NBER WORKING PAPER SERIES

THE TREND OF BMI VALUES OF US ADULTS BY CENTILES, BIRTH COHORTS 1882-1986

John Komlos
Marek Brabec

Working Paper 16252
http://www.nber.org/papers/w16252

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2010

The work was partly supported by the Czech Institutional Research Plan AV0Z10300504 ‘Computer Science for the Information Society: Models, Algorithms, Applications'. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Trend of BMI Values of US Adults by Centiles, birth cohorts 1882-1986
John Komlos and Marek Brabec
NBER Working Paper No. 16252
August 2010, Revised September 2010
JEL No. I10

ABSTRACT

Trends in BMI values are estimated by centiles of the US adult population by birth cohorts 1886-1986 stratified by ethnicity. The highest centile increased by some 18 to 22 units in the course of the century while the lowest ones increased by merely 1 to 3 units. Hence, the BMI distribution became increasingly right skewed as the distance between the centiles became increasingly larger. The rate of change of BMI centile curves varied considerably over time. The BMI of white men and women experienced upsurges after the two World Wars and downswings during the Great Depression and again after 1970. However, among blacks the pattern is different during the first half of the century with men's rate of increase in BMI values decreasing substantially and that of females remaining unchanged at a relatively high level until the Second World War. However, after the war the rate of change of BMI values of blacks resembled that of the whites with an accelerating phase followed by a slow down around the 1970s. In sum, the creeping nature of the obesity epidemic is evident, as the technological and lifestyle changes of the 20th century affected various segments of the population quite differently.

John Komlos
Department of Economics
Ludwig-Maximilians University
Ludwigstrasse 33/IV
Munich D-80539
GERMANY
and CESifo
and also NBER
john.komlos@gmx.de

Marek Brabec
Academy of Sciences of the Czech Republic
Pod Vodarenskou vezi 2
182 07, Praha 8
Czech Republic
mbrabec@szu.cz

The Trend of BMI Values of US Adults by Centiles, birth cohorts 1882-1986

Keywords: BMI, US, NHANES, obesity, overweight, semiparametric modelling, gamlss model, percentile estimation

Introduction

Komlos and Brabec (2010) recently estimated the trend in the mean BMI values of US-born adults by birth cohorts to find that they have been increasing continuously throughout the 20th century. This "creeping" nature of the trend is quite contrary to the received wisdom which tends to place the onset of the obesity epidemic in the final quarter of the previous century. However, they also found that the rate of increase in BMI values varied quite a bit with two periods of particularly rapid acceleration in BMI values following the two World Wars. Insofar as they have discussed in that paper the advantages and disadvantages of the birth-cohort approach as opposed to the period effects that has been the overwhelming focus of research up to now, we shall not reiterate the issues here. The current aim is to expand those results which explored exclusively the mean BMI values by estimating trends by centiles for four categories of adults, for whites and blacks by gender using the same NHANES data sets collected between 1959 and 2006.

Method

For modeling the BMI distribution and its dependence on several covariates, we use the approach based on the generalized additive model for location, scale, and shape (GAMLSS), developed by Rigby and Stasinopoulos (2005, 2006, 2007). In principle, this can be seen as a generalization of the generalized linear model (GLM) (McCullagh, Nelder 1989), as well as of the generalized additive model (GAM) (Hastie, Tibshirani 1990), or even of the LMS[1] approach (Cole 1988). The advantage of GAMLSS is that it enables one to fit not only the mean of the distribution as a function of the covariates, as is usual in linear, nonlinear, or nonparametric regression, but also other characteristics. Similarly as in GAM, variability can be modeled in detail, as well. Yet, in GAMLSS, the modeling is more flexible as it allows other moments (i.e., skewness and kurtosis) to change with the covariates. This is necessary if one is interested in realistic and flexible description of the whole BMI distribution and its changes with several explanatory variables. The distribution itself can be characterized by centiles and their changes over the range of the selected covariates. Because this is precisely our aim, we need to allow for departures from normality and for estimation of several characteristics of the distribution simultaneously: i.e., mean, variability, skewness and kurtosis.

In particular, after some experimentation, we model the BMI distribution using the Box-Cox t family, BCT(μ,σ,ν,τ) (Rigby and Stasinopoulos 2006). This is a parametric but very flexible family of distributions having parameters μ,σ,ν,τ . Variable Y with positive ($\Re^+$) support[2] has the BCT(μ,σ,ν,τ) distribution if the transformed variable Z has the following form:

$$
\begin{aligned}
Z &= \frac{1}{\sigma\nu}\left[\left(\frac{Y}{\mu}\right)^{\nu}-1\right], \quad if\ \ \nu \neq 0 \\
&= \frac{1}{\sigma}\log\left(\frac{Y}{\mu}\right), \qquad\quad if\ \ \nu = 0
\end{aligned}
\tag{1}
$$

Z is a truncated standard t distribution with τ degrees of freedom (where $\tau > 0$ does not need to be an integer). Truncation at zero is induced by the positivity of Y . In our case of BMI, the amount of truncation is very small. As shown in Rigby and Stasinopoulos (2006), under such circumstances, μ can be interpreted approximately as the median of Y , σ as the interquartile-range-based coefficient of variation as a measure of relative variability,[3] ν controls skewness, and τ controls kurtosis, or just how heavy the tails of Y are.

Our model allows the BCT(μ,σ,ν,τ)'s parameters to change with the covariates in a flexible, nonparametric way. Specifically, we use the cubic spline family (Eubank 1988, Green and Silverman 1994, Rigby and Stasinopoulos 2007) to model dependence of

μ, σ, ν and τ on covariates. In other words, we model the link-transformed[4] parameter as cubic splines in continuous variables plus effects of factors in the ANOVA style (Graybill 1976, Rawlings 1988) for discrete variable coding the education level of a particular person. We use identity link for μ, ν and log link for σ, τ parameters. In other words, we model μ, ν, $\log(\sigma)$ and $\log(\tau)$ by cubic splines. We also assume independence among individual responses. Strictly speaking, this is not reflecting the clustering induced by the survey sampling design used in NHANES data, but we use this as a reasonable approximation. We considered the extent to which adding the random primary sampling unit effect affects the model – and found that it did not change the estimates substantially.

Thus, our model is described by the following equations:

$$BMI_i \sim BCT(\mu_i, \sigma_i, \nu_i, \tau_i) \tag{2}$$

$$\mu_i = cs_\mu(Age_i, 4) + cs_\mu(Birth\_yr_i, 5) + cs_\mu(PIR_i, 2) + \sum_{m=1}^{3} \alpha_{\mu m} I(E_i = m)$$

$$\log(\sigma_i) = cs_\sigma(Age_i, 1) + cs_\sigma(Birth\_yr_i, 2) + cs_\sigma(PIR_i, 1) + \sum_{m=1}^{3} \alpha_{\sigma m} I(E_i = m)$$

$$\nu_i = cs_\nu(Age_i, 1) + cs_\nu(Birth\_yr_i, 1) + cs_\nu(PIR_i, 1) + \sum_{m=1}^{3} \alpha_{\nu m} I(E_i = m)$$

$$\log(\tau_i) = cs_\tau(PIR_i, 1),$$

where $I(.)$ is the indicator function which equals 1 if the condition in its argument is true, and 0 otherwise. E_i is the level of education of the subject. $cs(x, d)$ is the cubic spline in a variable x with d degrees of freedom.[5] BMI_i is the BMI for the i-th person. Similarly, Age_i is the age in years, $Birth\_yr_i$ is the birth year, PIR_i is the Poverty Income Ratio for the i-th person.

μ, σ, ν, τ change from individual to individual, but only through changes in various covariates. Unlike the others, τ changes only with a single covariate, PIR. Nevertheless, both spline parts involved in $\mu_i, \sigma_i, \nu_i, \tau_i$ as well as in the educational effects $\alpha_{\mu m}, \alpha_{\sigma m}, \alpha_{\nu m}, m = 1,2,3$ are (simultaneously) estimated via the Rigby, Stasinopoulos 2005) algorithm from the data. In particular, they are not assumed a priori as they would be if, for example, one would assume normality.

The degrees of freedom for splines in various variables are very important in that they control the smoothness of the fit. Therefore, they ought not be set arbitrarily. Instead, they were selected using GAIC, or generalized Akaike information criterion (Rigby and Stasinopoulos 2005). Only integer values of the degrees of freedom were considered in the search. Compared to the model of Komlos and Brabec (2010), model (2) allows different smoothness in different variables as well as different smoothness in the same explanatory variables for different characteristics e.g. $cs_{\mu}(Age_i, 4)$ and $cs_{\sigma}(Age_i, 1)$). Note that generally, for more complicated characteristics (from μ to τ), the curves are less complex (basically smoother), as expected.

We show the results of the weighted analyses (weighting by reciprocal variances). Shape of the centile curves does not change substantially if we recompute the model in unweighted fashion, however. We do the computations using the gamlss package (Rigby, Stasinopoulos 2007) from the “R” software environment (R 2010), together with some additional code written by us.[6] Those individuals with missing values of either BMI and/or any of the explanatory variables were excluded from the estimation. We explored the model fit by means of centile residuals considering various plots, similar to those used in standard regression, e.g. residuals vs. fitted, Q-Q plots, histograms of residuals, and also at worm plots (van Buuren and Fredriks 2001).

Results

That the persistent increase in BMI values began already among the birth cohorts of the late 19th century is confirmed by these estimates in all four groups (Figures 1-4). There are a number of similarities and differences in the experience of the four groups under consideration. In all four groups the shapes traced out by the BMI centiles can be characterized as having a shape of a half-fan in the sense that the upper centiles move up as the ridges of a fan while the lower ones remained essentially unchanged. Consider that the highest centiles increased by some 20, 20, 18, and 22 units (WM, WF, BM, BF) during the period under consideration while the lowest ones increased by merely 3, 1.5, 1, and 2 units. This is an indication that the distribution did not shift outward uniformly. Its shape has been deformed considerably and continuously so that it became increasingly skewed to the right.

Figures 1-4 about here

Another way of describing this pattern is to consider the variation across the centiles. These also indicate that the variance increased continuously as the centiles rotated upward (Figure 5). Obviously, the increase in variance is accompanied by a substantial skewing of the distribution toward the more obese range, rather than by a uniform increase in the whole BMI spectrum. One can also consider, moreover, the dates at which various centiles of a birth cohort which reached 30 BMI units, the conventional definition of obesity, as a measure of this upward rotation (Table 1 and Figure 6). The rate of rotation was rather similar among white males and females, and black males. Among men it took on average about 19 years for an additional centile to reach a BMI value of 30 while among white and black women it took 17 years and 13 years respectively. The black females were often 30-40 years ahead of the other three groups in reaching the level of obesity in a particular centile.

Insert Table 1 and Figures 5-6 about here

Rate of change of the centile curves

The rate of change of BMI centile curves were obtained by numerical differentiation of the centile functions estimated by model (2) with respect to the date of birth. These varied substantially over time in all the four groups under study (Figures 7-10). Initially, the rate of change was lowest among white men born in the 19th century and remained constant until the turn of the 20th century. This was followed by a rapid acceleration in BMI values around World War I. The acceleration was accompanied by a marked divergence among the centiles (leading to increased BMI variability), particularly in the upper ones, a divergence that continued during the remainder of the century (leading to increased skewness to the right of the BMI distribution). However, the rate of change peaked in the mid-1920s and decelerated during the Great Depression, reaching a nadir during the Second World War (Figure 7). During the war the rate of change among white men was still positive in most of the centiles, though at the lower centiles the rate dipped below that experienced in the late 19th century. However, in the upper centiles the rate was well above those of the 19th century even during the war. Another turning point was reached in the early 1950s as BMI values accelerated once again similarly to the pattern obtained after the First World War. Yet, the second upswing of acceleration in the lower centiles was both considerably shallower than the first one and reached a plateau quickly in the 1950s. By the birth cohorts of the early 1960s the rate of change of BMI values was constant or even negative among the lower centiles. Only in the higher centiles did the acceleration persist until the present day and pass the previous peak rate reached in the mid 1920s (Figure 7).

Figure 7 about here

In many respects the rate of change of white female BMI values has a similar pattern to that of white men (Figure 8). It remained fairly constant in the 19th century; it also accelerated around the two World Wars. However, the World War I acceleration lasted longer: the peak rate in the top centiles was reached in the mid-1930s instead of the mid-1920s as among their male counterparts. Moreover, the deceleration of the Great Depression was shallower and also lasted longer, - until the very end of the war. The subsequent acceleration also began at mid-century, as among men, and lasted until about 1970 at which time the rate of change either remained constant or declined somewhat particularly in the lower centiles. A similar flattening of the curves at least in the lower centiles among men occurred in the mid 1970s. In short, the salient pattern is similar among white men and women. The main difference is in the lengths and turning points of the cycles.

Figure 8 about here

In contrast, among blacks the pattern is quite different from that of whites in the pre-World War II era but becomes quite similar after mid-century. Among black men (Figure 9), the rate of change began at a higher level but declined practically continuously until World War II. The inter-centile range was as large as among the white women to begin with, but did not increase at all until after World War II. Furthermore, in contrast to that experienced by whites, the World War I upswing was inconsequential and meant only a short interruption of the persistent decline in the rate of change. Moreover, the post-World War II upswing began earlier than among whites, i.e., in the early 1940s, and lasted until the mid-1960s, when a decline set in, somewhat earlier than that among white women.

Figure 9 about here

The pattern among black women (Figure 10) was equally unique in the first half of the century insofar as the rate of change was high already to begin with and continued almost uninterrupted at that high level until mid century. The range between the lowest and highest centile was large at the beginning and, as among black men, did not widen at all in the first half of the century, contrary to the pattern among the whites. The post-World War II upswing started around 1960 among the highest centile women, but was a bit delayed among the lowest centiles. The peak rate of change was reached around 1969 among the highest centile black women, in 1971 among white women and in 1960 among black men. The highest centile white men did not have a local maximum during the post-World War II era as rates continued to rise until the end of the period.

Figure 10 about here

Confidence Intervals

The 95% confidence intervals were obtained for a given percentile curve as envelope bands, based on 500 bootstraps from a simplified model, without weighing.[7] They are often asymmetric (Figure 11) - but the degree of asymmetry varies across the different centiles. This reflects the amount of information that is contained in the data for the estimation of a given centile. When we are estimating a central centile (i.e. that close to the 50th) the the data are close to being symmetric, and hence the CI is more symmetric as well (unless there is strong asymmetry in the BMI distribution itself). More extreme centiles are restricted by the data much more asymmetrically and hence, for them we can typically observe very asymmetric CI's.

Figure 11 about here

Conclusion

We estimate the trends by centiles as well as the rate of change of the trends in BMI values of US adults by birth cohorts stratified into four groups: white men, white women, black men and black women between c. 1896 and 1986. We find that the BMI values were increasing as far back as our data allow us to go, namely, the late 19th century. Moreover, the centiles shifted outward like the veins of a fan implying that the distribution became increasingly skewed to the right over time. The BMI values in the lowest centiles hardly increased at all. Even among black women, who were the most susceptible to the obesity epidemic, the lowest centile increased only 2 units during the whole period under consideration. However, the highest centiles increased by as much as 18-22 units in the four groups considered. After World War II the low centile BMI values were either stagnant or practically so and the only BMI values that increased rapidly were in the upper centiles. Consequently, the spread between the lowest and highest centiles practically tripled from approximately 8 to 25 BMI units in three of the groups while among black women the spread increased from 10 to 30 BMI units in the course of the period considered (Figures 1-4).

There was considerable variation over time in the rate of change of BMI centile curves. Among whites, both men and women, BMI values accelerated around the birth cohorts of the two World Wars and decelerated among those of the Great Depression. The rate of change differed markedly among blacks and whites in the first half of the century but became quite similar after mid-century. Among black men the rate of change slowed during the first half of the century and then accelerated after World War II, while among black women it remained constant at a high level until World War II when it accelerated as in the other groups. After the war the rate of change in BMI values of blacks came to resemble that of whites with a post-war acceleration followed by a substantial deceleration around the late 1960s.

In sum, the obesity epidemic is hardly the making of the last few decades of the 20th century as the conventional wisdom would have it. Our estimates indicate that the transition to post-industrial BMI values occurred gradually in the course of the 20th century and probably started much earlier than the consensus asserts, with black women outpacing the other three groups from the very beginning. Thus, the transition to a post-industrial lifestyle over time affected an increasing portion of the BMI distribution. Only the bottom two centiles managed to stay below overweight status among white men, white women, and black men, and among black women only the lowest centile escaped the grips of the creeping epidemic. This also implies that lifestyle changes of the 20th century affected various segments of the population quite differently and that 10-20% of the population was completely immune to it.

Identifying the causes of this long-run trend is outside of the scope of this study, but we do note that the persistently “creeping” nature of the epidemic does suggest that its roots were embedded deep in the social fabric, slowly changing as the population responded to a vast array of seemingly irresistible socio-economic and technological forces. The question still remains to explore why the various ethnic groups, genders and the different segments of the BMI distribution responded so differently to these forces impinging on the life-style of the population.

References

Cole TJ. 1988. Fitting smoothed centile curves to reference data. J Roy Stat Soc A, 151, 3, 385-418

Eubank RL. 1988. Spline smoothing and nonparametric regression. Marcel Dekker. New York.

Graybill FA. 1976. Theory and Application of the Linear Model. Wadsworth & Brooks–Cole. Pacific Grove.

Green PJ, Silverman BW. 1994. Nonparametric regression and generalized linear models. Chapman & Hall, London.

Hastie T, Tibshirani R. 1990. Generalized Additive Models. London: Chapman and Hall.

Komlos J, Brabec, M. 2010. The Trend of Mean BMI Values of US Adults, birth cohorts 1882-1986 indicates that the obesity epidemic began earlier than hitherto thought. *Am J of Hum Biol*, 22, forthcoming; NBER Working Paper no. 15862.

McCullagh P, Nelder JA. 1989. Generalized Linear Models. London: Chapman and Hall.

R 2010. Statistical software package, http://cran.at.r-project.org/

Rawlings JO. 1988. Applied regression analysis: A research tool. Wadsworth & Brooks Cole. Pacific Grove.

Rigby RA. Stasinopoulos DM. 2005. Generalized additive models for location, scale and shape (with discussion). Applied statistics 54, 3:507-554.

Rigby RA, Stasinopoulos DM. 2006. Using the Box-Cox *t* distribution in GAMLSS to model skewness and kurtosis. Statistical modeling 6:209-229.

Rigby RA, Stasinopoulos DM. 2007. Generalized additive models for location, scale and shape GAMLSS) in R. J of Stat Software, 23, 7:1-46.

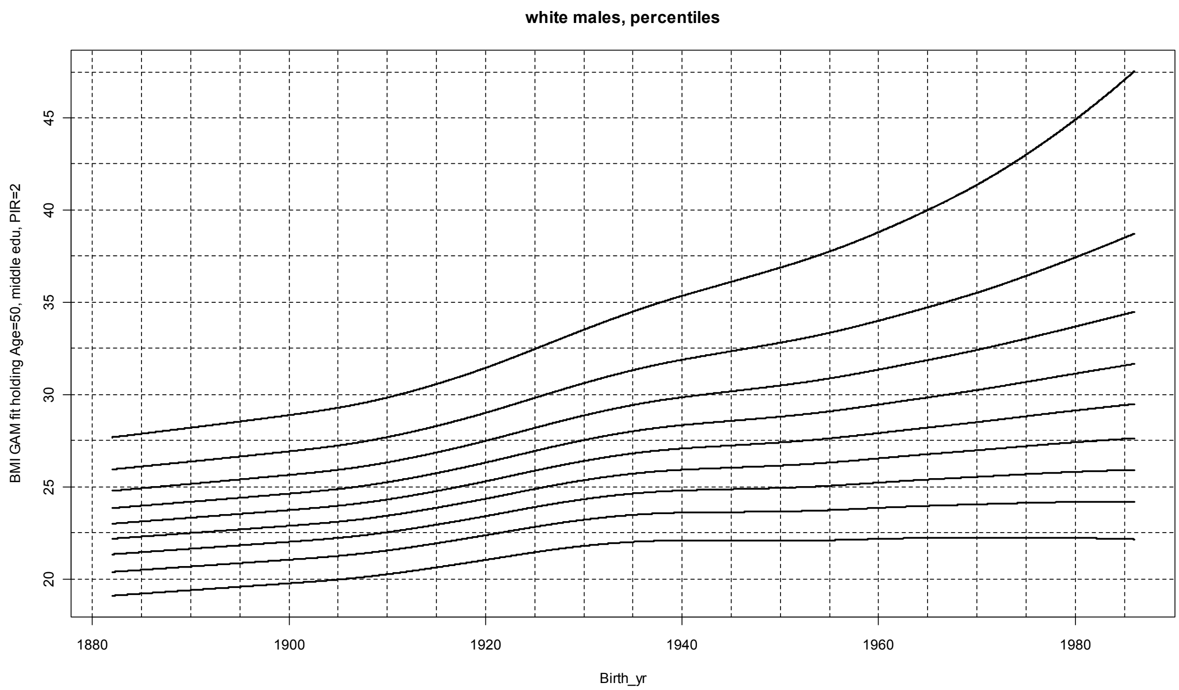
van Buuren S, Fredriks M. 2001. Worm plot: a simple diagnostic device for modelling growth reference curves. Stat in Med, 20:1259–1277.

Table 1. Dates by which given centile reached a BMI value of 30 (Birth Cohort)

	White		Black	
Centile	Males	Females	Males	Females
9th	1911	1912	1907	1897
8th	1926	1931	1924	1905
7th	1942	1946	1950	1917
6th	1967	1964	1962	1927
5th	na	1980	1982	1942
4th	na	na	na	1959

Note: Among white men and women, and black men, the 5th, 4th and 3rd centiles have not reached the BMI value of 30 during the observation period

Figure 1. Trend of BMI centile curves of US-born White Men by birth cohorts



Note: All figures show model (2) estimates evaluated at PIR = 2 at age 50 and with a High School Diploma.

Figure 2. Trend of BMI centile curves of US-born White Women by birth cohorts

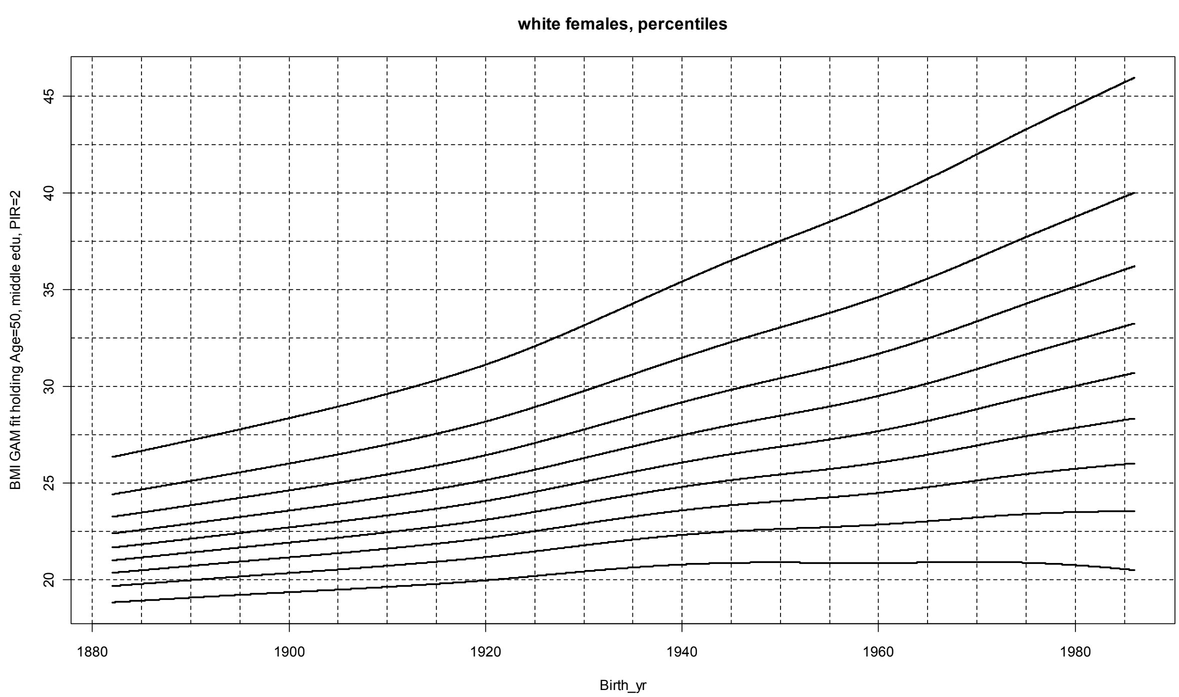


Figure 3. Trend of BMI centile curves of US-born Black Men by birth cohorts

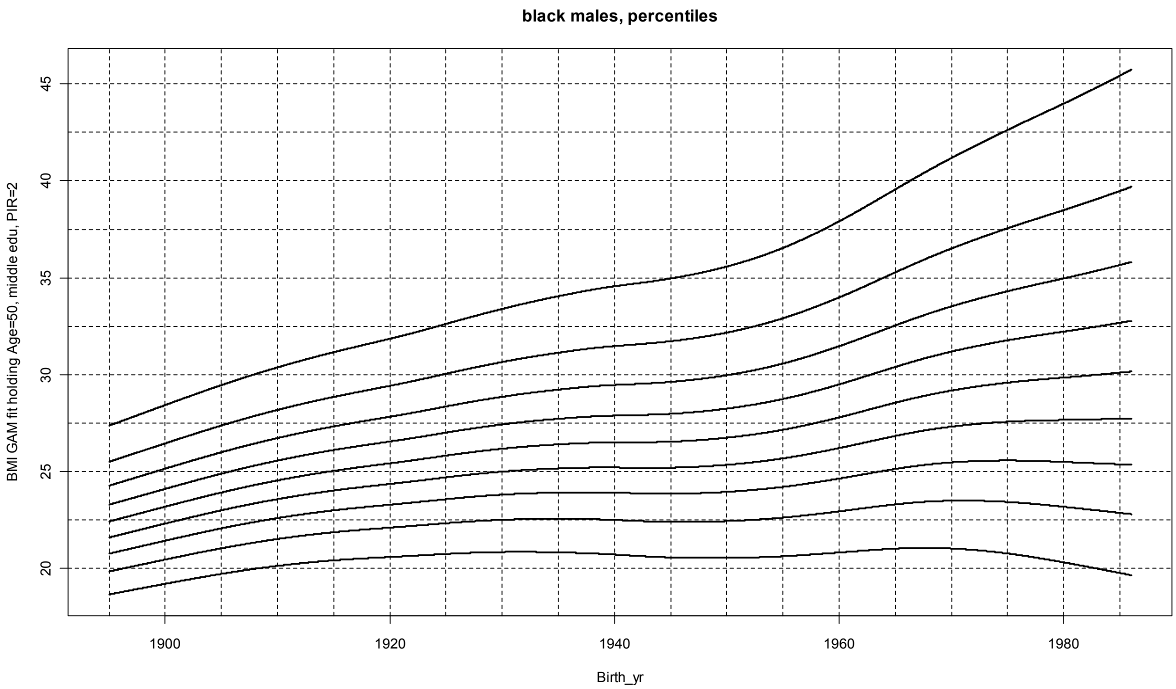


Figure 4. Trend of BMI centile curves of US-born Black Women by birth cohorts

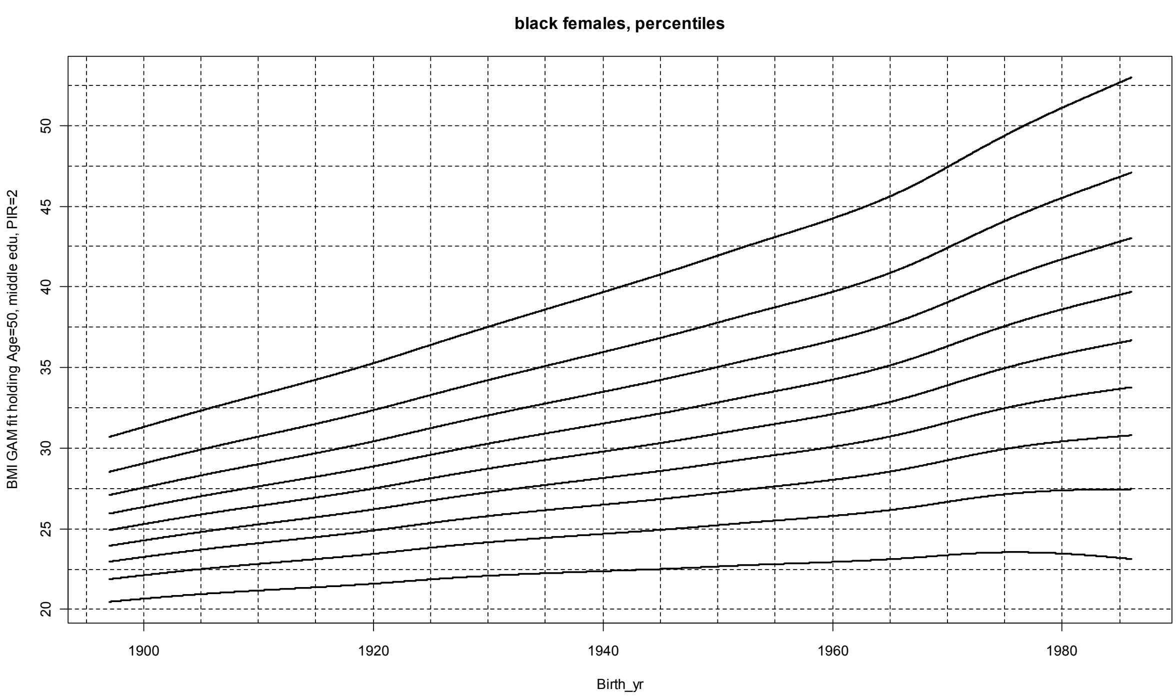


Figure 5. Variability of BMI Values over time, the σ function by sex and ethnic groups.

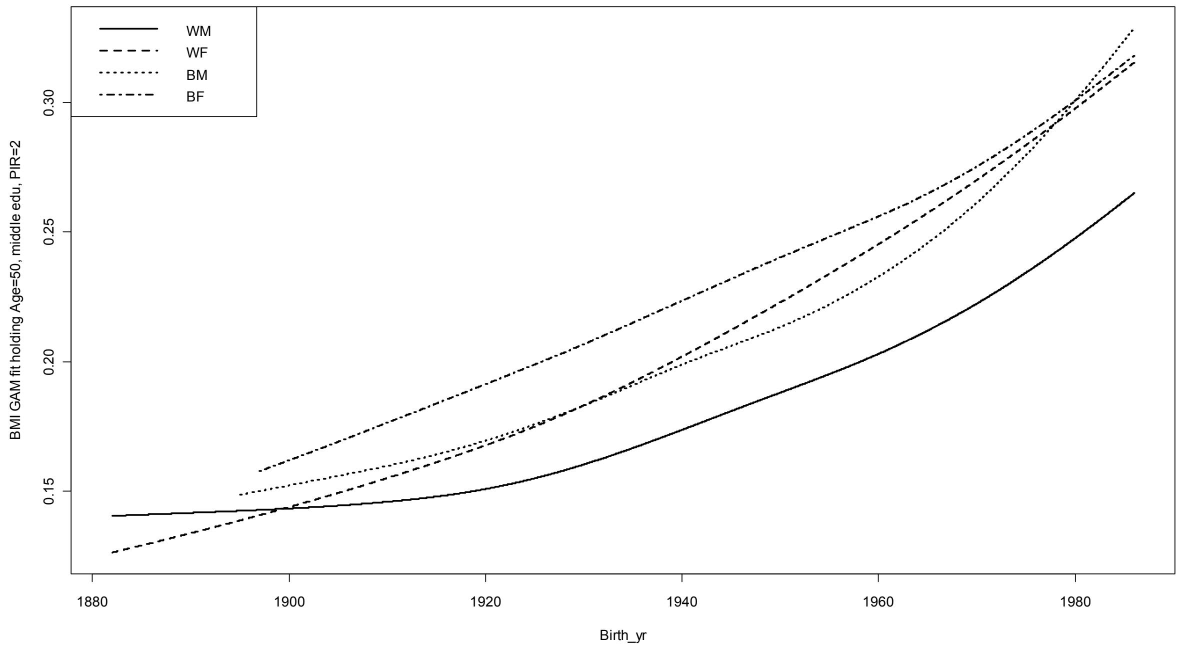
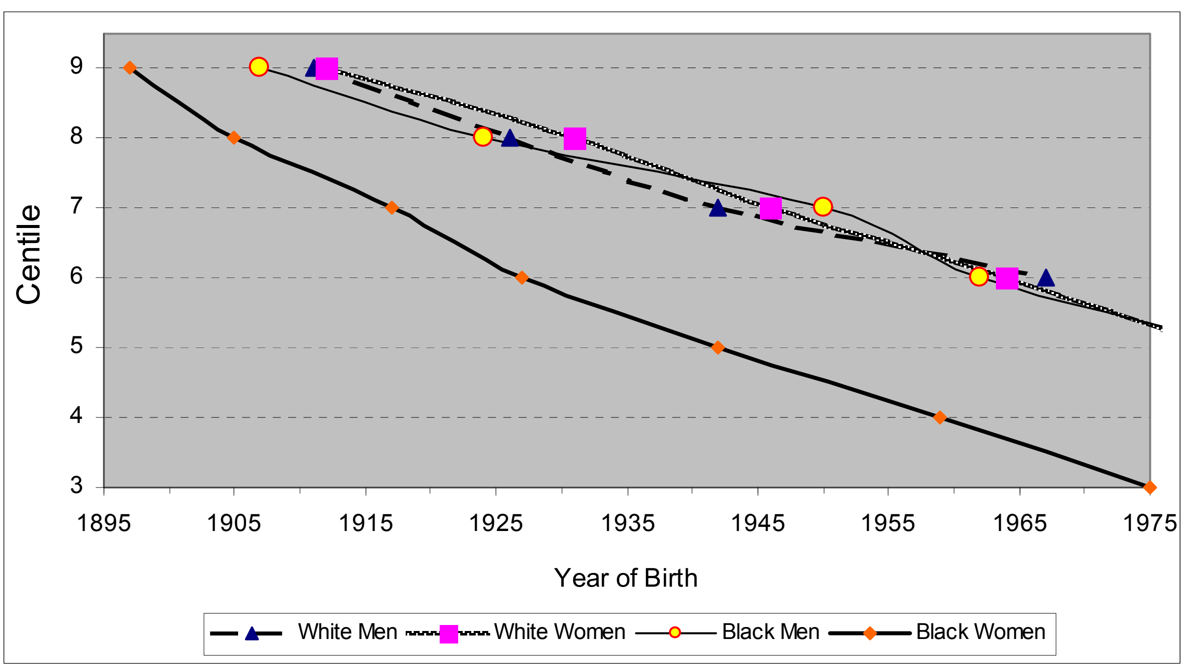


Figure 6. The dates by which given centile reached a mean BMI value of 30



Source: Table 1. Note: Among white men and women, and black men, the 5th, 4th and 3rd centiles have not reached the BMI value of 30 during the observation period.

Figure 7. Rate of Change of BMI centile curves of White Men by birth cohort in Figure 1

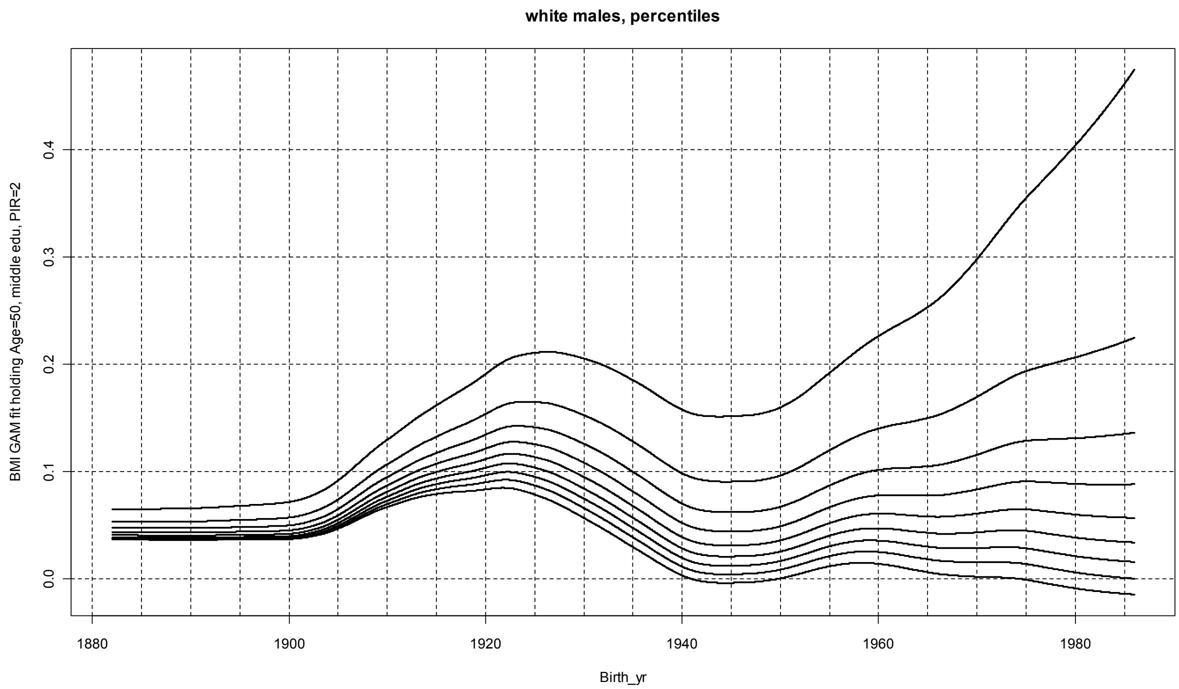


Figure 8. Rate of Change of BMI centile curves of White Females by birth cohort in Figure 2

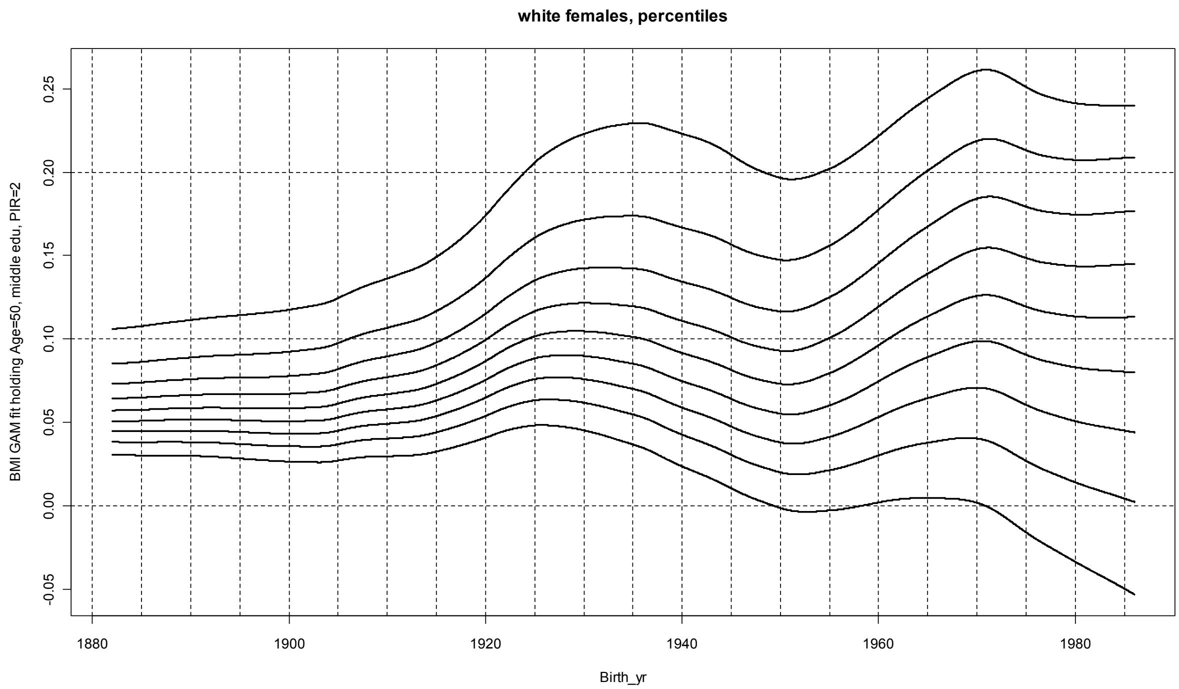


Figure 9. Rate of Change of BMI centile curves of Black Men by birth cohort in Figure 3

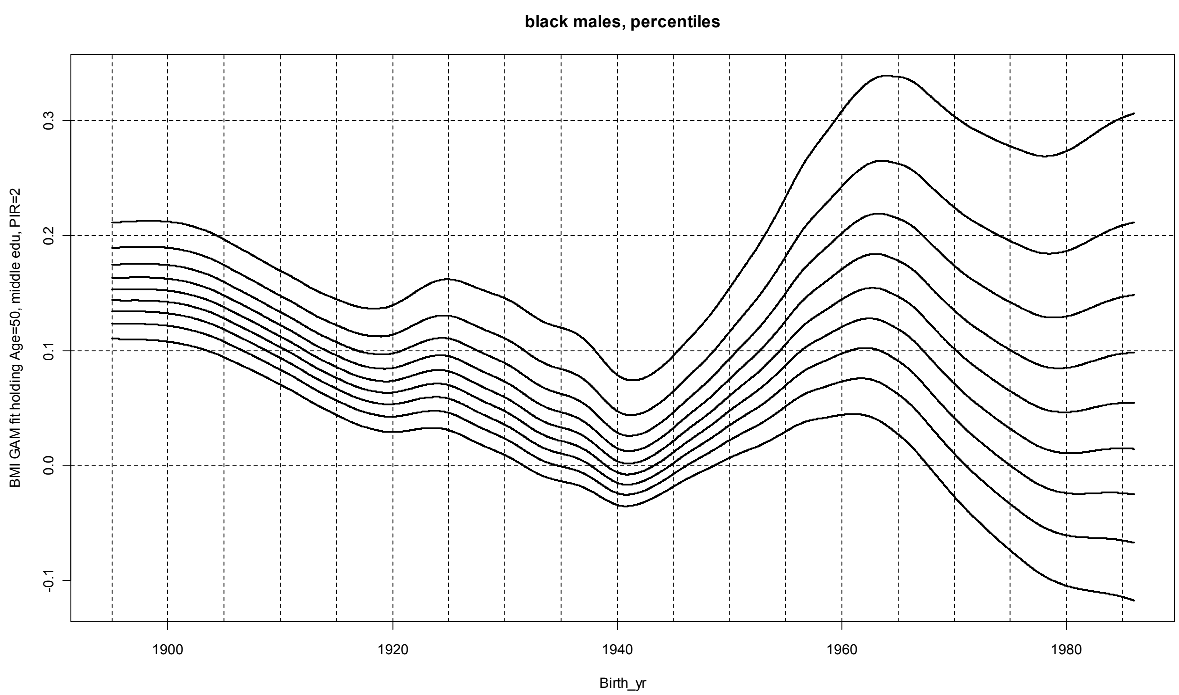


Figure 10. Rate of Change of BMI centile curves of Black Females by birth cohort in Figure 4

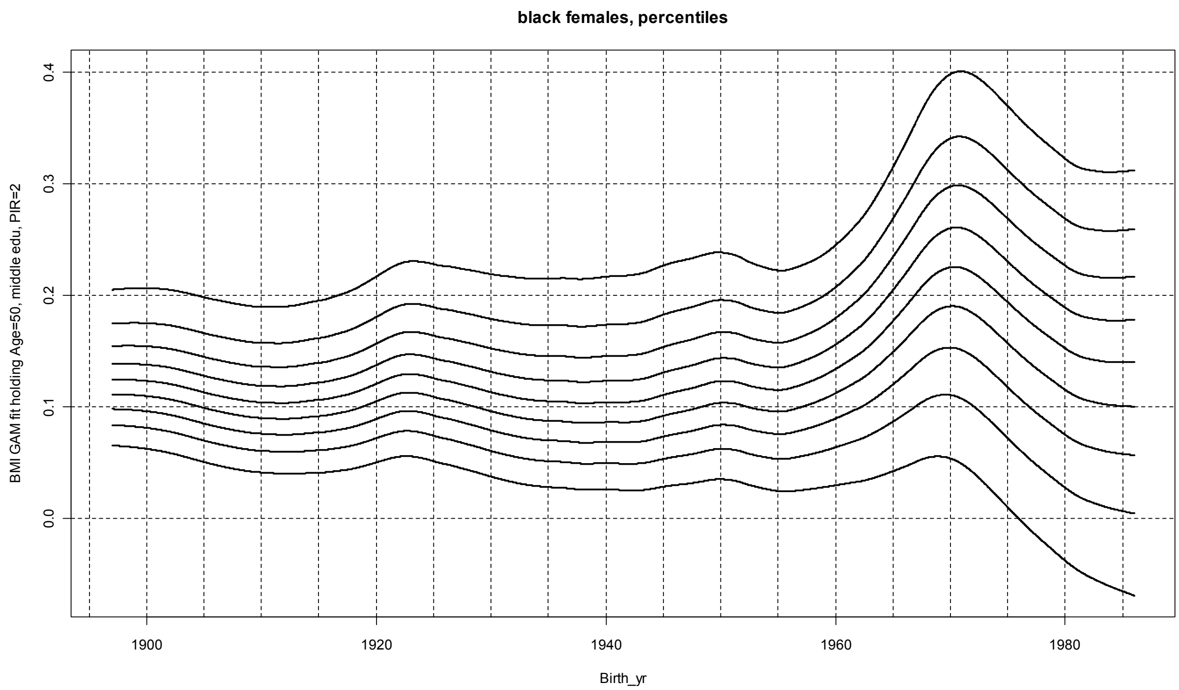
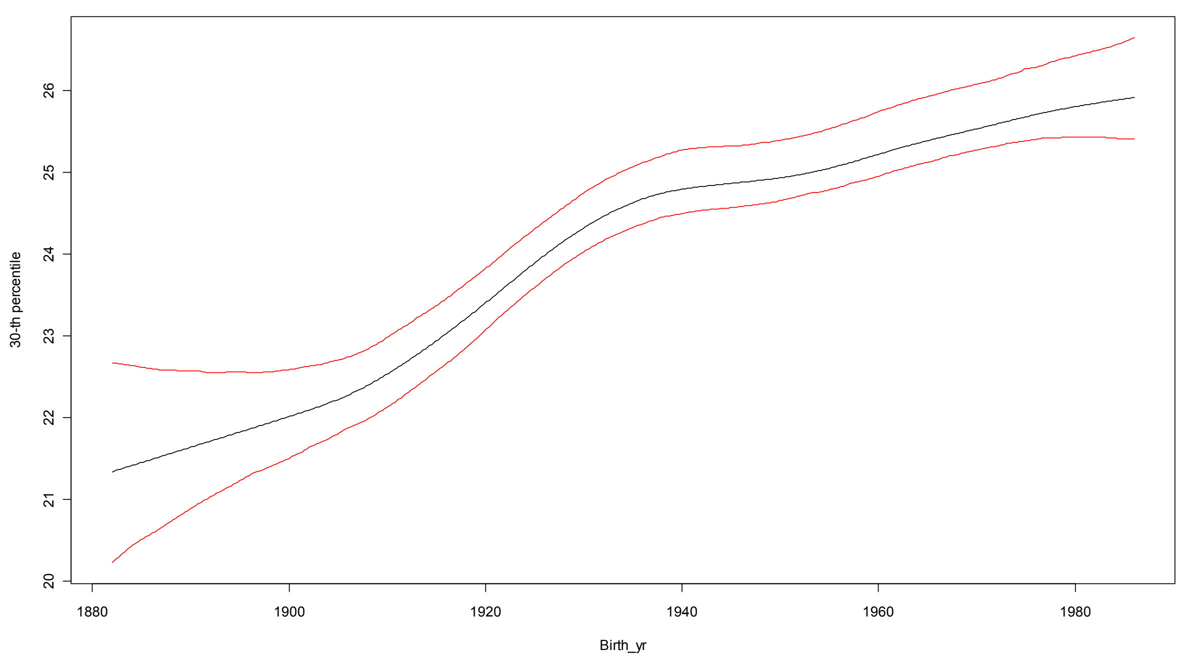


Figure 11. 95% confidence Intervals for BMI values of White Men in the 30th Percentile estimated by a bootstrap procedure.



Endnotes

[1] The LMS method is a Box-Cox transformation based spline smoothing with which median, coefficient of variation and Box-Cox transformation parameter are modeled as smooth functions of a covariate, using splines.

[2] Support is the closure of set where the density of the random variable of interest is positive.

[3] Sigma is related to the coefficient of variation, CV. Rigby and Stasinopoulos (2006) derive the following approximate formula: CV ≈ $\sigma[1 + 0.36/\tau]$. Nevertheless, the coefficient of variation is defined somewhat differently from what is used normally. Usually, one uses $CV = \frac{std.deviation}{mean}$. Here, one uses the so called centile-based coefficient of variation, namely: $CV = \frac{3.IQR}{4.median}$, where $IQR = Q_3 - Q_1$, interquartile range is the difference between third and first and quartiles of the distribution. One can consider it just another way of computing CV as a measure of relative variability, in which the mean is replaced by median and standard deviation by (appropriately scaled) interquartile range. The factor of ¾ comes from the fact that under normality, one needs such scaling to have an unbiased estimate of the standard deviation. In fact, under normality, to have unbiasedness, $\hat{\sigma} \cong 1.4826 MAD = \frac{1.4826 IQR}{2} \cong \frac{3}{4} IQR$.

[4] As in the case of the generalized linear models (McCullagh, Nelder 1989), here we deal with a model that is inherently nonlinear (in parameters). It is of relatively tame nonlinear class, however. Specifically, the linear predictor (i.e. linear combination of covariates or explanatory variables with unknown coefficients as parameters) does not model the μ, σ, ν or τ directly. Instead, it models its one-to-one function. The function is called a link.

[5] d's were selected separately for each cubic spline term in the model (1), based on the GAIC criterion described on the next page. Generally, the larger is the degree of freedom for a spline, the less smooth and more complex the spline function is.

[6] In particular, we do not use the centiles function built into the gamlss package, because we have several covariates in the model.

[7] CI's were estimated together over times and quantiles. To be preceise, we bootstrapped the model (actually the simplified model without weighting but with the same covariate structure

for μ,σ,ν,τ) 500 times. Each resample out of these 500 gives model parameters that allow for computation of all quantile curves for all times (and much more). Then we searched for 2.5 and 97.5 th percentiles over the 500 bootstrap resamples time point by time point, for each percentile (10, 20, ..., 90). This gives a sort of "envelope" band that has the property that it covers 95% percentile curves iver the bootstraps, for a given percentile.