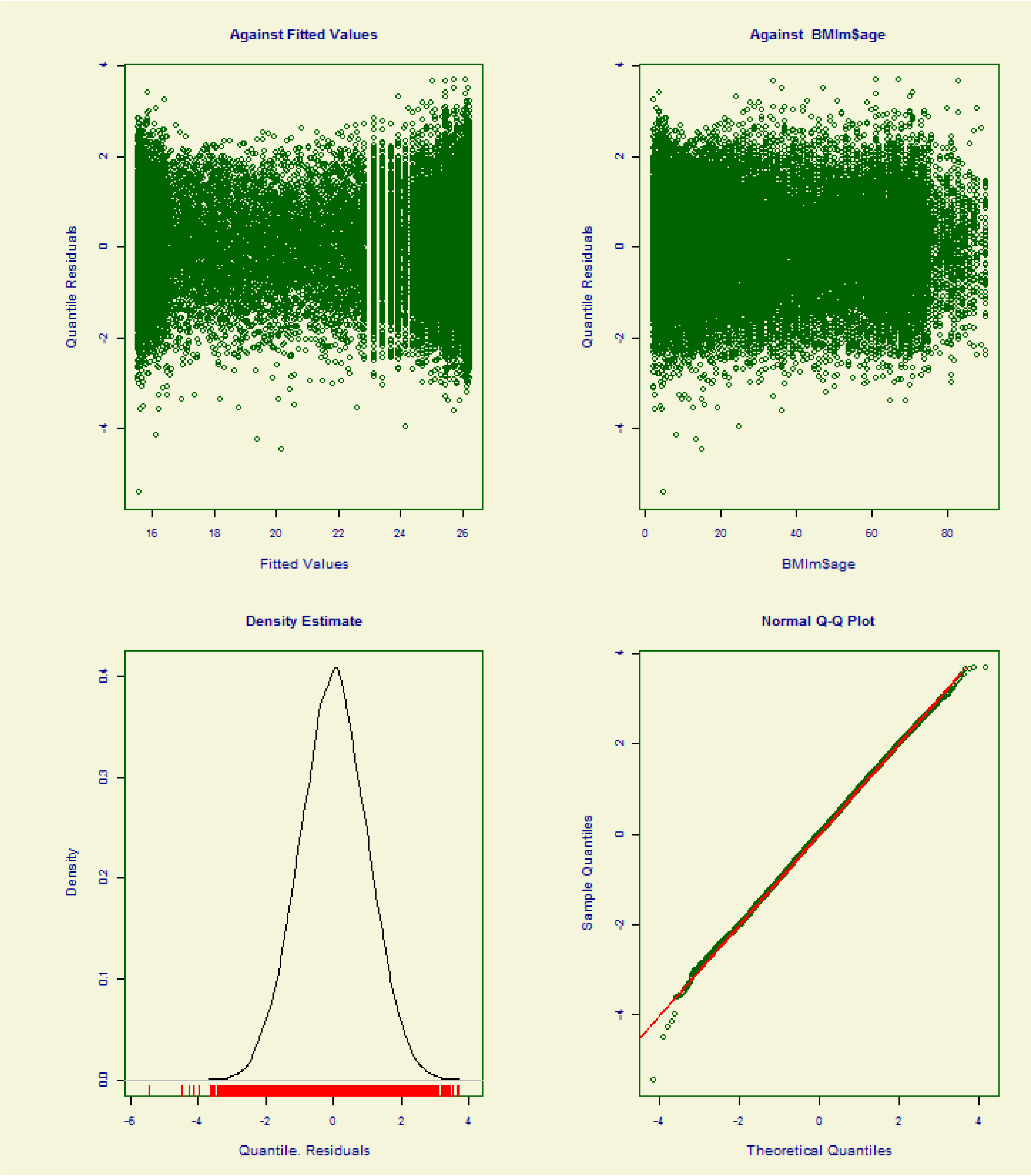


10.2 Graphical tests for normality of transformed BMI data

For women



For men

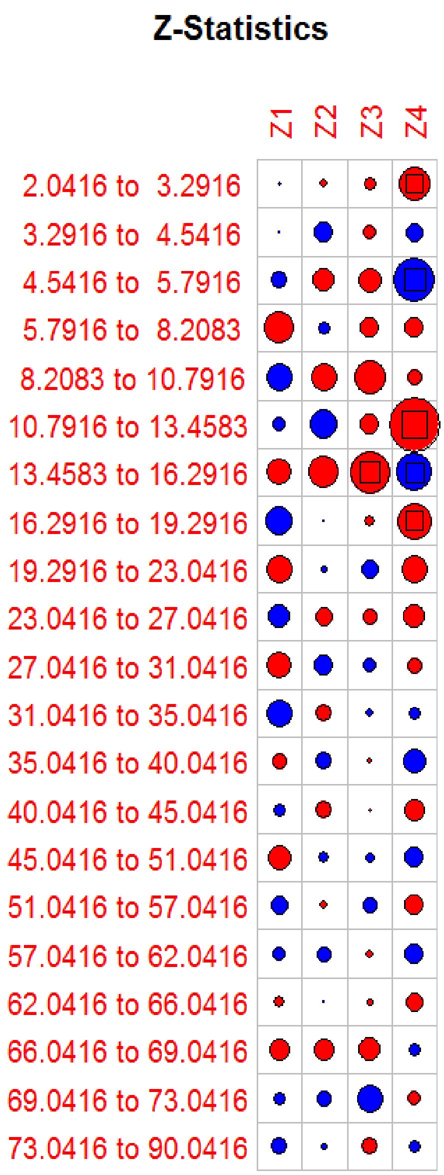


Against Fitted Values
Quantile Residuals
Fitted Values
Against BMIm$age
Quantile Residuals
BMIm$age
Density Estimate
Density
Quantile. Residuals
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles

10.3 Z-statistic tests

Z1 and Z2 test for standard normal mean and variance. Z3 and Z4 are test for skewness and kurtosis.

For men (left) and women (right)



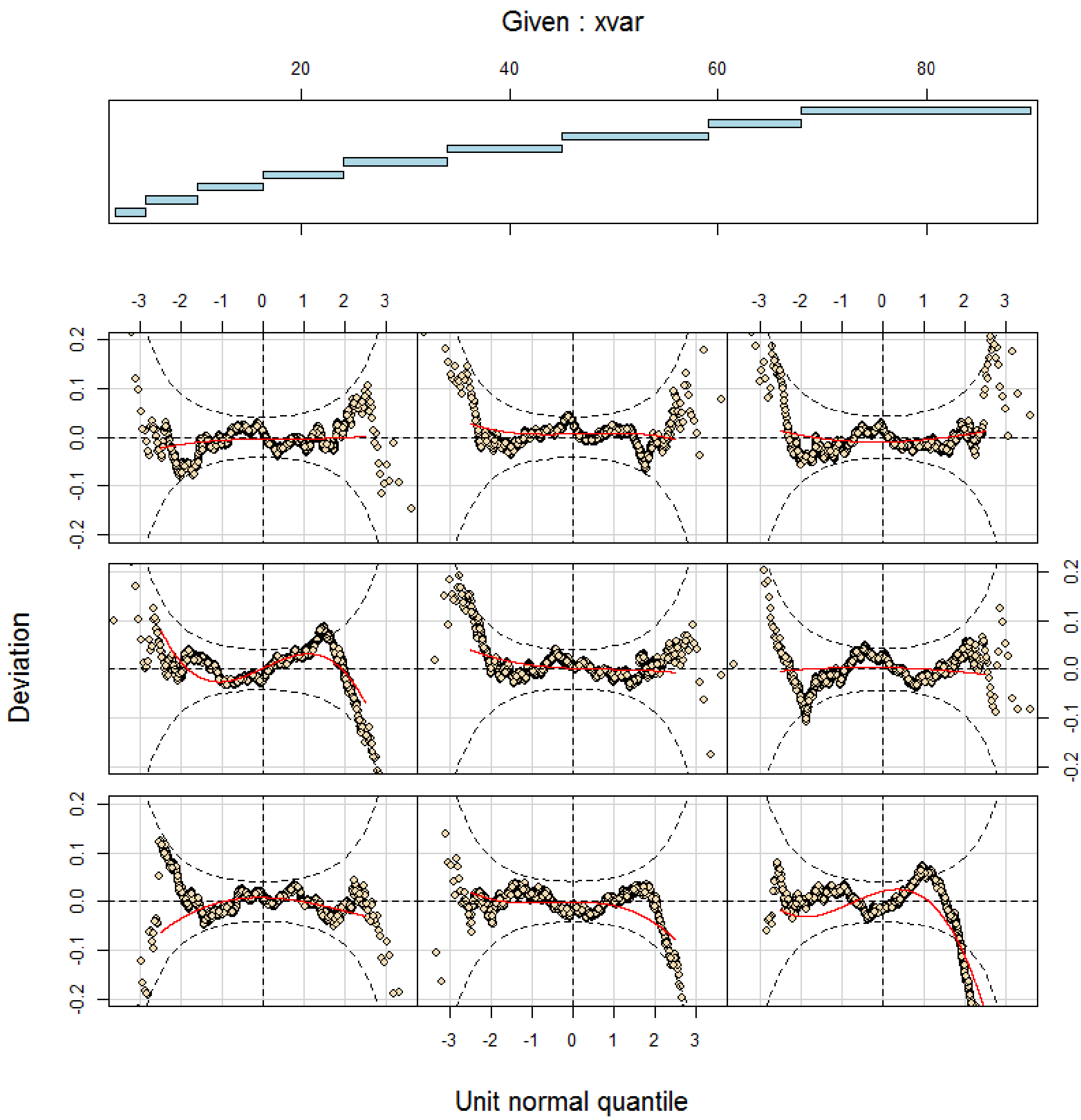
Any Z-statistic > 2 indicates model violation. Positive/negative values indicate higher/lower mean,variance, skewness, kurtosis than a standard normal.

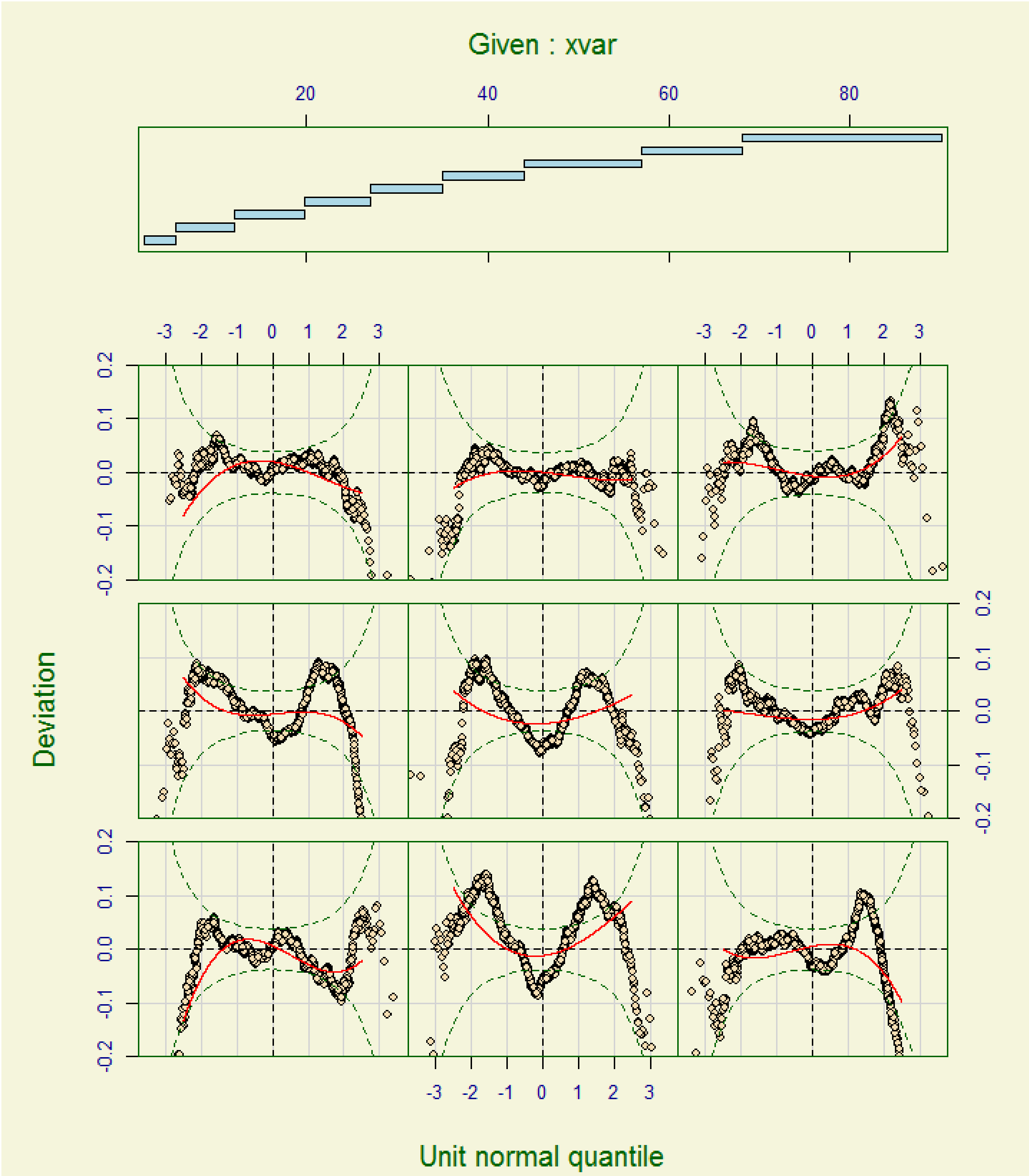
And compare the Agostino k2 statistic to the 5% value of a chi-squared distribution with 2 degrees of freedom, i.e. 6.

```
> round(Q.stats(am, xvar=BMIm$age),3)
                          Z1     Z2     Z3     Z4 AgostinoK2     N
 2.0416 to  3.3749 -0.143 -0.283 -0.062 -1.572      2.476  1715
 3.3749 to  4.6250  0.441  0.631 -0.623  0.671      0.837  1533
 4.6250 to  6.0416 -0.787 -0.696 -1.225  2.392      7.221  1685
 6.0416 to  8.7083 -0.019 -0.003 -0.463 -0.513      0.477  1546
 8.7083 to 11.3749  1.636 -1.209 -1.714 -1.934      6.679  1616
11.3749 to 14.3749 -1.021  1.323 -1.414 -3.185     12.142  1658
14.3749 to 17.2916 -0.631 -2.019 -3.014  2.068     13.362  1609
17.2916 to 21.0416  0.356  0.545  0.338 -1.979      4.031  1829
21.0416 to 25.0416  0.040 -0.266 -0.339 -1.434      2.172  1688
25.0416 to 29.0416  0.049  0.226  0.692 -1.053      1.587  1625
29.0416 to 33.0416  0.124 -0.165  0.105 -0.643      0.425  1457
33.0416 to 38.0416 -0.579  0.315 -0.388  2.077      4.465  1599
38.0416 to 43.0416  0.368 -0.441  0.081 -0.638      0.413  1550
43.0416 to 49.0416  0.284  0.716  0.104 -0.334      0.122  1655
49.0416 to 55.0416 -0.486 -0.081  0.167 -0.077      0.034  1567
55.0416 to 61.0416  0.576  0.006  0.312  1.098      1.302  1676
61.0416 to 65.0416 -0.760 -0.326  0.300 -0.226      0.141  1530
65.0416 to 69.0416 -0.302 -0.217 -1.393 -0.775      2.543  1862
69.0416 to 73.0416 -0.073  0.018  1.853 -0.317      3.533  1571
73.0416 to 90.0416  0.783  0.241 -0.629  0.274      0.471  1373
TOTAL Q stats       7.507  9.723 23.005 41.429     64.434 32344
df for Q stats      2.536 10.309  4.227 12.182     16.409     0
p-val for Q stats   0.039  0.493  0.000  0.000      0.000     0
> round(Q.stats(af, xvar=BMIf$age),3)
                          Z1     Z2     Z3     Z4 AgostinoK2     N
 2.0416 to  3.5416 -0.877  0.906 -4.262  2.412     23.983  1813
 3.5416 to  5.0416 -0.075 -1.306  0.402  2.279      5.354  1837
 5.0416 to  6.9583  0.006  0.286  2.137  1.871      8.069  1743
 6.9583 to 10.0416  1.532 -0.484  4.274 -1.623     20.904  1771
10.0416 to 13.1249 -0.378  1.085 -2.907 -1.216      9.933  1765
13.1249 to 16.3750 -0.006 -1.633 -1.275 -1.500      3.876  1806
16.3750 to 19.7916 -0.548  0.802  0.851  0.858      1.461  1764
19.7916 to 23.0416  0.408  0.453 -0.482 -0.624      0.621  2167
23.0416 to 26.0416 -0.370 -0.697  0.821  0.466      0.891  1718
26.0416 to 29.0416 -0.650 -1.128  1.025 -0.697      1.536  1698
29.0416 to 32.0416  0.343  1.341 -0.458  0.669      0.658  1663
32.0416 to 36.0416 -1.527 -0.084  1.605  0.269      2.647  1952
36.0416 to 40.0416  0.090  0.112  1.131 -0.814      1.942  1862
40.0416 to 44.0416 -1.151  0.011  0.730  1.217      2.013  1768
44.0416 to 49.0416  0.955  0.027 -1.945  3.378     15.193  1594
49.0416 to 55.0416 -1.057 -0.710 -0.196 -1.539      2.407  1813
55.0416 to 61.0416  0.325 -0.132 -2.002  1.974      7.905  1849
61.0416 to 65.0416  0.224  0.670 -0.937  1.000      1.876  1622
65.0416 to 69.0416  0.524 -0.047  1.494  0.046      2.235  2006
69.0416 to 73.0416 -0.641  0.137 -0.262  1.027      1.124  1710
73.0416 to 90.0416  0.251 -0.126  0.782  0.543      0.906  1576
TOTAL Q stats      11.010 12.119 70.140 45.394    115.534 37497
df for Q stats      2.357 12.165 18.997 11.187     30.185     0
p-val for Q stats   0.006  0.450  0.000  0.000      0.000     0
```

10.4 Worm plots

For men




Given : xvar
20
40
60
80
-3 -2 -1 0 1 2 3
0.2
0.1
0.0
-0.1
-0.2
Deviation
Unit normal quantile

Coefficient parameters for each of the 9 cubic polynomials fitted for nine detrended QQ plots in the worm plot.

Suggested cut-off points for identifying model violations are beta0>01., beta1>0.1, beta2>0.05 and beta3>0.03 (Buuren and Fredriks 2001). None of the parameters below are above the threshold.

	[beta,1]	[beta,2]	[beta,3]	[beta,4]
[1,]	0.009067451	-1.446059e-03	-0.0086200641	0.0018603106
[2,]	0.001560811	6.015645e-03	-0.0046811218	-0.0051045235
[3,]	0.018195652	1.708290e-02	-0.0233144719	-0.0100272671
[4,]	0.003924669	3.716032e-02	-0.0001733043	-0.0105059178
[5,]	0.001316348	-4.083230e-03	0.0028418160	-0.0006065150
[6,]	0.004401942	-5.232666e-05	-0.0026046184	-0.0003994103
[7,]	-0.004784666	4.403923e-03	-0.0013381785	0.0003791700
[8,]	0.008335320	2.227839e-03	0.0006404087	-0.0014578242
[9,]	-0.012199908	-9.480698e-04	0.0042178044	-0.0000119379

For women

	[beta,1]	[beta,2]	[beta,3]	[beta,4]
[1,]	0.006673344	-0.0303628820	-0.013373135	0.0084208563
[2,]	-0.011946871	0.0076820534	0.018169159	-0.0020412624
[3,]	0.005881616	0.0131799439	-0.008764777	-0.0052230315
[4,]	-0.006091077	0.0063925591	0.002123181	-0.0044983211
[5,]	-0.023633816	0.0035051579	0.009113526	-0.0008356451
[6,]	-0.015987125	0.0006343269	0.005868203	0.0011515480
[7,]	0.020259020	-0.0101260238	-0.012615657	0.0029671243
[8,]	0.000853557	-0.0072210857	-0.003396215	0.0016859991
[9,]	-0.007040849	-0.0076689054	0.008033274	0.0027350141