ELSEVIER
Economics and Human Biology 8 (2002) 1–13
ECONOMICS AND HUMAN BIOLOGY

Non-parametric regression with BayesX: a flexible estimation of trends in human physical stature in 19th century America

Stefan Lang [a], Marco Sunder [b,*]

[a] *Department of Statistics, University of Munich, Munich 80539, Germany*
[b] *Institute of Economic History, University of Munich, Munich 80539, Germany*

Abstract

The recently developed computer program BayesX provides a Bayesian approach to the estimation of non-parametric additive models. Such models can be useful in applications when the effect of metrical covariates (such as time) are to be estimated while controlling for other factors. In an application of this methodology, trends in the height of West Point cadets in the 19th century are estimated. The results indicate that the biological standard of living of the "middle class" increased relative to the rest of the American society during the Antebellum years.

JEL classification: C5; N31

Keywords: Non-parametric regression; BayesX; Middle class

1. Introduction

We introduce a recently developed software for the estimation of flexible additive models within a Bayesian framework. The focus is on models in which a usual OLS regression could be used and in which some of the covariates are metrical. A typical example is a time trend. The OLS model imposes linearity of the influence of independent variables which might be too restrictive in many cases. There are several ways to relax this assumption, including specification with dummy variables coding distinct intervals of the variable. However, this specification entails arbitrary time interval lengths and the estimated variances of the regression parameters are often quite large due to an overparameterization of the model. An alternative is to use non-parametric methods, such as penalized least squares (Hastie and

* Corresponding author. Tel.: +49-89-2180-2754.
E-mail address: macro@econhist.de (M. Sunder).

PII: S1570-677X(02)00009-6

Tibshirani, 1990). The impact of metrical covariates can be non-linear in such a framework, but the effects are assumed to be smooth, thereby avoiding the problem of large variances described above. An overview of non-parametric regression can be found in Fahrmeir and Tutz (2001). In this paper, we describe a Bayesian approach to non-parametric regression and an easy-to-use implementation of these techniques in the software package BayesX. The freeware program can be downloaded for Windows at http://www.stat.uni-muenchen.de/∼lang. We illustrate the approach and the use of BayesX with an application to human stature in 19th century America. In this application, the height of West Point cadets is used as an indicator of net-nutritional attainment during childhood and adolescence. A time trend, the effect of age and other covariates on the height of the cadets is estimated using BayesX.

2. Methodological background

2.1. Penalized regression

Suppose we have given observations $y_i, x_i,\ i = 1, \dots, n$, of a metrical response variable y (in our case the height of a cadet) and a metrical explanatory variable x (e.g. the age of the cadet or the calendar time). Traditionally, the effect of x is modeled using a Gaussian linear regression model, i.e.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \tag{1}$$

where the errors are assumed to be independent and Gaussian $\epsilon_i \sim N(0, \sigma^2)$ with a common variance σ^2 across subjects. A requirement is that a linear relationship between y and x is reasonable. In many historical applications, the effect of x is modeled by a set of dummy variables $x_{ij},\ j = 1, \dots, J$ (e.g. yearly dummies if x corresponds to calendar time) to take possible non-linearities into account. This leads to a linear regression model of the form

$$y_i = \beta_1 x_{i1} + \cdots + \beta_J x_{iJ} + \epsilon_i, \tag{2}$$

where β_j is the regression parameter of the jth dummy. The model can be estimated by minimizing the residual sum of squares

$$S(\beta) = \sum_{i=1}^{n} (y_i - \beta_1 x_{i1} - \cdots - \beta_J x_{iJ})^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{J} \beta_j x_{ij} \right)^2, \tag{3}$$

with respect to $\beta = (\beta_1, \dots, \beta_J)$. The approach, however, suffers from dramatically increased variances for the estimated parameters due to the overparameterization of the model. To regularize the problem (i.e. to decrease the variances), a common approach is to replace ordinary least squares (3) by *penalized least squares* where strong jumps between neighboring regression parameters are penalized (e.g. parameters of age dummies if x is covariate age). Possible penalizations are given for example by either

$$S_1(\beta) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{J} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=2}^{J} (\beta_j - \beta_{j-1})^2 \rightarrow \min_{\beta}, \tag{4}$$

or

$$S_2(\beta) = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{J}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=3}^{J}((\beta_j - \beta_{j-1}) - (\beta_{j-1} - \beta_{j-2}))^2 \to \min_{\beta}, \quad (5)$$

The additional terms introduced are first- (4) or second-order (5) *difference penalties*. First-order differences penalize abrupt jumps $\beta_j - \beta_{j-1}$ between successive regression parameters and second-order differences penalize deviations from the linear trend $2\beta_j - \beta_{j-2}$. The trade-off between fidelity to the data and smoothness is controlled by the *smoothing parameter* $\lambda > 0$. Small values of λ give large weight to the first terms in (4) and (5). Accordingly, large differences between neighboring parameters are allowed. In the limit ($\lambda \to 0$) the data are interpolated and ordinary least squares is obtained as a special case. Large values of λ give large weight to the penalty terms in (4) and (5) and only small jumps between neighboring parameters are allowed.[1]

We follow Bayesian[2] versions of the penalized least squares approaches (4) and (5), described in detail in Fahrmeir and Lang (2001a,b). A Bayesian approach has several advantages over the specification (4) or (5). For instance, the amount of smoothness controlled by λ can be estimated *simultaneously* with the regression coefficients β_j which is usually quite difficult within the traditional frequentist methodology. In the Bayesian approach, all unknown parameters are assumed to be stochastic and appropriate prior distributions must be specified. The direct stochastic analogue to the penalty term in (4) is a first-order random walk for the regression coefficients, i.e.

$$\beta_j = \beta_{j-1} + u_j, \quad j = 2, \ldots, J, \quad (6)$$

with $u_j \sim N(0, \tau^2)$. For the initial value β_1 we assume a diffuse prior, i.e. $\beta_j \propto$ const. The analogue to the penalty term in (5) is a second-order random walk, i.e.

$$\beta_j = 2\beta_{j-1} - \beta_{j-2} + u_j, \quad j = 3, \ldots, J, \quad (7)$$

with $u_j \sim N(0, \tau^2)$ and diffuse priors for the initial values β_1 and β_2. The analogue to the smoothing parameter λ in (4) and (5) is the variance parameter τ^2. More (less) smoothness is obtained with decreasing (increasing) variance τ^2. To be able to estimate the amount of smoothing (i.e. τ^2) simultaneously with the regression parameters an additional prior $p(\tau^2)$ is specified for τ^2. Mainly for mathematical simplicity, the conjugate prior[3] for τ^2 is usually assumed. It is an inverse gamma distribution, i.e. $\tau^2 \propto IG(a, b)$ with fixed (non-stochastic) hyperparameters a and b.[4] A possible choice is $a = 1$ and $b = 0.005$ or $a = 0.001$ and $b = 0.001$ resulting in relatively non-informative priors for τ^2. Bayesian inference for the unknown parameters β and τ^2 is based on the posterior distribution $p(\beta, \tau^2|y)$. According

[1] In the limit ($\lambda \to \infty$), estimated parameters β_j are all equal, i.e. $\beta_j = c$, if (4) is used. If (5) is used the parameters β_j follow a straight line.

[2] For a comprehensive introduction to Bayesian estimation and inference written for social scientists, see Simon Jackman's website: http://jackman.stanford.edu/mcmc.

[3] A prior is conjugate if the posterior follows the same distribution family as the prior.

[4] The χ^2 distribution is a special case of the gamma distribution.

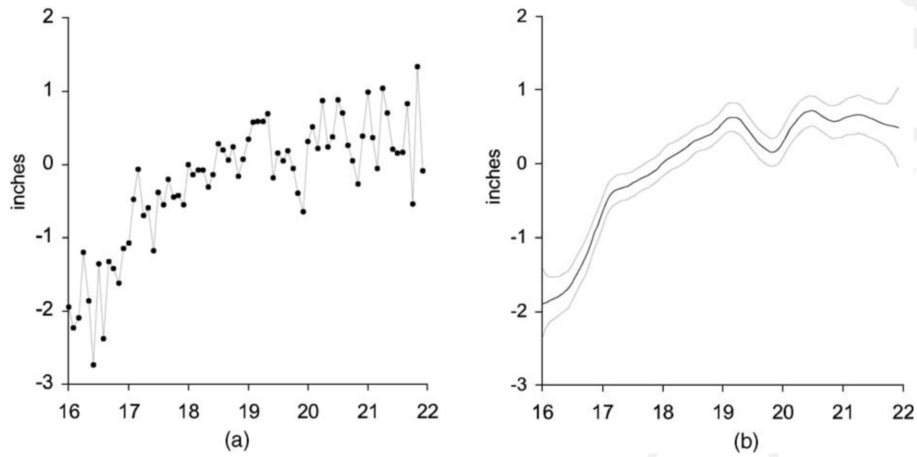

Fig. 1. Estimated effects of cadets age on height (in.), relative to 18-year-old cadets, based on an ordinary dummy variables approach (a) and penalized dummies (b).

to Bayes' theorem it is given by

$$p\left(\beta, \tau^2|y\right) = cL(y|\beta)\,p(\beta|\tau^2)\,p\tau^2, \tag{8}$$

where c is a normalizing constant and $L(y|\beta)$ is the likelihood. Because of independence assumptions about the errors ϵ_i in (1), the likelihood is simply a product of normal densities. Point estimators for β are obtained by the posterior mean $E(\beta|y)$ or the posterior mode. Credible intervals (the Bayesian analogue to confidence intervals) for the parameters β_j with a nominal level of $1 - \alpha$ are obtained by computing their posterior $\alpha/2$ and $1 - \alpha/2$ quantiles.

In many practical situations, however, the posterior distribution is numerically intractable. A common technique to overcome these problems are Markov Chain Monte Carlo (MCMC) simulation methods. These allow the drawing of *random numbers* from the numerically intractable posterior distribution and in this way, the estimation of characteristics of the posterior like means, standard deviations or quantiles via their *empirical analogies*. The main idea is that instead of drawing directly from the posterior (which is impossible in most cases anyway) a Markov Chain is created, whose iterations of the transition kernel coverage to the posterior distribution. In this way a sample of dependent random numbers of the posterior is obtained. As a rule, the first drawings from this sample of parameter values is discarded to take into account the time the algorithm needs for convergence to the posterior. This part is known as burn-in period.[5]

To demonstrate the usefulness of the penalization, we contrast in Fig. 1 the estimated effect of the age on the cadets' height based on a simple dummy variable approach (Fig. 1a) and on (Bayesian) penalized dummies (Fig. 1b). We see that the penalization results in much more stable estimates, which are also easier to interpret. However, the sudden drop of the effect at

[5] A nice introduction to MCMC methods can be found in Green (2001).

the age of 20 years is somewhat unreliable because we would expect a monotonic increase of height with age. This comes about obviously due to the small number of observations after age 20 (the number of observations is only 10–44 per month). We will investigate further improvements of our approach in the next section.

2.2. Penalized splines

In the previous section possible non-linearities of the effect of a covariate x are modeled with a dummy variable approach. If we think of the effect of x as a (non-linear) function f of x we can write (2) as

$$y_i = f(x_i) + \epsilon_i, \tag{9}$$

where $f(x_i) = \beta_1 x_{i1} + \cdots + \beta_J x_{iJ}$ is piecewise constant. In this section, we generalize our approach by assuming more general functions f. More specifically, we assume that f is a *polynomial spline*. Suppose that the range of x is divided into non-overlapping intervals with equal length through $r + l$ cutpoints ζ_j with $x_{\min} = \zeta_0 < \zeta_1 < \cdots < \zeta_{r-1} < \zeta_r = x_{\max}$. A polynomial spline of degree d with respect to the cutpoints ζ_j is a function f with the following properties:

- On each of the intervals $(\zeta_0, \zeta_1), \ldots, (\zeta_{r-1}, \zeta_r)$, f is a polynomial of degree d.
- At the cutpoints ζ_j the spline f is $d - 1$ times continuously differentiable.

Usually, the cutpoints ζ_j are called the knots of the spline.

Every spline can be written in terms of a linear combination of J basis functions $B_j(x)$ spanning the spline space,[6] i.e.

$$f(x)=\beta_1 B_1(x) + \cdots + \beta_J B_J(x),$$

and we can replace (9) by

$$y_i = \beta_1 B_1(x_i) + \cdots + \beta_J B_J(x_i) + \epsilon_i. \tag{10}$$

With $d = 0$ we obtain the dummy variable approach from the previous section as a special case. Here, the basis functions are given by $B_j(x_i) = x_{ij}$. Model (10) can be estimated by ordinary least squares. However, the choice of the *number of knots* is crucial. For a small number of knots the resulting spline space may be not flexible enough to capture the variability of the data. For a large number of knots, estimated curves may tend to overfit the data. As a solution to these problems, we follow Eilers and Marx (1996) and Lang and Brezger (in press) who suggest a moderately large number of knots (usually between 20 and 40) to ensure enough flexibility. In complete analogy to the dummy variable approach of the previous section we define a roughness penalty based on differences of adjacent regression coefficients to guarantee sufficient smoothness of the fitted curves.

[6] Splines represent a finite dimensional vector space. Hence, every spline can be written in terms of a finite set of basis vectors, which are functions in our case. The basis functions are not unique. For numerical reasons and to make sure that the penalization of the regression coefficients is meaningful, we use a local B-spline basis. A comprehensive treatment of polynomial splines is in De Boor (1978), an easy to read introduction is given in Green and Silverman (1994).

As an example, compare Fig. 2a which shows the estimated effect of age on the cadets' height based on a P(enalized) spline of degree 3 with 20 knots and a second-order random walk penalty. Compared to penalized dummies in Fig. 1b, we obtain a smoother estimate where the buckle around age 20 years has disappeared almost completely. Note that estimated effects are usually more or less unaffected by varying the number of knots provided that there are enough knots.

2.3. Additive and varying coefficient models

So far, we have considered only one metrical covariate *x*. In many applications, however, two or more metrical covariates are considered. We demonstrate possible extensions of the simple model with an example on the physical stature of West point cadets. The effect we are most interested in is a possible time trend in the height of the cadets. In an *additive model* (Hastie and Tibshirani (1990)) with the two covariates 'time' and 'age' we assume that the effect of both covariates is additively composed of two (non-linear) functions f_1(time) and f_2(age), i.e. the height of the *i*th cadet is modeled by

$$\text{height}_i = f_1(\text{time}_i) + f_2(\text{age}_i) + \epsilon_i. \tag{11}$$

Similar to the previous section, we can assume polynomial splines for the two functions f_1 and f_2 and penalize the regression coefficients to prevent overfitting. If an additional vector of *q* categorical covariates (dummy variables) w exists, we can easily extend our model by assuming usual linear effects on the heights and we obtain

$$\text{height}_i = f_1(\text{time}_i) + f_2(\text{age}_i) + \gamma' w_i + \epsilon_i. \tag{12}$$

where $\gamma = (\gamma_1, \ldots, \gamma_q)$ is a vector of further regression coefficients.

The West point cadets are divided into three occupational groups: farmers, cadets of the "middle class", and others. It is reasonable to assume *different* time trends for the three occupational groups. Defining the dummy variables farmer_i and middleclass_i, we obtain the *varying coefficient* model (Hastie and Tibshirani, 1993)

$$\begin{aligned} \text{height}_i = {} & f_1(\text{time}_i) + f_2(\text{time}_i) \times \text{farmer}_i + f_3(\text{time}_i) \times \text{middleclass}_i \\ & + f_4(\text{age}_i) + \gamma' w_i + \epsilon_i. \end{aligned} \tag{13}$$

In this model, the function f_i(time) captures the time trend for the occupational group "others". The functions f_2(time) and f_3 (time) are deviations from the time trend of the group "others". Hence, $f_1 + f_2$ corresponds to the time trend for the farmers and $f_1 + f_3$ to the trend for the middle class. The model is called a varying coefficient model because the effects of the dummy variables farmer and middleclass_i vary smoothly over the course of the covariate time.

Bayesian inference for additive or varying coefficient models is done by MCMC methods in a similar way as described in Section 2.1. Details can be found in Fahrmeir and Lang (2001a,b); Lang and Brezger (in press) and in Brezger et al. (2002).

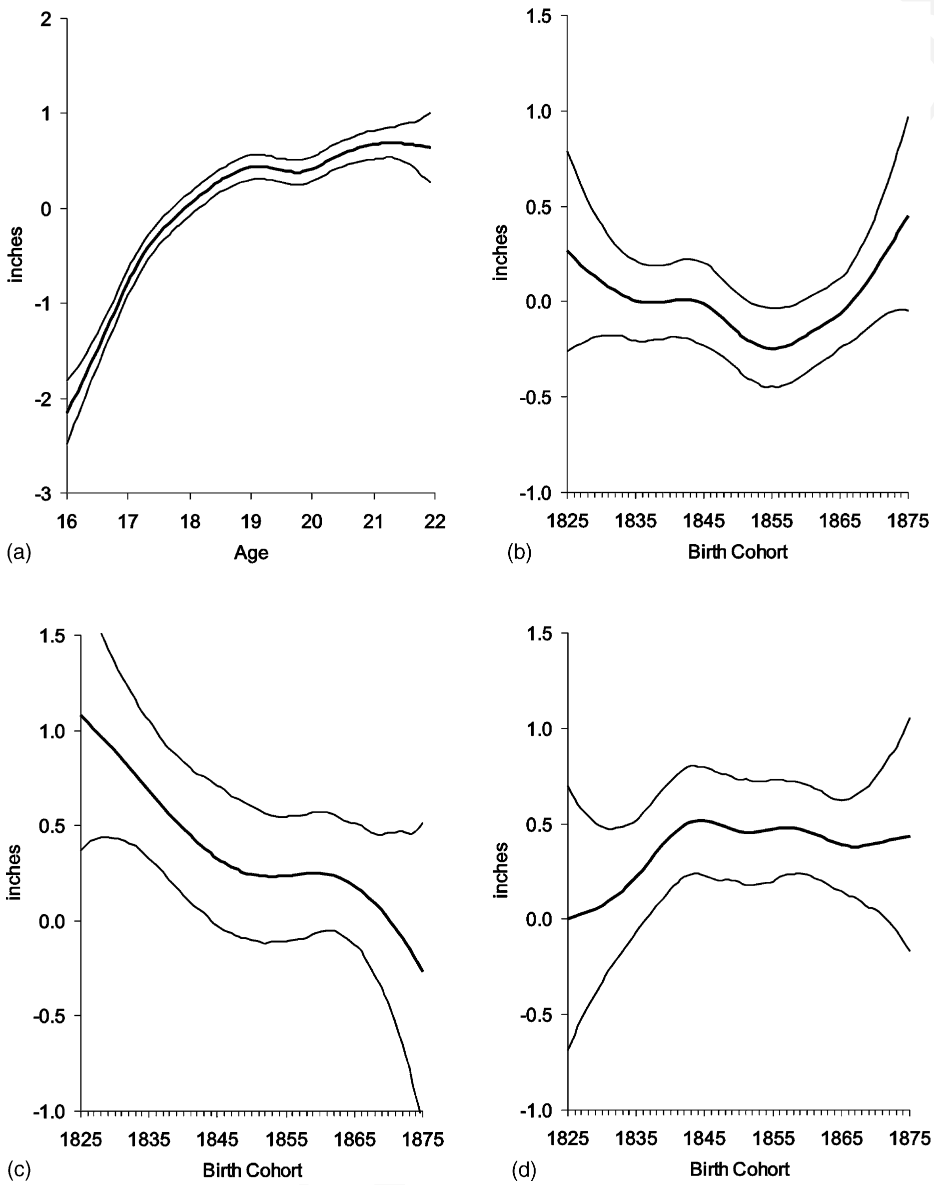

Fig. 2. Estimation results for non-linear functions. Depicted are the mean and the 10th/90th percentile: (a) age effect; (b) baseline effect; (c) farmer effect; (d) middle class effect.

2.4. Modeling spatial heterogeneity

In many applications, observations pertain to different locations. For instance, for the West Point cadets we know the state where they were born. In such situations, it may be necessary to control for possible spatial heterogeneity caused by unobserved location specific covariates, e.g. different economic conditions. BayesX allows to take spatial heterogeneity into account by a *mixed model* approach (Fahrmeir and Tutz, 2001). Thereby, the strategy is similar to the smoothing techniques described above. In analogy to the dummy variable approach for metrical covariates, we estimate for every location $l \in \{1, \dots, L\}$ (e.g. every state in the West Point example) one parameter b_l and penalize parameters to prevent overfitting. A possible assumption is that the location specific parameters b_l are independent and Gaussian with a common variance

$$b_l \sim N(0, \upsilon^2),$$

implying only weak smoothing conditions. Because of the common variance υ^2, parameters are shrunk towards zero. For the variance υ^2 we assume, similar to random walks, an inverse gamma prior. Much stronger smoothing is implied if we assume that the location specific parameters are spatially correlated, i.e. that neighboring parameters are more alike than others. A prior for the parameters can be defined by generalizations of one dimensional random walks to two dimensions. Such a prior is called a Markov random field (Besag et al., 1991) and is also supported by BayesX. Details on non-parametric regression models with spatially correlated effects and examples are in Fahrmeir and Lang (2001a) and Lang and Brezger (in press).

3. An application to the heights of West Point cadets

3.1. The data and the model

We provide a brief demonstration of the methods described above by revisiting the physical stature of cadets of the West Point academy in the 19th century.[7] These data were first analyzed by Komlos (1987, 1996) and Cuff (1993),[8] and they provide one of the pillars of the "Antebellum Puzzle:" as human height is an indicator of net-nutritional attainment during childhood and adolescence (Steckel, 1995), it is astounding that a decline in physical stature of the (non-slave) American male population occurred at a time of increasing per capita incomes (Komlos, 1998). Other evidence is provided by data on Union Army soldiers (Margo and Steckel, 1983; A'Hearn, 1998; Haines et al., 2000; Lauderdale and Rathouz, 1999; Cuff, 1998). Similar patterns have been documented among European societies at the same time. The explanation suggested by Komlos (1987, 1998) includes the adverse effects of urbanization that accompanied industrialization. The number of city dwellers who statistically depended on a farmer for nutrition was rising faster than productivity in agriculture.

[7] Data are available from ICPSR data archive (http://www.icpsr.umich.edu), data set no. 9468.

[8] Woitek (in press) analyzed the time series properties of this sample.

Many farmers switched from self-sufficiency to serving the new urban markets. Prior to refrigerating however, some foodstuffs could not be transported over large distances, e.g. skimmed milk as a by-product of making cheese. With an inelastic world supply of food, this led to an increase in the (relative) price of food compared to other goods. To be sure, Antebellum towns additionally suffered from deficient sanitation and a demanding epidemio logical environment (Costa and Steckel, 1997).

The West Point sample comprises approximately 4200 cadets who entered the academy between 1843 and 1894. We include in our analysis only individuals born in the US between 1825 and 1875 and aged 16–21 years. Of the remaining 3973 records, information on 2721 persons was matched with characteristics on family background (Komlos, 1987, p. 899). Our analysis is restricted to this matched subsample which has information on the size of the place of residence and on father's occupation, thus permitting a rough social stratification.[9] We distinguish between sons of farmers, cadets of the "middle class", and others, mainly sons of blue collar workers (Komlos, 1996, p. 204). Based on the deliberations of the previous section we estimate the model

$$\begin{aligned}\text{height}_i = \gamma_0 + f_i(\text{time}_i) + f_2(\text{time}_i) \times \text{farmer}_i + f_3(\text{time}_i) \times \text{middleclass}_i \\ + f_4(\text{age}_i) + \gamma_1 \times \text{urban}_i + b_{\text{state}_i} + \epsilon_i. \qquad (14)\end{aligned}$$

The functions f_1–f_4 are assumed to be smooth, and we choose P-splines of degree 3 and a second-order random walk penalty to approximate them. In addition, spatial effects are considered by including a dummy variable for urban residence and uncorrelated spatial effects b_{state_i} that pertain to the state where the ith cadet was born.

3.2. Estimation in BayesX

BayesX is command line oriented. To estimate the model (14) in BayesX we either have to write a batch file containing all necessary commands or the commands could be entered and executed directly in BayesX. In our example, the batch file consists of the following statements: *dataset w*

w.infile using c:\data\westpoint.raw bayesreg b
*b.regress height = time (psplinerw2) + farmer*time(psplinerw2)*
*+ middleclass*time (psolinerw2) + age (psplinerw2) + urban + state (random),*
family = gaussian iterations = 52000 *burnin* = 2000 *using w.*

With the first two statements we read in the data. The next two statements are used to specify the model to be estimated and the number of MCMC iterations the algorithm should perform, including the burn-in stage. The batch file is executed by typing '*useful filename*' in BayesX. The program also provides several possibilities to visualize estimation results, details can be found in the user manual (Brezger et al., 2002).

[9] When estimating a model for an average cohort trend only, the restricted sample yielded a similar pattern as the full sample, hence we are confident that limiting the scope to the matched subsample does not impose severe selectivity issues on its own.

Table 1
Estimated constant effects (in.)

	10th percentile	Mean	90th percentile
Intercept	66.85	67.08	67.31
Urban	−0.505	−0.323	−0.133

Intercept pertains to 18-year-old cadets.

3.3. Results

Table 1 and Figs. 2 and 3 provide estimation results for our model (14). For the ease of interpretation, the implied cohort trends are depicted in Fig. 4. Urban cadets were significantly shorter on average with a point estimate of −0.32 in. (Table 1). Such an urban penalty is consistent with the literature on anthropometric history in the US before the 20th century. The trends in stature of the three occupational strata are very different. At the beginning of the period under consideration, farmers were the tallest. The notion that “propinquity to the source of food” was conducive to human stature is indeed well established in the literature (Komlos, in press). However, this advantage declined subsequently. The increasing *relative* price of food and the decline in transportation costs, especially with the construction of the railroad network, could have induced farmers to trade away larger shares of their produce. The middle class cadets, on the other hand, gained considerably in the late 1830s, just at the time when the height of the rest of the society was beginning to

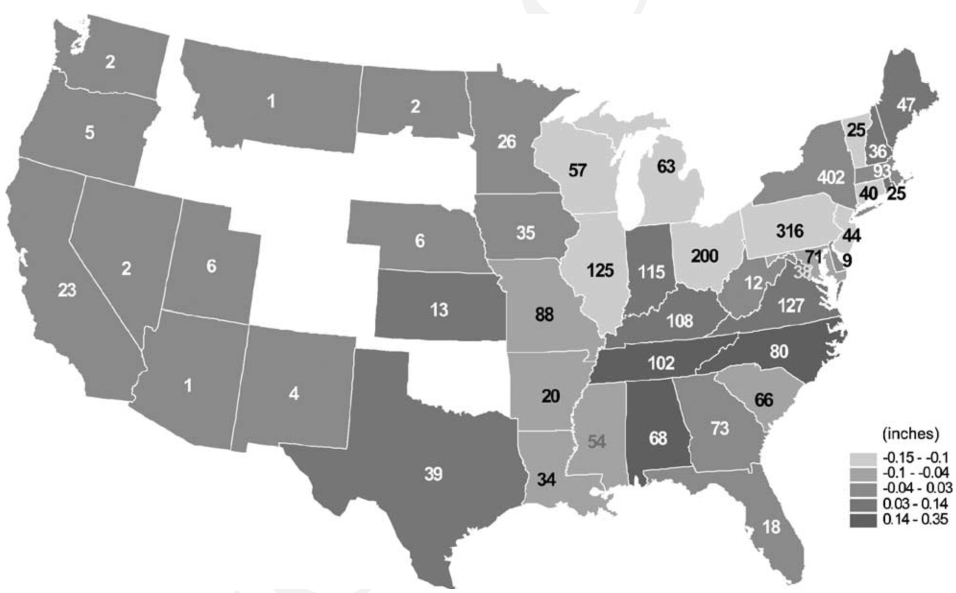

Fig. 3. Birth state random effects on a map of 1880. Numbers refer to case count.

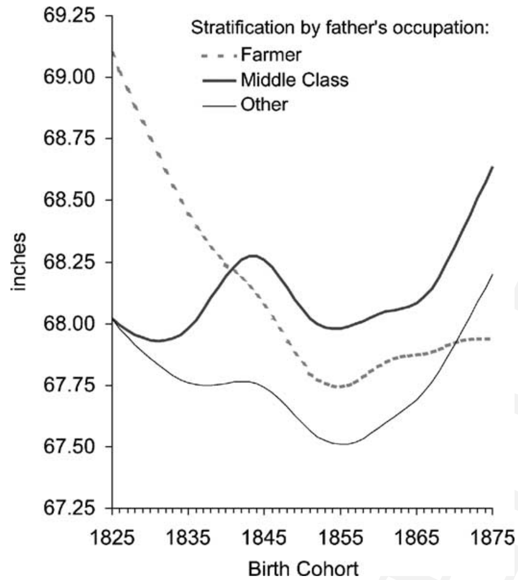

Fig. 4. Predicted trends in height, calibrated for 21-year-old rural cadets.

decline. In other words, there is no evidence of the "Antebellum Puzzle" pertaining to the middle class. The height advantage of the middle class was largest between 1844 and 1860. Heights did decline among the cohorts experiencing the impact of Civil War on the cost of living during their adolescence. This cycle in heights does, not differ from the other groups, but those from the middle class maintained the lead that they had attained in the late 1830s (Fig. 4).

We also estimated a model confined to 1903 urban (non-farmer) cadets (of whom 1430 were middle class). The middle class height advantage (Fig. 5) was positive and became significant in 1838 (after the economic recession of 1837) until the end of Civil War. The increasing height of the urban middle class at a time of increasing inequality (Lindert, 1991), when other segments of the society were becoming shorter, supports the view that the Antebellum Puzzle was brought about by changing economic circumstances (Lauderdale and Rathouz, 1999; Woitek, in press; Komlos, 1987).[10] Had the phenomenon been caused solely by a worsening disease environment, the height of the middle classes should have been affected as well. We can infer that the increase in middle class incomes must have been large enough to compensate for the rising relative prices of (protein-rich) foods.[11]

[10] The West Point sample is the only one so far that permits a closer investigation into the biological standard of living of the middle classes in 19th century America.

[11] In addition, if the disease environment had worsened, then the inference is warranted that the increase in their income was sufficiently large to increase their nutrient intake and thereby compensate for the rise in the demands of the disease environment.

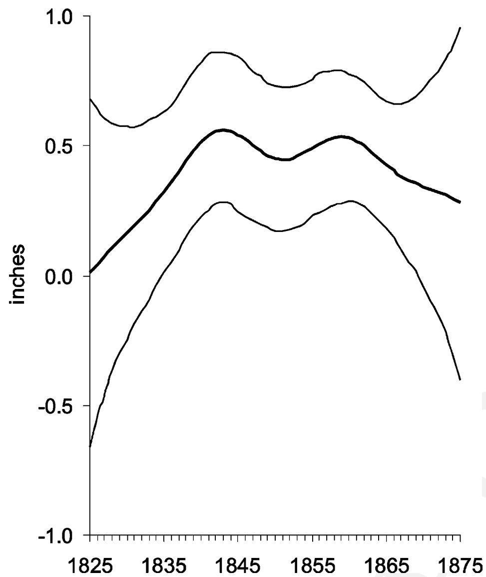

Fig. 5. Height "premium" of the urban middle class relative to urban non-middle class/non-farmer cadets (from a separate regression): mean estimate and 10th/90th percentile.

4. Conclusion

In many empirical regression applications, metrical covariates can exert non-linear effects. While it is possible to implement non-linearity into traditional linear models, more flexible approaches such as the additive models framework are quite attractive, since the progress in the speed of computing has expanded the practical limits considerably. The Bayesian approach to additive models described here has the advantage that it does not rely on cross-validation of models with regard to different combinations of smoothness parameters, as in traditional frequentist implementations. Thus, it is possible—given enough data—to estimate several of those functions in one model. Actually, the possibilities offered in the BayesX software reach beyond those applied here.[12]

In our application, we have demonstrated that this methodology can be useful to estimate historical time trends from micro data. Trends in human stature in 19th century America have been shown to be quite different among different occupational groups. Our results indicate a substantial gain in the "biological standard of living" of the middle class in comparison to the rest of the society during the Antebellum decades, reflecting the increasing per capita income of this group as well as the rising inequality during the period.

[12] Link functions may be incorporated to estimate (multinomial) logit, (ordered) probit, and other models of the traditional GLM family. Moreover, spatial correlation in data organized by geographic units can be taken into account.

References

A'Hearn, B., 1998. The Antebellum Puzzle revisited: a new look at the physical stature of Union Army recruits during the Civil War. In: Komlos, J., Baten, J. (Eds.), The Biological Standard of Living in Comparative Perspective. Franz Steiner Verlag, Stuttgart, pp. 250–267.

Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration with two applications in spatial statistics (with discussion). Annals of the Institute of Statistical Mathematics 43, 1–59.

Brezger, A., Kneib, T., Lang, S., 2002. BayesX. Software for Bayesian Inference based on Markov Chain Monte Carlo simulation techniques, version 0.9. University of Munich, Munich.

Costa, D.L., Steckel, R.H., 1997. Long-term trends in health, welfare, and economic growth in the United States. In: Steckel, R.H., Floud, R. (Eds.), Health and Welfare During Industrialization. University of Chicago Press, Chicago, pp. 47–89.

Cuff, T., 1993. The body mass index values of nineteenth century West Point cadets: a theoretical application of Waaler's curves to a historical population. Historical Methods 26 (4), 171–183.

Cuff, T., 1998. Variation and trends in the stature of Pennsylvanians. In: Komlos, J., Baten, J. (Eds.), The Biological Standard of Living in Comparative Perspective. Franz Steiner Verlag, Stuttgart, pp. 208–235.

De Boor, C., 1978. A Practical Guide to Splines. Springer, New York.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). Statistical Science 11 (2), 89–121.

Fahrmeir, L., Lang, S., 2001a. Bayesian inference for generalized additive mixed models based on Markov random field priors. Journal of the Royal Statistical Society C (Applied Statistics) 50, 201–220.

Fahrmeir, L., Lang, S., 2001b. Bayesian semiparametric regression analysis of multicategorical time-space data. Annals of the Institute of Statistical Mathematics 53, 10–30.

Fahrmeir, L., Thtz, G., 2001. Multivariate statistical modelling based on generalized linear models. Springer, New York.

Green, P.J., 2001. A primer in Markov Chain Monte Carlo. In: Barudorff-Nielsen, O.E., Cox, D.R., Klüppelberg, C. (Eds.), Complex Stochastic Systems. Chapmann and Hall, London, pp. 1–62.

Green, P.J., Silverman, B., 1994. Nonparametric regression and generalized linear models. Chapman and Hall, London.

Haines, Michael R., Craig, L.A., Weiss, T., 2000. Development, Health, Nutrition, and Mortality: The Case of the 'Antebellum Puzzle' in the United States. Historical Working Paper 130. National Bureau of Economic Research.

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

Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. Journal of the Royal Statistical Society B 55, 757–796.

Komlos, J., 1987. The height and weight of West Point cadets: dietary change in Antebellum America. Journal of Economic History 47 (4), 897–927.

Komlos, J., 1996. Anomalies in economic history: toward a resolution of the 'Antebellum Puzzle'. Journal of Economic History 56 (1), 202–214.

Komlos, J., 1998. Shrinking in a growing economy? The mystery of physical stature during the industrial revolution. Journal of Economic History 58 (3), 779–802.

Komlos, J., in press. Access to food and the biological standard of living: perspectives on the nutritional status of native Americans. American Economic Review.

Lang, S., Brezger, A., in press. Bayesian P-splines. Journal of Computational and Graphical Statistics.

Lauderdale, D.S., Rathouz, P.J., 1999. Evidence of environmental suppression of familial resemblance: height among US Civil War brothers. Annals of Human Biology 26 (5), 413–426.

Lindert, P.H., 1991. Toward a comparative history of income and wealth inequality. In: Brenner, Y.S., Kaelble, H., Thomas, M. (Eds.), Income Distribution in Historical Perspective. Cambridge University Press, Cambridge, 212–231.

Margo, R., Steckel, R., 1983. Heights of native born whites during the Antebellum period. Journal of Economic History 43 (1), 167–174.

Steckel, R.H., 1995. Stature and the standard of living. Journal of Economic Literature 33 (4), 1903–1940.

Woitek, U., in press. Height cycles in the 18th and 19th centuries. Economic Appliquee.