The Pennsylvania State University
The Graduate School
Eberly College of Science

MODEL TESTING FOR PARTIALLY LINEAR MODELS

A Thesis in
Statistics
by
Lynn Waterhouse

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2012

The thesis of Lynn Waterhouse was reviewed and approved* by the following:

Michael G. Akritas
Professor of Statistics
Thesis Adviser

David Hunter
Professor of Statistics
Head of the Department of Statistics

Bing Li
Professor of Statistics
Graduate Chair Department of Statistics

*Signatures are on file in the Graduate School.

Abstract

In a partially linear model some covariates have a linear effect, while the effect of others may be nonlinear. We perform Monte Carlo simulations to compare two methods for testing the significance of covariates with (possibly) nonlinear effects. Both methods use the residuals from fitting only the covariates that have a linear effect. One of the methods is based on the sliced inverse regression (SIR) procedure of Li (1991) applied on the residuals. The other is an ANOVA-type procedure modeled after Wang, Akritas, and Van Keilegom (2008). We also use the two methods for testing two datasets, historical spirit consumption in the UK and a CO_2 study, both of which have been described in the literature with partially linear models. This study also serves as a preliminary investigation as to whether the asymptotic theory developed by Li (1991) for the SIR procedure is also relevant when residuals are used.

Contents

List of Tables

List of Tables

List of Figures

List of Figures

Acknowledgements

Acknowledgements

I thank Dr. C. Yi for granting us permission to use the data related to temperature and net ecosystem CO_2 exchange (NEE), which was provided by Dr. Runze Li.

Additionally, I would like to futher thank my adviser Dr. Akritas for helping me as an adviser, mentor, and friend in my thesis, course work, and teaching duties. I would also like to thank the following persons: Jenn Parkes for being an immense asset in the Statistics Department, without her, the department literally would not run; Dr. Gus Colangelo, for helping me find a second home at Smeal and allowing me to participate in truly amazing projects; Dr. Naomi Altman, for helping me to navigate my way through my Statistics degree, Dr. Durland Shumway, for further opening my eyes to the wonderful world of consulting, and Dr. Don Richards, for providing advice in difficult times of teaching and for leading lively discussions that I believe should be taking place in places of higher education. Also, I would like to thank Dr. David Hunter and Dr. Bing Li for reading this thesis. Finally, I would like to thank my parents and my brother for proofreading my thesis, even though, the contents found within may have been less than appealing.

Chapter 1. Introduction

Partially linear models define a general class of models in which one component of the regression function is a linear function of unknown parameter and the other component is an unknown function. In particular, this model specifies that the response, Y_i, is related to covariates $\mathbf{X}_i$ and $\mathbf{T}_i$ through the equation

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + g(\mathbf{T}_i) + \xi_i, \ = 1, \ldots, n, \tag{0.1}$$

where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ir})$ and $\mathbf{T}_i = (T_{i1}, \ldots, T_{is})$. The pairs $(\mathbf{X}_i, \mathbf{T}_i)$ can be either independent and identically (IID) distributed random variables or fixed design points, the error variables ξ_i are IID with zero mean and finite variance and are uncorrelated from $(X_i, \mathbf{T}_i)$, the $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)$ is the vector of unknown parameters, and $g()$ is an unknown function that maps $\mathbb{R}^s$ into $\mathbb{R}^1$. Engle, Granger, Rice and Weiss (1986) were among the first to consider this model and used it to analyze the monthly electricity sales (Y) for four cities using as covariates the price of electricity (X_1), income (X_2), and the average daily temperature (T).

In this thesis we will consider a heteroscedastic version of this model by modeling the error variables ξ_i as:

$$\xi_i = \sigma(\mathbf{X}_i, \mathbf{T}_i)\epsilon_i,$$

where the ϵ_i are IID with zero mean and finite variance σ_ϵ^2, and $\sigma(\mathbf{X}_i, \mathbf{T}_i)$ is the standard deviation of the error variable ξ_i in (0.1). It will be assumed that $\sigma(\mathbf{X}_i, \mathbf{T}_i)$ remains bounded over the range of $(\mathbf{X}_i, \mathbf{T}_i)$.

Partially linear models, which belong in the class of semiparametric models,

have a broad range of applications including microeconomics and time series analysis; see Härdle et al. 2000, Härdle et al. 2007, Zhu 2005, Engle et al. 1986. Their practical appeal is further enhanced by the fact that the parametric components can be estimated at a rate of $\sqrt{n}$, and thus can avoid the curse of dimensionality which plagues the fully nonparametric models, where the nonparametric function estimation precision decreases rapidly as the dimension of the nonlinear variable, $\mathbf{T}$, increases (Härdle et al. 2000).

One such example is the analysis of the consumption of spirits in the United Kingdom (UK) from 1980 to 1938 by You and Zhou (2005). They looked at the relationship between annual per capita spirit consumption, Y_t, as a result of per capita income, X_{t1}, and price per spirit, X_{t2} (both income and price were deflated by a general price index and all data is in logarithmic form). The original dataset was studied by A. R. Prest (1949), discussed in Durbin and Watson (1951), and the dataset can be found in Fuller (1976).

Another example involves the analysis of how temperature affects the relationship between net ecosystem CO_2 exchange (NEE) and the photosynthetically active radiation (PAR). The data consists of 1997 observation of temperature (T), NEE, and PAR from 1999. A partially linear model was fit to the data by Li and Nie (2008).

In this thesis we consider the problem of testing the significance of the nonparametric component of the model (0.1). The two methods that will be compared are: (1) ANOVA-type test of Wang, Akritas, and Van Keilegom (2008) and (2) Li's (1991) test based on sliced inverse regression (SIR). The performance of these methods will be evaluated on the basis of their type I error rate and power. When the null hypothesis that there is no nonlinear component holds, the type I error rate should be close to the nominal level of significance, which is set at 0.05. Under the alternative hypothesis,

when the model consists also of a nonlinear component, the power should be as large as possible. The simulation settings use univariate $\mathbf{T}$, but the extension to multivariate $\mathbf{T}$ will be discussed.

The simulations performed also serve as a preliminary investigation as to whether the asymptotic theory developed by Li (1991) for the SIR procedure is also relevant when residuals are used and when the regression model is heteroscedastic. This is of particular interest in the context of partial linear models, because fitting of the $\mathbf{T}$ covariates is much more problematic when $\mathbf{T}$ is high dimensional. However, it is recognized that further numerical studies should be performed before a clear picture emerges. In particular, the present simulations use only a one-dimensional $\mathbf{T}$ covariate. In this case, Li's (1991) test procedure takes a particularly simple form and the results of the present simulations cannot be extrapolated to higher dimensional $\mathbf{T}$. Moreover, the present simulations use residuals obtained under the null hypothesis that the $\mathbf{T}$ covariates have no effect. A method for fitting the $\mathbf{X}$ covariates when the $\mathbf{T}$ covariates are also in the model is briefly discussed but not used in the present simulations.

This thesis is organized as follows: Section describes the ANOVA-type test of Wang, Akritas, Van Keilegom (2008); Section describes the Sliced Inverse Regression test; Section details the methods used for the Monte Carlo simulations and the two data examples; Section presents the results from the Monte Carlo simulations and the two data examples; and Section contains the final discussion for the thesis including a brief introduction to the alternative type of residuals.

Chapter 2. The ANOVA-Type Test

The name "ANOVA-type" is due to the fact that this test is a regression adaptation of the test for testing the equality of group (or cell) means in a high-dimensional heteroscedastic one-way ANOVA design developed in Akritas and Papadatos (2004). A brief description of this test is given in the next subsection, while its regression adaptation is presented in Subsection

High-Dimensional Heteroscedastic ANOVA

Suppose we have responses $Y_{ij},\ i = 1, \ldots, n,\ j = 1, \ldots, k_i$, where for each i the $Y_{ij},\ i = 1, \ldots, k_i$, are IID with mean μ_i and variance σ_i^2, and are interested in testing the equality of the means. Decomposing the means as

$$\mu_i = \mu + \alpha_i, \text{ where } \mu = \frac{1}{n}\sum_{i=1}^{n} \mu_i, \text{ and } \alpha_i = \mu_i - \mu$$

the null hypothesis can be written as $H_0 = \alpha_i = 0, \forall i$. Assuming normality and homoscedasticity,

$$F = \frac{MST}{MSE} \underset{H_0}{\sim} F_{n-k,n(k-1)}, \tag{0.2}$$

where $MST = \dfrac{\sum_{i=1}^{n} k_i[(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})]^2}{(n-1)}$, with $\bar{Y}_{i\cdot} = k_i^{-1}\sum_{j=1}^{k_i} Y_{ij}$ and $\bar{Y}_{\cdot\cdot} = (\sum_{i=1}^{n} k_i)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{k_i} Y_{ij}$, and $MSE = \dfrac{1}{n}\sum_{i=1}^{n}\dfrac{1}{k_i - 1}\sum_{j=1}^{k_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$.

Relation (0.2) is approximately correct also under heteroscedasticity pro-

vided the design is balanced, i.e. $k_i = k$ for all i; cf. Sheffe (1959), Section 10.4.

Without normality we need to use the asymptotic approximation. Two types of asymptotic results are available, one where the number of groups is fixed and the group sizes tend to infinity, and one where the number of groups tends to infinity. In the brief description of the two results that follows, it is assumed that $k_i = k$ and also $\sigma_i^2 = \sigma^2$, for all i.

Consider first the case where n stays fixed and $k \to \infty$ (see Arnold 1981). In this case it is fairly easy to establish that, as $k \to \infty$,

$$MSE = \frac{1}{n}\sum_{i=1}^{n} S_i^2 \xrightarrow{p} \sigma^2 \quad \text{and} \quad MST = \frac{1}{n-1}\sum_{i=1}^{n}[\sqrt{k}(\bar{Y}_{i.} - \bar{Y}_{..})]^2 \xrightarrow[H_0]{d} \sigma^2\chi^2_{n-1}.$$

Thus, by Slutsky's theorem,

$$F = \frac{MST}{MSE} \xrightarrow[H_0]{d} \chi^2_{n-1}/(n-1), \quad \text{as} \quad k \to \infty.$$

Consider now the case where k stays fixed and $n \to \infty$. In this case it can be seen that, under H_0,

$$MSE \xrightarrow{p} \sigma^2 \quad \text{and} \quad MST \xrightarrow{p} \sigma^2.$$

Hence, k stays fixed and $n \to \infty$, $F \xrightarrow{p} 1$ under H_0, which implies that a test cannot be constructed. To construct a test in this case, one needs to establish an asymptotic theory for $\sqrt{n}(F-1)$. The basic result is

Theorem 0.1. *If $k \geq 2$ stays fixed and $n \to \infty$,*

$$\sqrt{n}(F-1) \xrightarrow[H_0]{d} N\left(0, \frac{2k}{k-1}\right).$$

This and additional results including the heteroscedastic case can be found in Akritas and Papadatos (2004).

The ANOVA-Type Test for Model Checking in Regression

Suppose we have data (X_i, Y_i), $i = 1, \ldots, n$, which follow the regression model

$$Y_i = m(X_i) + \epsilon_i,$$

and we wish to test the hypothesis $H_0 = m(x) = C$ for all x, where C is a constant. A conceptual connection between the regression model and the high-dimensional one-way ANOVA can be established by considering each level X_i of the covariate as a factor level. With this convention, the hypothesis of a constant regression function corresponds to the ANOVA hypothesis of equality of the means. However, the test of Akritas and Papadatos (2004) cannot be used because, typically, there is only one observation per covariate value, whereas the asymptotic theory described in Theorem 0.1 requires at least two observations per factor level.

To get around this difficulty, one constructs artificial replications for each "group" (i.e., each specific value of the covariate) (see Wang et al. 2008). The groups are constructed by creating a window, W_i, around each covariates value X_i of size k, which contains X_i and the $(k-1)/2$ covariates X_j

that are closest to X_i on either side. More precisely we consider W_i as a set of indices defined by

$$W_i = \left\{ j : |\hat{F}_X(X_j) - \hat{F}_X(X_i)| \leq \frac{k-1}{2n} \right\},$$

where $\hat{F}_X$ is the empirical distribution function of the covariate values.

In the definition of MST and MSE, the Y_{ij} will be replaced with the Y_j, for $j \in W_i$, and here we will denote them as $V_{i1}, \ldots, V_{ik}$ the k independent observations from group i. Then we have:

$$F_n = \frac{MSE}{MST},$$

where

$$MST = \frac{k}{n-1} \sum_{i=1}^{n} (\bar{V_{i\cdot}} - \bar{V_{\cdot\cdot}})^2, MSE = \frac{1}{N-n} \sum_{i=1}^{n} \sum_{j=1}^{k} (V_{ij} - \bar{V_{i\cdot}})^2.$$

Under certain conditions, detailed Lemma 2.1 in Wang et al. 2008, we can look at the asymptotic distribution in two cases:
(1) If $k_n = k$ is fixed, then as $n \to \infty$,

$$n^{1/2}(MST - MSE) \to N\left(0, \frac{2k(2k-1)}{3(k-1)}\tau^2\right),$$

where $\tau^2 = \int_0^1 \sigma^4(x) r(x) dx$. (2) If $n \to \infty$ and $k_n \to \infty$ such that $k_n n^{-1} \to 0$, then with τ^2 defined above,

$$\left(\frac{n}{k_n}\right)^{1/2} (MST - MSE) \to N\left(0, \frac{4}{3}\tau^2\right),$$

An estimator for τ^2 can be used (Wang et al. 2008),

$$\widehat{\tau}^2 = \frac{1}{4(n-3)} \sum_{j=2}^{n-2} R_j^2 R_{j+2}^2,$$

where $R_j = Y_j - Y_{j-1}, j = 2, \ldots, n$, denote the local residuals.

Chapter 3. Sliced Inverse Regression

When the dimension of the covariate dataset is greater than the dimension of the response variable this can often lead to computation issues. One way to alleviate problems of this sort is through a process called dimension reduction. Dimension reduction refers to the process of reducing the number of covariates under consideration, and thus, reducing the dimensions. Some methods for dimension reduction include: principle component analysis (PCA), nonlinear dimensionality reduction methods, and cluster analysis, to name a few. Here the use of sliced inverse regression (SIR) will be evaluated. SIR is different from PCA in that in takes the response into account.

Ker-Chau Li wrote a paper detailing the method of sliced inverse regression as it applies to dimension reduction (Li 1999). For a true (unknown) model such that,

$$Y = f(\boldsymbol{\beta}_1 \mathbf{X}, ..., \boldsymbol{\beta}_K \mathbf{X}, \varepsilon) \tag{0.3}$$

with the $\boldsymbol{\beta}$ are unknown row vectors, f is an unknown function in $\mathbb{R}^{K+1}$, and we make the strong assumption $\varepsilon \perp\!\!\!\perp \mathbf{X}$.

Looking at the expected value, $E(\mathbf{b}^T\mathbf{X}|\boldsymbol{\beta}_1^T\mathbf{X},...,\boldsymbol{\beta}_K^T\mathbf{X}) = c_o + c_1\boldsymbol{\beta}_1^T\mathbf{X} + ... + c_K\boldsymbol{\beta}_K^T\mathbf{X} = c_0 + \mathbf{c}^T\mathbf{B}^T\mathbf{X}$.

Theorem 0.2. *Theorem 3.1 from Li (1991) states*

$$[E(\mathbf{X}|Y) - E(\mathbf{X})] \in span(\boldsymbol{\beta}_k^T \sum_{xx}, k = 1, ..., K).$$

We can assume, without a loss of generality, that $E(\mathbf{X}) = 0$. Let there be a $\mathbf{b}$ s.t. $\boldsymbol{\beta}_k^T \sum_{xx} \mathbf{b} = 0, \forall k$. To show $\mathbf{b}^T E(\mathbf{X}|Y) = 0$, we have,

$$\begin{aligned}
\mathbf{b}^T E(\mathbf{X}|Y) &= b_1 E(X_1|Y) + ... + b_p E(X_p|Y) \\
&= E(b_1 X_1 + ... + b_p X_p|Y) \\
&= E(\mathbf{b}\mathbf{X}|Y) \\
&= E[E(\mathbf{b}^T\mathbf{X}|\boldsymbol{\beta}_1^T\mathbf{X},...,\boldsymbol{\beta}_K^T\mathbf{X}, Y)|Y] \\
&= E[E(\mathbf{b}^T\mathbf{X}|\boldsymbol{\beta}_1^T\mathbf{X},...,\boldsymbol{\beta}_K^T\mathbf{X})|Y].
\end{aligned}$$

For the last step we use equation 0.3 and the assumption of $\varepsilon \perp\!\!\!\perp \mathbf{X}$.

To show that this is zero, it suffices to show that:

$$E(\mathbf{b}^T\mathbf{X}|\boldsymbol{\beta}_1^T\mathbf{X},...,\boldsymbol{\beta}_k^T\mathbf{X}) = 0,$$

or

$$E[E(\mathbf{b}^T\mathbf{X}|\mathbf{B}^T\mathbf{X})^2] = 0,$$

or

$$E[E(\mathbf{b}^T\mathbf{X}|\mathbf{B}^T\mathbf{X})\mathbf{b}^T\mathbf{X}],$$

or

$$E[c_0 + \mathbf{c}_1^T\mathbf{B}^T\mathbf{X})\mathbf{X}^T\mathbf{b}] = E(c_0\mathbf{X}^T\mathbf{b}) + \mathbf{c}^T\mathbf{B}^T E(\mathbf{X}\mathbf{X}^T)\mathbf{b} = \mathbf{c}\mathbf{B}^T \sum_{xx} \mathbf{b} = 0.$$

Here we make use of the fact that $E(c_0\mathbf{X}^T\mathbf{b}) = 0$ because $E(\mathbf{X}) = 0$.

The test for sliced inverse regression on $(\mathbf{Y}_i, \mathbf{X}_i), i = 1, \ldots, n$, is conducted in the following way:

1. Standardize $\mathbf{X}$ by an affine transformation to get $\widetilde{\mathbf{X}_i} = \widehat{\Sigma_{\mathbf{XX}}^{-1/2}}(\mathbf{X}_i - \overline{\mathbf{X}})$, for $i = 1, ..., n$, where $\widehat{\Sigma_{\mathbf{XX}}}$ and $\overline{\mathbf{X}}$ are the sample covariance matrix and sample mean of $\mathbf{X}$ respectively.
2. Divide the range of $\mathbf{Y}$ into H slices, $I_1, \ldots, I_H$; let the proportion of the $\mathbf{Y}_i$ that falls into slice h be $\hat{p_h}$; that is, $\hat{p_h} = \frac{1}{n}\sum_{i=1}^{n} \delta_h(\mathbf{Y}_i)$, where $\delta_h(\mathbf{Y}_i)$ takes the values 0 or 1 depending on whether $\mathbf{Y}_i$ falls into the hth slice, I_h, or not.
3. Within each slice, compute the sample mean of the $\overline{\mathbf{X}_i}$'s, denoted by $\hat{m_h}$, for $h = 1, \ldots, H$, so that $\hat{m_h} = (1/n\hat{p_h}) \sum_{\mathbf{Y}_i \in I_h} \widetilde{\mathbf{X}_i}$.
4. Conduct a (weighted) principal component analysis for the data $\hat{m_h}$, for $h = 1, \ldots, H$, in the following way: Form the weighted covariance matrix, $\widehat{V} = \sum_{h=1}^{H} \hat{p_h}\hat{m_h}{\hat{m_h}}^T$, then find the eigenvalues and the eigenvectors for $\widehat{V}$.
5. Let the K largest eigenvectors (row vectors) be $\widehat{\eta_k}$, for $k = 1, \ldots, K$. Output $\widehat{\beta_k} = \widehat{\eta_k}\widehat{\Sigma_{\mathbf{XX}}^{-1/2}}$, for $k = 1, \ldots, K$ and $H \geq K + 1$.

That the K dimensions for reduction have been successfully picked can be

checked using the companion output eigenvalues $\widehat{V}$ from Step (4) of the SIR process. The asymptotic distribution of the average of the smallest $p-K$ eigenvalues, denoted by $\bar{\lambda}_{(p-K)}$, for $\widehat{V}$ can be derived based on perturbation theory for finite-dimensional spaces. For normal $\mathbf{X}$, we have the following result.

Theorem 0.3. *Theorem 5.1 from Li (1991) states, if* $\mathbf{X}$ *is normally distributed, then* $n(p-K)\bar{\lambda}_{(p-K)}$ *follows a* χ^2 *distribution with* $(p-K)(H-K-1)$ *df asymptotically.*

We apply the methodology of Li on the residuals (from least squares regression), which is not yet justified theoretically. As such, this is just a preliminary investigation as to the effectiveness of this methodology. As previously stated, the goal is to reduce the dimensionality of the nonlinear portion.

For the purposes of this thesis we will use 2 slices, so we know K to be 1. As such, the test simplifies into a two sample t-test.

Chapter 4. Methods

The software package R (R Development Core Team 2010) was used to conduct Monte Carlo simulations to evaluate the performance of each testing method and for the data example analyses. We focused on the simplest case here, that is when the nonlinear component is one dimensional. This was done to ease the computations.

For both the Monte Carlo simulations and the data example first the linear

component of the model was fit using least squares and then the residuals were used to test if the nonparametric component existed, using: (1) ANOVA-type test of Wang, Akritas, and Van Keilegom (2008) and (2) sliced inverse regression. We discuss alternative methods of estimation for the linear and nonparametric components in the discussion section.

Monte Carlo Simulations

In each case 1,000 simulations were performed.

The $\mathbf{X}$ values were generated uniformly from $\frac{1}{100}$ to 1 at intervals of $\frac{1}{100}$ and then standardized. The $\mathbf{T}$ values were generated from the normal distribution with a specified mean and standard deviation. The $\mathbf{Y}$ vector was generated such that $\mathbf{Y} = \alpha + \beta\mathbf{X} + m\mathbf{X}^d\mathbf{Z} + g(\mathbf{T})$, where α is the intercept from the linear portion, β is the slope from the linear portion, $m\mathbf{X}^d\mathbf{Z}$ are the heteroscedastic errors, and $g(\mathbf{T})$ is the non-linear function of $\mathbf{T}$. The heteroscedastic errors, $m\mathbf{X}^d\mathbf{Z}$, where m is a multiplier, d is a power value, and $\mathbf{Z}$ are IID from the normal distribution with a mean of zero and a specified variance. The simplest case is when the null hypothesis is true, i.e., $g(\mathbf{T}) = 0$. Additional simulations were run with the $g(\mathbf{T})$ function becoming more different from the null hypothesis. More details of the function, $g(\mathbf{T})$ can be found in Table 2. For the ANOVA-type test the group size, or window, was 5.

Data Example: UK Spirit Consumption

Data were collected on the consumption of spirits in the United Kingdom from 1870 to 1938. The dependent variable is the annual per capita spirit

consumption, $\mathbf{y}_t$, and the covariates are per capita income, $\mathbf{x}_{t1}$, and price per spirit, $\mathbf{x}_{t2}$ (both income and price were deflated by a general price index and all data is in logarithmic form). The original dataset was studied by A. R. Prest (1949), discussed in Durbin and Watson (1951), and the dataset can be found in Fuller (1976, on page 427).

The data were read into R and then the linear component of the model was fit using the built in least squares package, 'lm'. We then tested the null hypothesis that the nonlinear component did not exist, with residuals as the dependent variable and time (year) as the independent variable.

Data Example: Net Ecosystem CO_2 Exchange

Data were collected at various sites at different elevations from 1999 to 2002. A subsection of that data from a subapline forest (approximately 3050 meters above sea level) in 1999 was analyzed. The data consist of measurements of the net ecosystem CO_2 exchange (NEE), temperature (T), and the photosynthetically active radiation (PAR). A proposed model for this relationship is $NEE = R(T)+m(PAR)+\epsilon$, where $R(T)$ is a simple linear model as a function of temperature and $m(PAR)$ is a nonlinear function of PAR (Li and Nie 2008).

Chapter 5. Results

Monte Carlo Simulations

When the null hypothesis is true, i.e., $g = 0$, both the ANOVA-type method and the testing method based on Li's SIR have p-values near the desired

level of 0.05 (see Table 1).

When the alternative hypothesis is true, i.e., $g \neq 0$, we see that the form of g matters (see Tables 2, 3, and 4). When g is a trigonometric function, such as *cos* or *sin*, we see that the ANOVA-type test has higher power than the test based on Li's SIR. Additionally, when g is an even power function the ANOVA-type test has higher power that the test based on Li's SIR. However, when g is a power function where the power is an odd number then the test based on Li's SIR does nearly as well as the ANOVA-type test in terms of having high power, in fact, sometimes the test based on Li's SIR does better than the ANOVA-type test.

Futhermore, we see that the strength of within group variability affects the ANOVA-type test's ability to detect a pattern (see Figures 1, 2, 3, 4, 5, and 6). When the g function is $sin(.2\mathbf{T})^4$ there is less within group variability, for the ANOVA-type test, compared to when $g = sin(.4\mathbf{T})^4$), and as a result the power is higher for the former (see Tables 2, 3, and 4).

The ANOVA-type test appears to conservative under H_0, that is when the nonlinear portion does not exist. When we add in homoscedastic errors, the ANOVA-type test becomes less conservative and the test based on Li's Sliced Inverse Regression becomes more conservative.

Data Example: Spirit Consumption in the UK

You and Zhou (2005) fit a partially linear model to the UK spirit consumption data and found that this model fit the spirit consumption data better than the time series model of Fuller (1976). So prior to testing there was a strong indication that the null hypothesis, that the nonlinear portion did not exist, would be rejected. Only the ANOVA-type test results in the

rejection of the null hypothesis, with a p-value of 0.0000 (see Table 6). The test based on Li's SIR fails to reject the null hypothesis, with a p-value of 0.1463. We can see that the residuals have a pattern that the test based on Li's SIR cannot detect, since the range of the $\mathbf{T}$ values (year) will be similar in each of the slices (Figure 7).

Data Example: Net Ecosystem CO_2 Exchange

Li and Nie (2008) fit a partially linear model to the net ecosystem CO_2 exchange data, in which NEE was related to temperature through a linear component and related to PAR by a nonparametric function (Li and Nie 2007, 2008). The data are shown in Figure 8. Both the ANOVA-type test and the Sliced Inverse Regression test result in the rejection of the null hypothesis, with both tests resulting in a p-value of 0.0000 (see Table 7).

Chapter 6. Discussion

For this thesis we used the residuals obtained by least squares fitting of the $\mathbf{X}$ covariates under the null hypothesis which specifies that the $\mathbf{T}$ covariates have no effect. Another type of residuals comes from estimating the methods of Speckman (1988) and Robinson (1988).

Following their method we can take equation 0.1 and rearrange it to yield, $Y_i - \mathbf{X}_i^T\boldsymbol{\beta} = g(\mathbf{T}_i) + \xi_i$. From this we can see that an estimator of the nonlinear component is $\widehat{g}(\mathbf{T}_i) = \sum_{j=1}^n (Y_j - \mathbf{X}_j^T\boldsymbol{\beta})W_{ij}$, for a suitable set of weights $W_{ij} = W_j(\mathbf{T}_i)$. Solving for Y_i we get, $Y_i = \mathbf{X}_i^T\boldsymbol{\beta} + \sum_{j=1}^n Y_j\mathbf{W}_j(\mathbf{T}_i) - (\sum_{j=1}^n \mathbf{W}_j(\mathbf{T}_i)\mathbf{X}_j^T)\boldsymbol{\beta}$. Rearranging the equation we get $Y_i - \sum_{j=1}^n Y_j\mathbf{W}_j(\mathbf{T}_i) = (\mathbf{X}_i^T - \sum_{j=1}^n \mathbf{W}_j(\mathbf{T}_i)\mathbf{X}_j^T)\boldsymbol{\beta} + \xi_i$. This leads to a new

regression equation:

$$\widetilde{Y}_i = \widetilde{\mathbf{X}_i}^T \boldsymbol{\beta} + \xi_i, \tag{0.4}$$

where $\widetilde{Y}_i = Y_i - \sum_{j=1}^{n} Y_j \mathbf{W}_j(\mathbf{T}_i)$ and $\widetilde{\mathbf{X}_i}^T = \mathbf{X}_i^T - \sum_{j=1}^{n} \mathbf{W}_j(\mathbf{T}_i)\mathbf{X}_j^T$.

We suspect that the choice of residuals should not have a large effect on the type I error rate, but that the type II error rate will be effected by the choice of residuals especially if there is dependence between $\mathbf{X}$ and $\mathbf{T}$. We believe that the residuals from equation 0.4 will provide better power than the residuals from least squares, which is what was used for the purpose of this thesis.

Finally, there are some limitations of the work presented in this thesis. First, the nonlinear portion consisted of just one covariate so the procedure based on Li's SIR reduces to a simple t-test. If $\mathbf{T}$ were multivariate than we would have applied Li's actual procedure for SIR (described in Section). Additionally, we are fitting the residuals from least squares regression, when it would be better to use the residuals that come from fitting the linear model while taking the nonlinear portion into account (as described by Equation 0.4). Another minor limitation is that we only explored using the test based on Li's SIR when we have two slices.

Bibliography

References

[1] Akritas, M. G. and N. Papadatos. 2004. Heteroscedastic one-way ANOVA and lack-of-fit tests. Journal of the American Statistical Association. 99: 368-382.

[2] Arnold, Steven F. 1981. The theory of linear models and multivariate analysis. Wiley, New York.

[3] Durbin, J., and G. S. Watson. 1951. Testing for serial correlation in the least squares regression, II. Biometrika. 38: 159-178.

[4] Engle, R. F., Granger, C. W. J., Rice, J., and A. Weiss. 1986. Semiparametric estimates of the relation between weather and electricity sales. Journal of the American Statistical Association. 81: 310-320.

[5] Fuller, Wayne A. 1976. Introduction to Statistical Time Series. Jonn Wiley and Sons, Inc., New York.

[6] Härdle, W., Mori, Yuichi, and P. Vieu. 2007. Statistical methods for biostatistics and related fields. Springer, New York.

[7] Härdle, W., Liange, H., and J. Gao. 2000. Partially Linear Models. Physica-Verlang: A Springer-Verlang Company, New York.

[8] Li, Ker-Chau. 1991. Sliced inverse regression for dimension reduction. Journal of the American Statistical Association. 86: 316-327.

[9] Li, R. and Nie, L. 2007. A new estimation procedure for a partially nonlinear model via a mixed-effects approach. The Canadian Journal of Statistics. 35: 399-411.

[10] Li, R. and Nie, L. 2008. Efficient statistical inference procedures for partially nonlinear models and their applications. Biometrics. 64: 904-911.

[11] Prest, A. R. 1949. Some experiments in demand analysis. Review of Economics and Statistics. 31: 33-49.

[12] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[13] Robinson, P. M. 1988. Root-n-consistent semiparametric regression. Econometrica. 56: 931-954.

[14] Scheffe, Henry. 1959. The analysis of variance. Wiley, New York.

[15] Speckman, P. 1988. Kernel smoothing in partial linear models. Journal of the Royal Statistical Society, Series B. 50: 413-436.

[16] Wang, L. , Akritas, M. G. , and I. Van Keilegom. 2008. An ANOVA-type nonparametric diagnostic test for heteroscedastic regression models. Journal of Nonparametric Statistics. 20: 365-382.

[17] You, Jinhong and Xian Zhou. 2005. Boostrap of a semiparametric partially linear model with autoregressive errors. Statistica Sinica. 15: 117-133.

[18] Zhu, Lixing. 2005. Nonparametric monte carlo tests and their applications. Springer, New York.

Table 1: Results from testing H_0, $g = 0$, from Monte Carlo simulations. Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5.

Data generating equation	P-value ANOVA-type test	P-value Li's test
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z}$, where $Z \sim$norm(0,1)	0.038	0.058
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ}$, where $Z \sim$norm(0,1)	0.038	0.058
$\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{XZ}$, where $Z \sim$norm(0,1)	0.038	0.058
$\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^2\mathbf{Z}$, where $Z \sim$norm(0,1)	0.023	0.041
$\mathbf{Y} = 2 + 5\mathbf{X} + 3\mathbf{X}^2\mathbf{Z}$, where $Z \sim$norm(0,1)	0.023	0.041
$\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^2\mathbf{Z}$, where $Z \sim$norm(0,10^2)	0.023	0.041
$\mathbf{Y} = 2 + 5\mathbf{X} + 3\mathbf{X}^2\mathbf{Z}$, where $Z \sim$norm(0,10^2)	0.023	0.041
$\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^2\mathbf{Z}$, where $Z \sim$norm(0,100^2)	0.023	0.041
$\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^3\mathbf{Z}$, where $Z \sim$norm(0,1)	0.021	0.048
$\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^4\mathbf{Z}$, where $Z \sim$norm(0,1)	0.017	0.057
$\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^3\mathbf{Z}$, where $Z \sim$norm(0,10^2)	0.021	0.048
$\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^3\mathbf{Z}$, where $Z \sim$norm(0,100^2)	0.021	0.048

Table 2: Results from testing H_0, $g = 0$, from Monte Carlo simulations when H_0 is not true. Here $\mathbf{T} \sim$N(0,36) and the form of $g(\mathbf{T})$ is defined in the table below. Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5.

Data generating equation	Power ANOVA-type test	Power Li's test
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + cos(.4\mathbf{T})^4$, where $Z \sim$norm(0,1)	0.416	0.059
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + sin(.4\mathbf{T})^4$, where $Z \sim$norm(0,1)	0.455	0.065
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + sin(.2\mathbf{T})^4$, where $Z \sim$norm(0,1)	0.463	0.063
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^2$, where $Z \sim$norm(0,1)	1.000	0.053
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^3$, where $Z \sim$norm(0,1)	0.998	1.000
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^4$, where $Z \sim$norm(0,1)	0.991	0.044
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^5$, where $Z \sim$norm(0,1)	0.982	0.994
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^6$, where $Z \sim$norm(0,1)	0.957	0.045
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^7$, where $Z \sim$norm(0,1)	0.946	0.889

Table 3: Results from testing H_0, $g = 0$, from Monte Carlo simulations when H_0 is not true. Here $\mathbf{T}$ ∼N(0,1) and the form of $g(\mathbf{T})$ is defined in the table below. Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5.

Data generating equation	Power ANOVA-type test	Power Li's test
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + cos(.4\mathbf{T})^4$, where Z ∼norm(0,1)	0.027	0.890
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + sin(.4\mathbf{T})^4$, where Z ∼norm(0,1)	0.022	0.045
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + sin(.2\mathbf{T})^4$, where Z ∼norm(0,1)	0.023	0.047
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^2$, where Z ∼norm(0,1)	0.266	0.055
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^3$, where Z ∼norm(0,1)	0.886	0.999
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^4$, where Z ∼norm(0,1)	0.968	0.041
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^5$, where Z ∼norm(0,1)	0.979	0.994
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^6$, where Z ∼norm(0,1)	0.955	0.044
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + \mathbf{T}^7$, where Z ∼norm(0,1)	0.948	0.890

Table 4: Results from testing H_0, $g = 0$, from Monte Carlo simulations when H_0 is not true. Here $\mathbf{T}$ ∼N(0,36) and the form of $g(\mathbf{T})$ is defined in the table below. Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5.

Data generating equation	Power ANOVA-type test	Power Li's test
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + cos(.4\mathbf{T})^4$, where Z ∼norm(0,1)	0.416	0.059
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + sin(.4\mathbf{T})^4$, where Z ∼norm(0,1)	0.455	0.065
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + sin(.2\mathbf{T})^4$, where Z ∼norm(0,1)	0.463	0.063
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + \mathbf{T}^2$, where Z ∼norm(0,1)	1.000	0.055
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + \mathbf{T}^3$, where Z ∼norm(0,1)	0.998	1.000
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + \mathbf{T}^4$, where Z ∼norm(0,1)	0.991	0.045
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + \mathbf{T}^5$, where Z ∼norm(0,1)	0.982	0.994
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + \mathbf{T}^6$, where Z ∼norm(0,1)	0.957	0.046
$\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + \mathbf{T}^7$, where Z ∼norm(0,1)	0.946	0.889

Table 5: UK spirit consumption data (reproduced from Durbin and Watson (1951))

Year	Consumption $\mathbf{Y}$	Income $\mathbf{X}_1$	Price $\mathbf{X}_2$	Year	Consumption $\mathbf{Y}$	Income $\mathbf{X}_1$	Price $\mathbf{X}_2$
1870	1.9565	1.7669	1.9176	1905	1.9139	1.9924	1.9952
1871	1.9794	1.7766	1.9059	1906	1.9091	2.0117	1.9905
1872	2.0120	1.7764	1.8798	1907	1.9139	2.0204	1.9813
1873	2.0449	1.7942	1.8727	1908	1.8860	2.0018	1.9905
1874	2.0561	1.8156	1.8984	1909	1.7945	2.0038	1.9859
1875	2.0678	1.8083	1.9137	1910	1.7644	2.0099	2.0518
1876	2.0561	1.8083	1.9176	1911	1.7817	2.0174	2.0474
1877	2.0428	1.8067	1.9176	1912	1.7784	2.0279	2.0341
1878	2.0290	1.8166	1.9420	1913	1.7945	2.0359	2.0255
1879	1.9980	1.8041	1.9547	1914	1.7888	2.0216	2.0341
1880	1.9884	1.8053	1.9379	1915	1.8751	1.9896	1.9945
1881	1.9835	1.8242	1.9462	1916	1.7853	1.9843	1.9939
1882	1.9773	1.8395	1.9504	1917	1.6075	1.9764	2.2082
1883	1.9748	1.8464	1.9504	1918	1.5185	1.9965	2.2700
1884	1.9629	1.8492	1.9723	1919	1.6513	2.0652	2.2430
1885	1.9396	1.8668	2.0000	1920	1.6247	2.0369	2.2567
1886	1.9309	1.8783	2.0097	1921	1.5391	1.9723	2.2988
1887	1.9271	1.8914	2.0146	1922	1.4922	1.9797	2.3723
1888	1.9239	1.9166	2.0146	1923	1.4606	2.0136	2.4105
1889	1.9414	1.9363	2.0097	1924	1.4551	2.0165	2.4081
1890	1.9685	1.9548	2.0097	1925	1.4425	2.0213	2.4081
1891	1.9727	1.9453	2.0097	1926	1.4023	2.0206	2.4367
1892	1.9736	1.9292	2.0048	1927	1.3991	2.0563	2.4284
1893	1.9499	1.9209	2.0097	1928	1.3798	2.0579	2.4310
1894	1.9432	1.9510	2.0296	1929	1.3782	2.0649	2.4363
1895	1.9569	1.9776	2.0399	1930	1.3366	2.0582	2.4552
1896	1.9647	1.9814	2.0399	1931	1.3026	2.0517	2.4838
1897	1.9710	1.9819	2.0296	1932	1.2592	2.0491	2.4958
1898	1.9719	1.9828	2.0146	1933	1.2635	2.0766	2.5048
1899	1.9956	2.0076	2.0245	1934	1.2549	2.0890	2.5017
1900	2.0000	2.0000	2.0000	1935	1.2527	2.1059	2.4958
1901	1.9904	1.9939	2.0048	1936	1.2763	2.1205	2.4838
1902	1.9752	1.9933	2.0048	1937	1.2906	2.1205	2.4636
1903	1.9494	1.9797	2.0000	1938	1.2721	2.1182	2.4580
1904	1.9332	1.9772	1.9952				

Table 6: Results from testing H_0, $g = 0$, using the spirit consumption data

Testing Method	P-value
ANOVA-type test	0.0000
Sliced Inverse Regression	0.1463

Table 7: Results from testing H_0, $g = 0$, using the net ecosystem CO_2 exchange

Testing Method	P-value
ANOVA-type test	0.0000
Sliced Inverse Regression	0.0000

Figure 1: Plots of the response data, $\mathbf{Y}$, and covariate with a nonlinear effect, $\mathbf{T}$, generated from Monte Carlo simulations testing H_0, $g = 0$, when H_0 is not true. The response is generated as $\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + g(\mathbf{T})$, where $\mathbf{Z}$ ~N(0,1) and $\mathbf{T}$ ~N(0,36). Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5. The form of $g()$ in each graph is as follows: (A) $g(\mathbf{T}) = cos(.4\mathbf{T})^4$; (B) $g(\mathbf{T}) = sin(.4\mathbf{T})^4$; (C) $g(\mathbf{T}) = sin(.2\mathbf{T})^4$; (D) $g(\mathbf{T}) = \mathbf{T}^2$; (E) $g(\mathbf{T}) = \mathbf{T}^3$; (F) $g(\mathbf{T}) = \mathbf{T}^4$; (G) $g(\mathbf{T}) = \mathbf{T}^5$; (H) $g(\mathbf{T}) = \mathbf{T}^6$; and (I) $g(\mathbf{T}) = \mathbf{T}^7$.

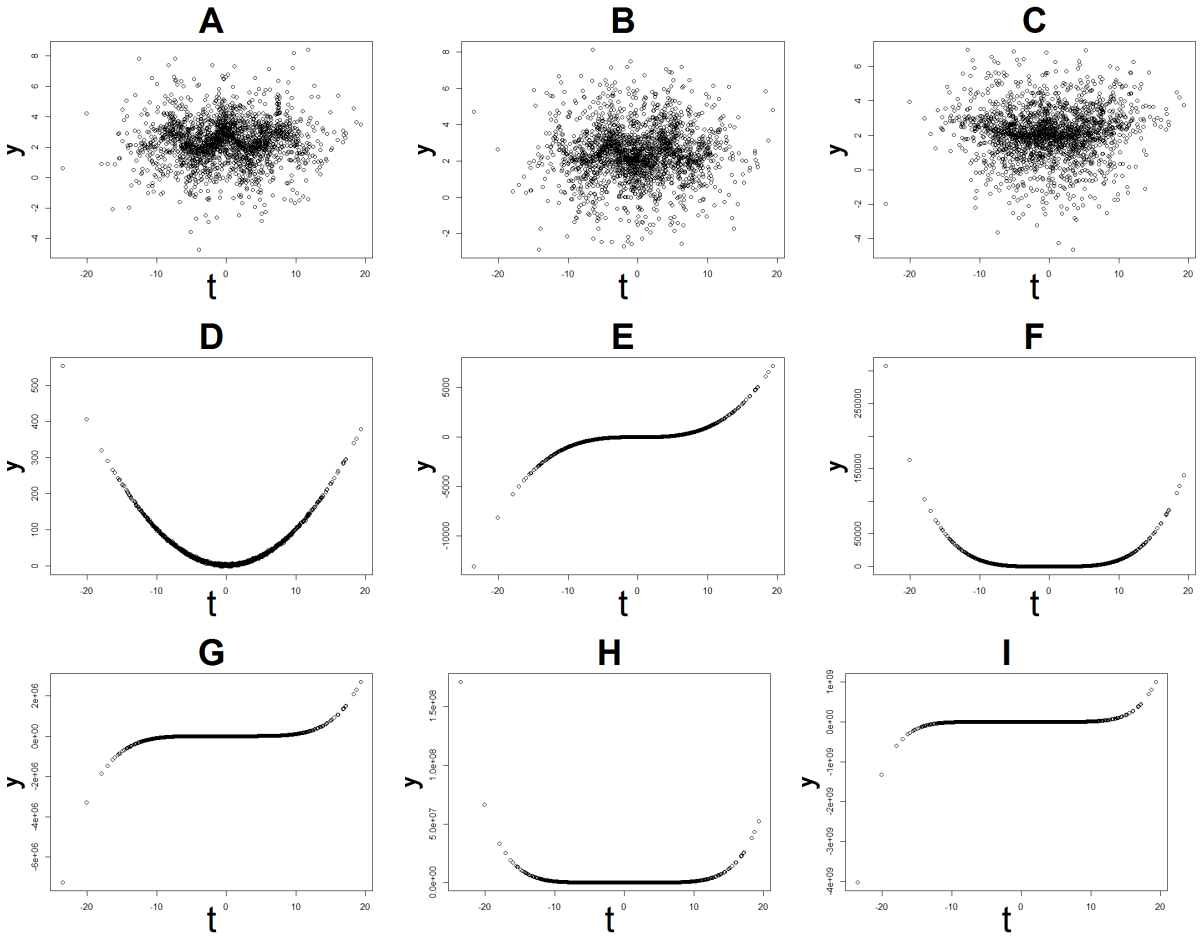

Figure 2: Plots of the residuals from least squares and the covariate with a nonlinear effect, $\mathbf{T}$, generated from Monte Carlo simulations testing H_0, $g = 0$, when H_0 is not true. The residuals come from fitting of linear model of $\mathbf{Y}$ on $\mathbf{X}$, where the response is generated as $\mathbf{Y} = 2 + \mathbf{X} + \mathbf{XZ} + g(\mathbf{T})$, where $\mathbf{Z}$ ~N(0,1) and $\mathbf{T}$ ~N(0,36). Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5. The form of $g()$ in each graph is as follows: (A) $g(\mathbf{T}) = cos(.4\mathbf{T})^4$; (B) $g(\mathbf{T}) = sin(.4\mathbf{T})^4$; (C) $g(\mathbf{T}) = sin(.2\mathbf{T})^4$; (D) $g(\mathbf{T}) = \mathbf{T}^2$; (E) $g(\mathbf{T}) = \mathbf{T}^3$; (F) $g(\mathbf{T}) = \mathbf{T}^4$; (G) $g(\mathbf{T}) = \mathbf{T}^5$; (H) $g(\mathbf{T}) = \mathbf{T}^6$; and (I) $g(\mathbf{T}) = \mathbf{T}^7$.

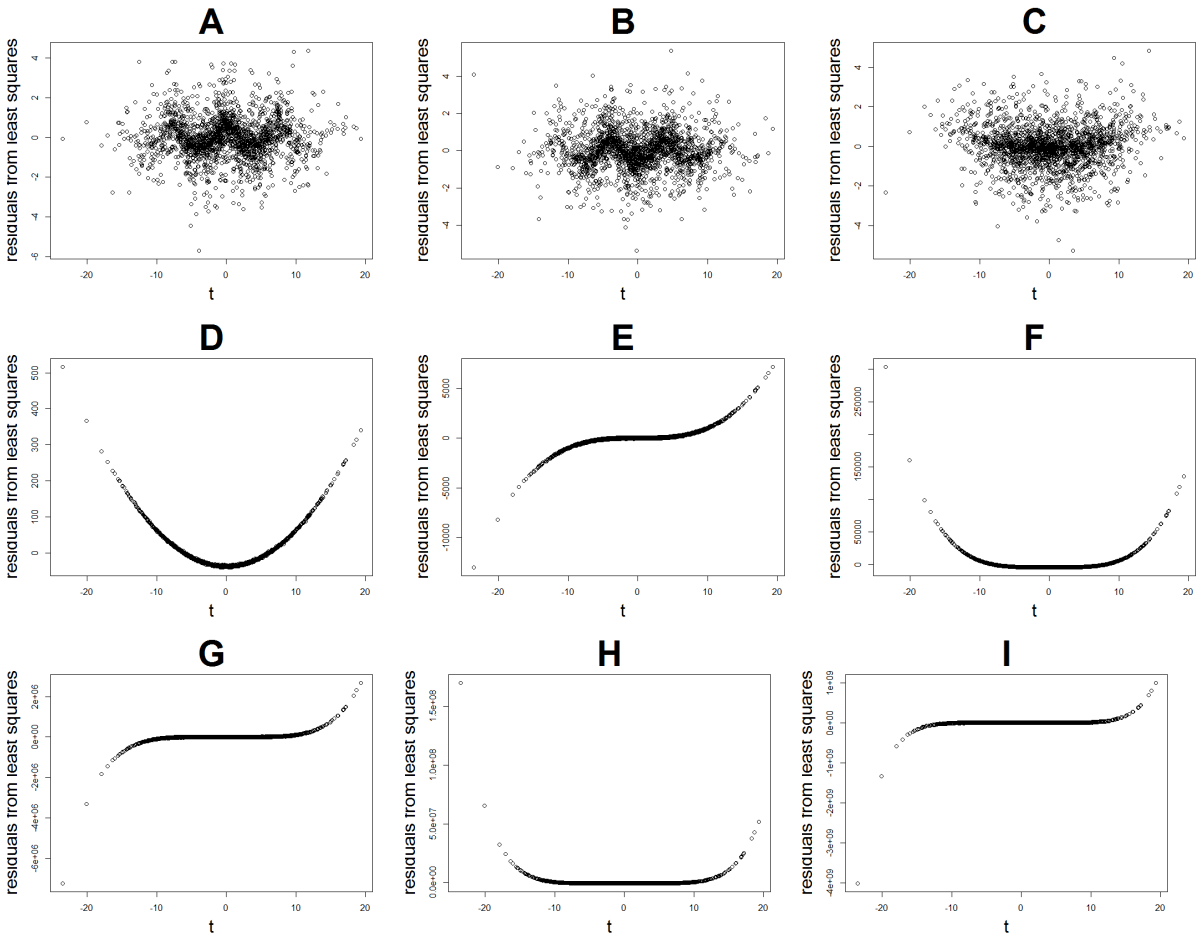

Figure 3: Plots of the response data, $\mathbf{Y}$, and covariate with a nonlinear effect, $\mathbf{T}$, generated from Monte Carlo simulations testing H_0, $g = 0$, when H_0 is not true. The response is generated as $\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^2\mathbf{Z} + g(\mathbf{T})$, where $\mathbf{Z}$ ~N(0,1) and $\mathbf{T}$ ~N(0,1). Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5. The form of $g()$ in each graph is as follows: (A) $g(\mathbf{T}) = cos(.4\mathbf{T})^4$; (B) $g(\mathbf{T}) = sin(.4\mathbf{T})^4$; (C) $g(\mathbf{T}) = sin(.2\mathbf{T})^4$; (D) $g(\mathbf{T}) = \mathbf{T}^2$; (E) $g(\mathbf{T}) = \mathbf{T}^3$; (F) $g(\mathbf{T}) = \mathbf{T}^4$; (G) $g(\mathbf{T}) = \mathbf{T}^5$; (H) $g(\mathbf{T}) = \mathbf{T}^6$; and (I) $g(\mathbf{T}) = \mathbf{T}^7$.

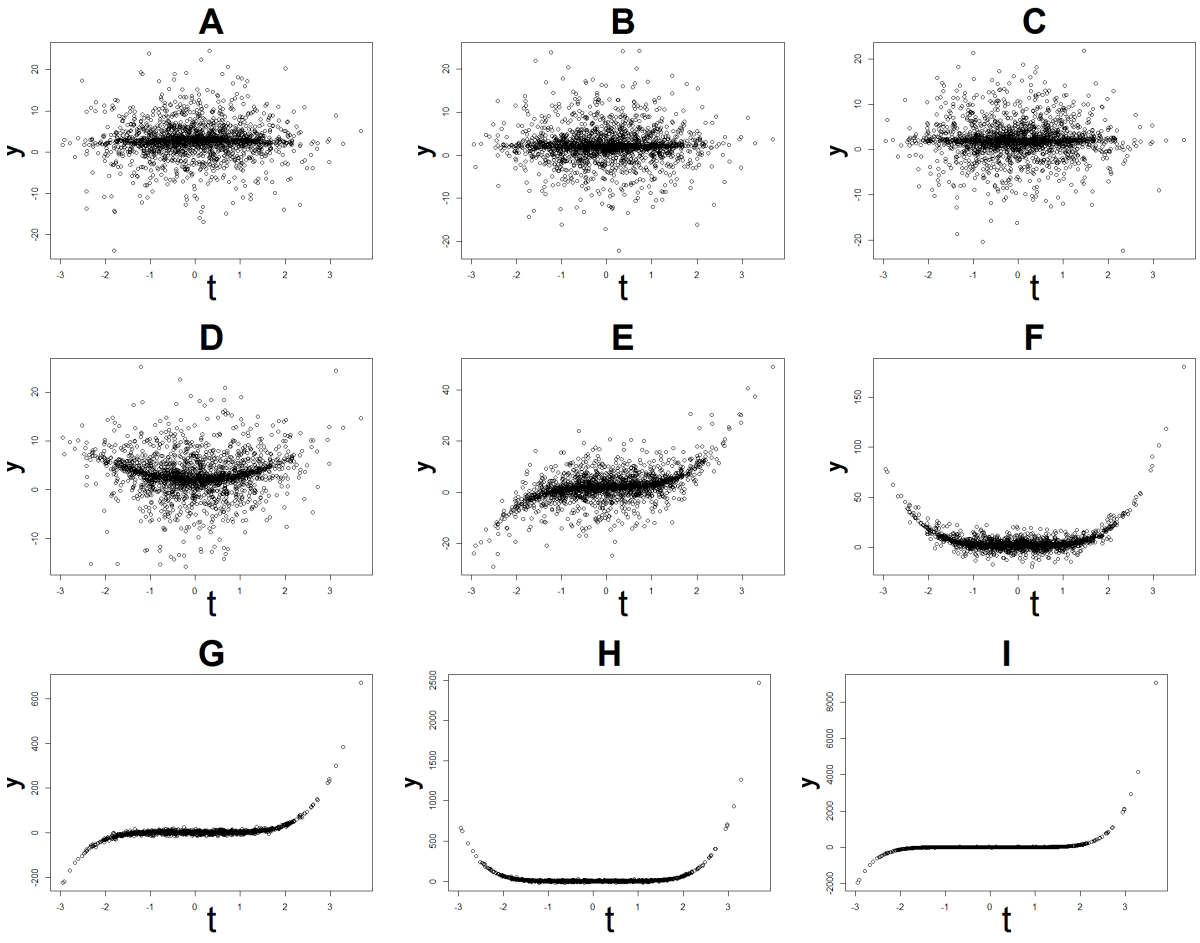

Figure 4: Plots of the residuals from least squares and the covariate with a nonlinear effect, $\mathbf{T}$, generated from Monte Carlo simulations testing H_0, $g = 0$, when H_0 is not true. The residuals come from fitting of linear model of $\mathbf{Y}$ on $\mathbf{X}$, where the response is generated as $\mathbf{Y} = 2 + \mathbf{X} + 3\mathbf{X}^2\mathbf{Z} + g(\mathbf{T})$, where $\mathbf{Z} \sim$N(0,1) and $\mathbf{T} \sim$N(0,1). Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5. The form of $g()$ in each graph is as follows: (A) $g(\mathbf{T}) = cos(.4\mathbf{T})^4$; (B) $g(\mathbf{T}) = sin(.4\mathbf{T})^4$; (C) $g(\mathbf{T}) = sin(.2\mathbf{T})^4$; (D) $g(\mathbf{T}) = \mathbf{T}^2$; (E) $g(\mathbf{T}) = \mathbf{T}^3$; (F) $g(\mathbf{T}) = \mathbf{T}^4$; (G) $g(\mathbf{T}) = \mathbf{T}^5$; (H) $g(\mathbf{T}) = \mathbf{T}^6$; and (I) $g(\mathbf{T}) = \mathbf{T}^7$.

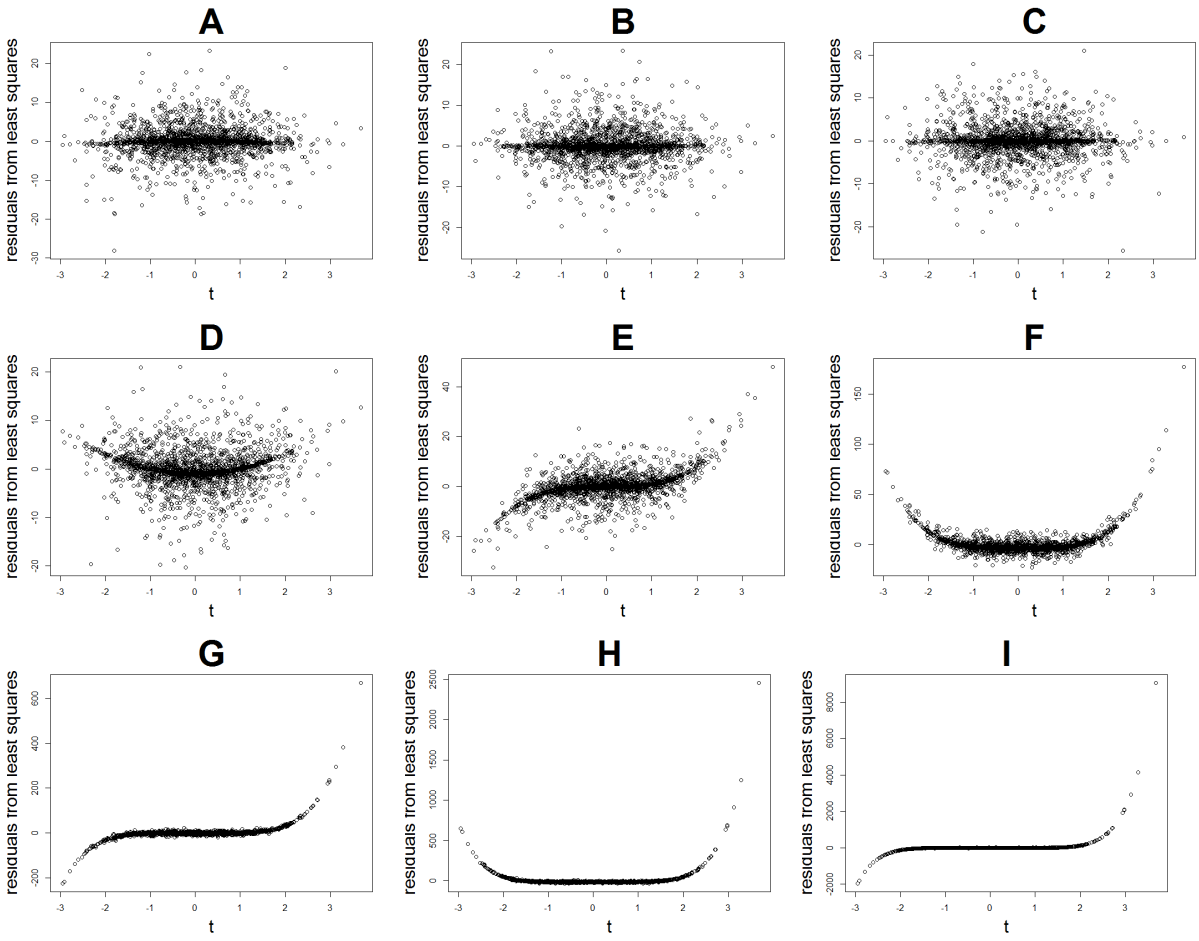

Figure 5: Plots of the response data, $\mathbf{Y}$, and covariate with a nonlinear effect, $\mathbf{T}$, generated from Monte Carlo simulations testing H_0, $g = 0$, when H_0 is not true. The response is generated as $\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + g(\mathbf{T})$, where $\mathbf{Z}$ $\sim$N(0,1) and $\mathbf{T}$ $\sim$N(0,36). Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5. The form of $g()$ in each graph is as follows: (A) $g(\mathbf{T}) = cos(.4\mathbf{T})^4$; (B) $g(\mathbf{T}) = sin(.4\mathbf{T})^4$; (C) $g(\mathbf{T}) = sin(.2\mathbf{T})^4$; (D) $g(\mathbf{T}) = \mathbf{T}^2$; (E) $g(\mathbf{T}) = \mathbf{T}^3$; (F) $g(\mathbf{T}) = \mathbf{T}^4$; (G) $g(\mathbf{T}) = \mathbf{T}^5$; (H) $g(\mathbf{T}) = \mathbf{T}^6$; and (I) $g(\mathbf{T}) = \mathbf{T}^7$.

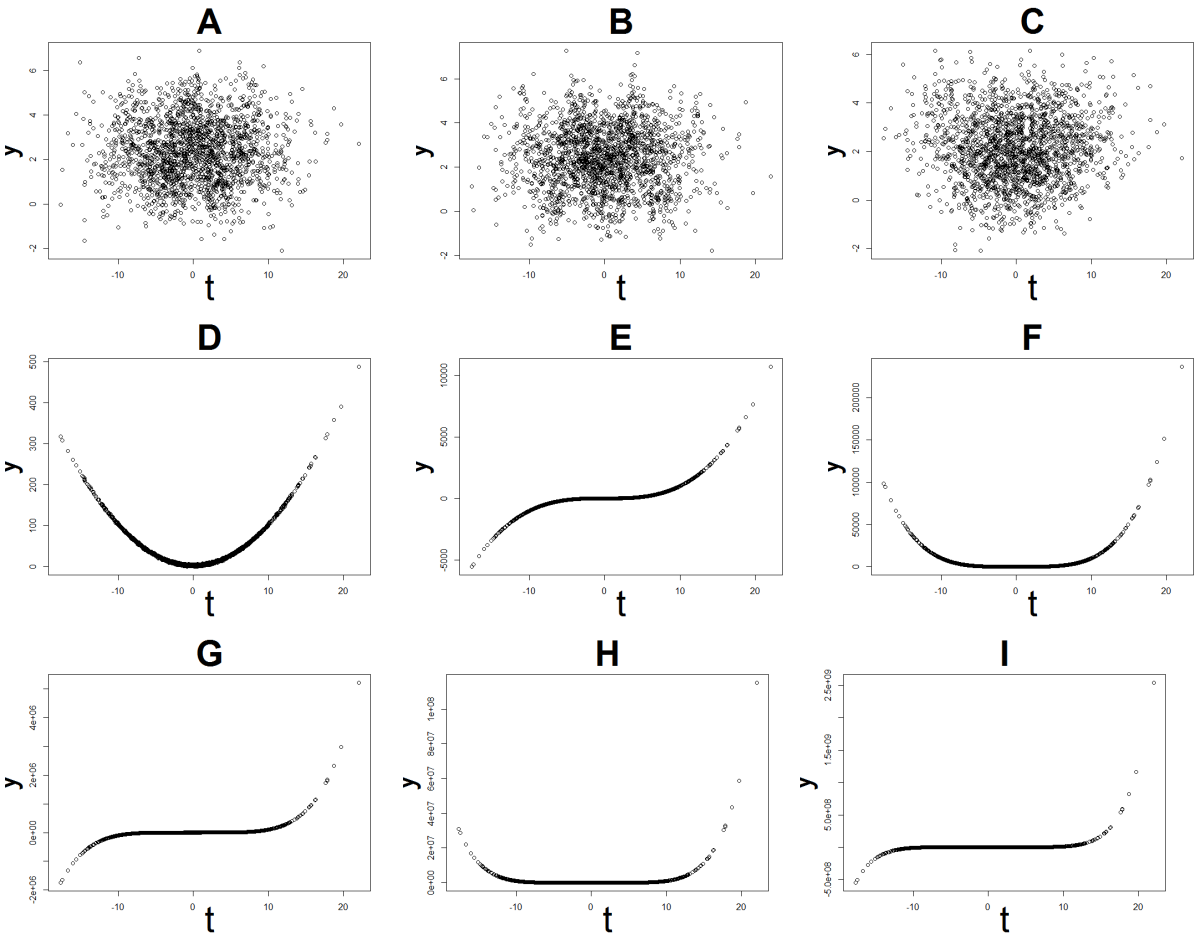

Figure 6: Plots of the residuals from least squares and the covariate with a nonlinear effect, $\mathbf{T}$, generated from Monte Carlo simulations testing H_0, $g = 0$, when H_0 is not true. The residuals come from fitting of linear model of $\mathbf{Y}$ on $\mathbf{X}$, where the response is generated as $\mathbf{Y} = 2 + \mathbf{X} + \mathbf{Z} + g(\mathbf{T})$, where $\mathbf{Z}$ ~N(0,1) and $\mathbf{T}$ ~N(0,36). Number of simulations run each time is 1,000. The sample size is 100 and for the ANOVA-type test the window width was 5. The form of $g()$ in each graph is as follows: (A) $g(\mathbf{T}) = cos(.4\mathbf{T})^4$; (B) $g(\mathbf{T}) = sin(.4\mathbf{T})^4$; (C) $g(\mathbf{T}) = sin(.2\mathbf{T})^4$; (D) $g(\mathbf{T}) = \mathbf{T}^2$; (E) $g(\mathbf{T}) = \mathbf{T}^3$; (F) $g(\mathbf{T}) = \mathbf{T}^4$; (G) $g(\mathbf{T}) = \mathbf{T}^5$; (H) $g(\mathbf{T}) = \mathbf{T}^6$; and (I) $g(\mathbf{T}) = \mathbf{T}^7$.

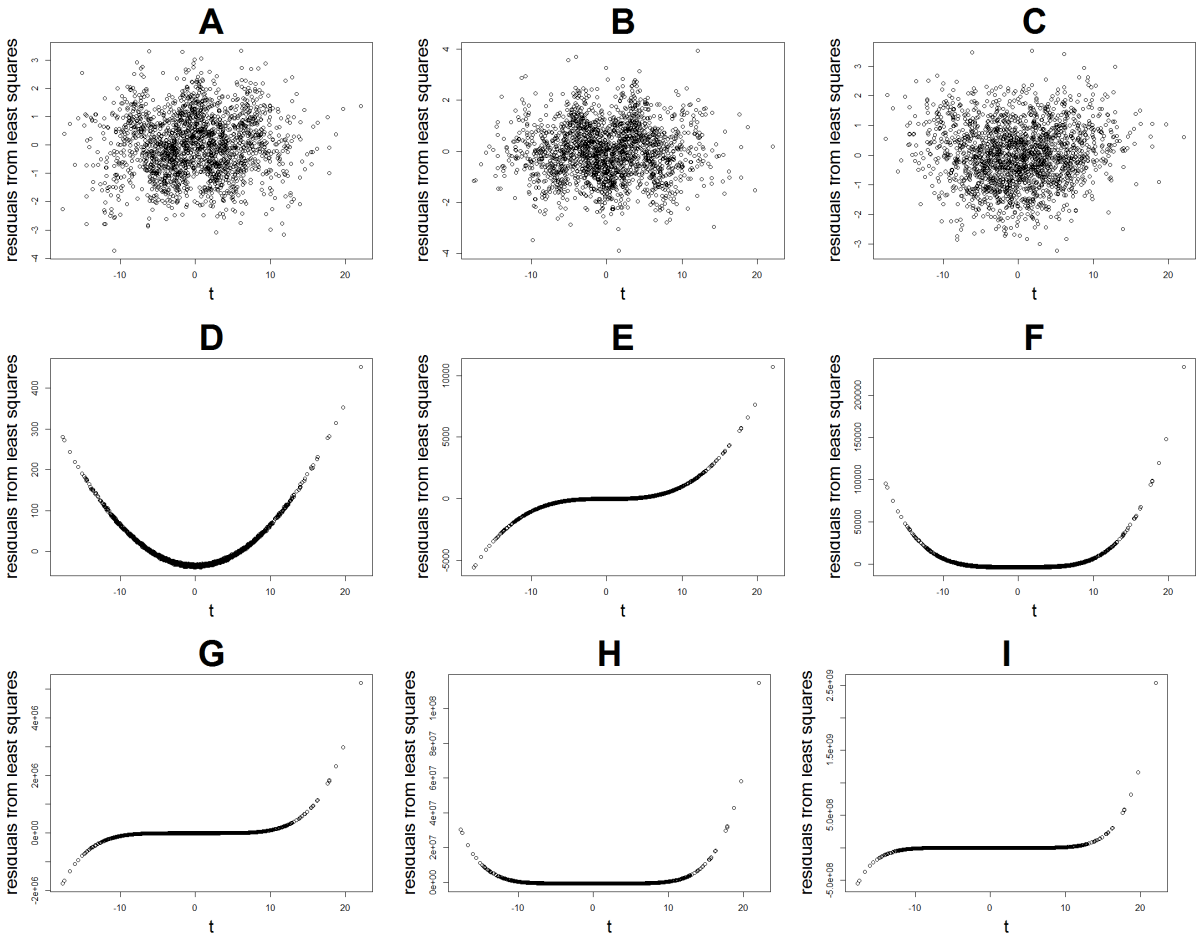

Figure 7: Plots from the UK spirit consumption data (dataset from Durbin and Watson (1951)). The top plot shows the consumption of spirits versus the covariate with the nonlinear effect, year. The bottom plot shows the residuals from least squares regression versus the covariate with the nonlinear effect, year.

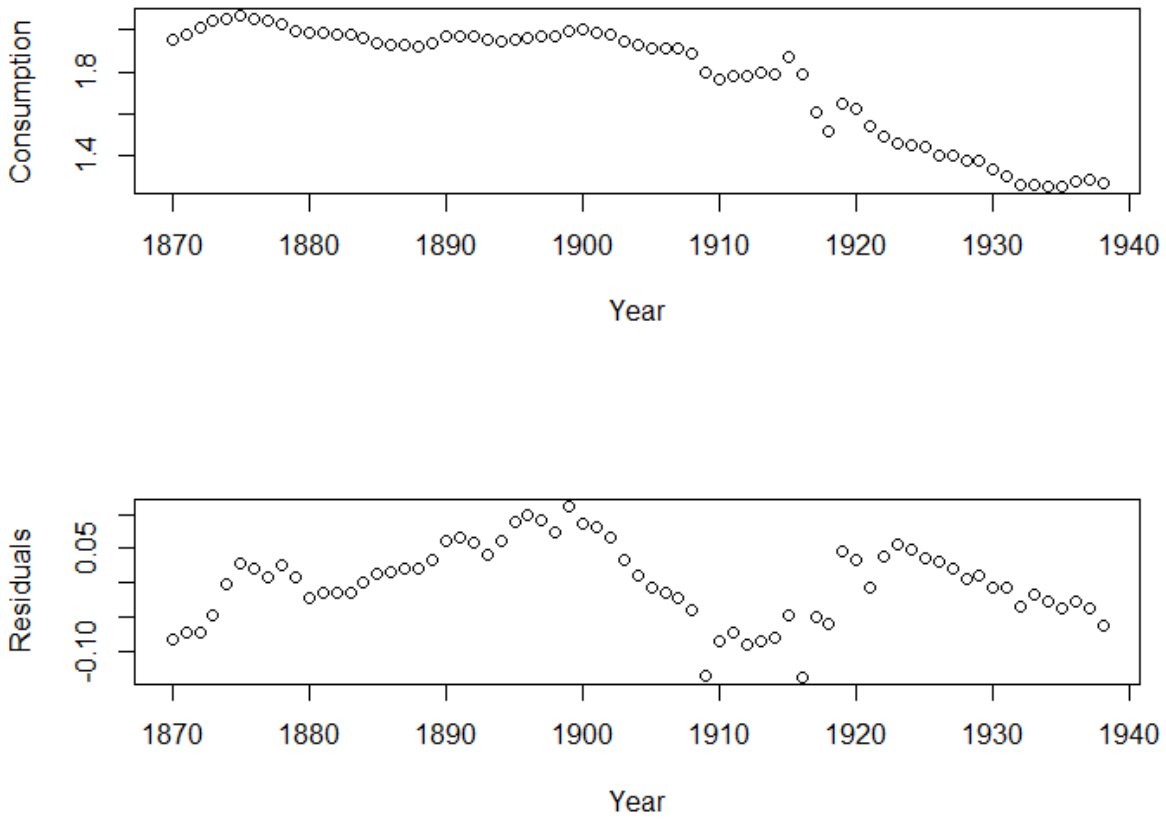

Figure 8: Plots from the net ecosystem CO_2 exchange data. The top plot shows the net ecosystem exchange of CO_2 versus the covariate with the nonlinear effect, year which has been standardized following the methodology of Li and Nie (2007, 2008). The bottom plot shows the residuals from least squares regression versus the covariate with the nonlinear effect, standardized year.