Covariate-Adjusted Regression

By DAMLA ŞENTÜRK AND HANS-GEORG MÜLLER

Department of Statistics, University of California, Davis, CA 95616, USA

dsenturk@wald.ucdavis.edu mueller@wald.ucdavis.edu

Summary

We introduce covariate-adjusted regression for situations where both predictors and response in a regression model are not directly observable, but are contaminated with a multiplicative factor that is determined by the value of an unknown function of an observable covariate. We demonstrate how the regression coefficients can be estimated by establishing a connection to varying-coefficient regression. The proposed covariate adjustment method is illustrated with an analysis of the regression of plasma fibrinogen concentration as response on serum transferrin level as predictor for 69 haemodialysis patients. In this example, both response and predictor are thought to be influenced in a multiplicative fashion by body mass index. A bootstrap hypothesis test enables us to test the significance of the regression parameters. We establish consistency and convergence rates of the parameter estimators for this new covariate-adjusted regression model. Simulation studies demonstrate the efficacy of the proposed method.

Some key words: Bootstrap; Diagnostics; Linear regression; Multiplicative effects; Smoothing; Varying-coefficient model.

1. Introduction

1.1 Preamble

We address the problem of parameter estimation in multiple regression when the actual predictors and response are not observable. Instead, one observes contaminated versions of these variables, where the distortion is multiplicative, with a factor that is a smooth unknown function of an observed covariate. The simultaneous dependence of response and predictors on the same covariate may lead to artificial correlation and regression relationships which do not exist between the actual hidden predictor and response variables. An example is the fibrinogen data of Kaysen et al. (2003), where the regression of fibrinogen level on serum transferrin level in haemodialysis patients is of interest. Both observed response and predictor are known to depend on body mass index, defined as weight/height^2, which thus has a confounding effect on the regression relation. The theme of this paper is to explore such confounding in regression and to develop appropriate adjustment methods.

1.2 Proposed covariate-adjusted regression model

Consider the simple linear regression model

$$Y_i = \gamma_0 + \gamma_1 X_i + e_i, \tag{1}$$

for the data (X_i, Y_i), $i = 1, \ldots, n$, where Y_i is the response for the ith subject in the sample, X_i is the predictor, e_i is an error term, and γ_0 and γ_1 are unknown parameters. A departure from the usual regression model is that X_i and Y_i are not observable. Instead one observes distorted versions $(\tilde{X}_i, \tilde{Y}_i)$, along with a univariate covariate U_i, where

$$\tilde{Y}_i = \psi(U_i) Y_i, \qquad \text{and} \qquad \tilde{X}_i = \phi(U_i) X_i, \tag{2}$$

and $\psi(\cdot)$ and $\phi(\cdot)$ are unknown smooth functions of the covariate U. For the above-mentioned fibrinogen data the confounding variable U is body mass index. In medical studies variables are commonly normalised by dividing them by a confounder such as body mass, implicitly assuming that the relationship between the confounder and the unobserved underlying variable is of a multiplicative nature. Equation (2) extends this to a more flexible and general multiplicative confounding model.

Identifiability constraints for $\psi(\cdot)$ and $\phi(\cdot)$ are implied by the natural assumption that the mean distorting effect should correspond to no distortion, i.e.

$$E\{\psi(U)\} = 1, \qquad E\{\phi(U)\} = 1. \tag{3}$$

We also assume that $(X_i, U_i, e_i)_{i=1,\ldots,n}$ are independent and identically distributed where $E(e_i) = 0$, $\text{var}(e_i) = \sigma^2$, and X, e and U are mutually independent. A central goal is to obtain consistent estimators of the regression coefficients in model (1), given the observations of the confounding variable U_i and the distorted observations $(\tilde{X}_i, \tilde{Y}_i)$ in (2). Under the identifiability conditions (3), given a consistent estimator $\hat{\gamma}_1$ of γ_1, the estimator $\hat{\gamma}_0 = n^{-1}\sum_i \tilde{Y}_i - \hat{\gamma}_1 n^{-1}\sum_i \tilde{X}_i$ will be consistent for γ_0. Thus it suffices to consider the estimation problem for γ_1 only. We refer to model (1) - (3) as the multiplicative distortion model or covariate-adjusted regression model.

1.3 Other distortion models

Adjustment for confounding variables per se is a classical problem. We start by investigating a sequence of nested models, for all of which standard adjustment methods already exist.

First, consider model (1) with additive instead of multiplicative distorting effects, i.e. $\tilde{Y} = Y + \psi_a(U)$ and $\tilde{X} = X + \phi_a(U)$. The identifiability constraints here are $E\{\psi_a(U)\} = E\{\phi_a(U)\} = 0$, for the distorting effects of U to average out to 0. A simple adjustment method for the consistent estimation of γ_1 in the additive distortion model is to use an estimator of the slope α_1 obtained by regressing $\tilde{e}_{\tilde{Y}|U}$ on $\tilde{e}_{\tilde{X}|U}$ by least squares, where $\tilde{e}_{W_1|W_2}$ is the set of residuals from the nonparametric regression of W_1 on W_2. However, as is shown in the Appendix, under (1)-(3), the estimator of α_1 is targeting the value

$$\xi_1 = \gamma_1 \Delta, \tag{4}$$

for $\Delta = E\{\psi(U)\phi(U)\}/E\{\phi^2(U)\}$, where Δ and therefore ξ_1 can assume any real value. Thus, while this simple adjustment works for the special case of an additive distortion model, it fails for the multiplicative distortion in the covariate-adjusted regression model.

The second model we consider is a special case of the additive effects model, where the distorting functions $\psi_a(\cdot)$ and $\phi_a(\cdot)$ are linear functions of U. In this case, a consistent

estimator of α_1 in the regression model $\tilde{Y} = \alpha_0 + \alpha_1 \tilde{X} + \alpha_2 U + e$ will also be consistent for γ_1 in model (1). This simple adjustment method however fails for the covariate-adjusted regression model, since, under (1)-(3), the target value ξ_2 of the estimator of α_1 will generally not satisfy $\xi_2 = \gamma_1$. Indeed it holds that $\xi_2 = \xi_1$, where ξ_1 is as given in (4); see the Appendix.

As a third model that is nested in all of the above models we consider the case of no distorting effect. This amounts to $\psi(\cdot) = \phi(\cdot) = 1$ in the covariate-adjusted regression model, and $\psi_a(\cdot) = \phi_a(\cdot) = 0$ in the additive model. In this case we would simply regress $\tilde{Y}$ on $\tilde{X}$, and use the slope estimator as a substitute for the estimator of γ_1. It is shown in the Appendix that, under (1)-(3), the slope estimator obtained from this regression model is targeting the value

$$\xi_3 = \frac{E\{\phi(U)\psi(U)\}\{\gamma_0 E(X) + \gamma_1 E(X^2)\} - \gamma_1\{E(X)\}^2 - \gamma_0 E(X)}{E\{\phi^2(U)\}E(X^2) - \{E(X)\}^2} \tag{5}$$

instead of γ_1, and that ξ_3 can assume any real value. Therefore, arbitrarily large biases may result if the confounding covariate is ignored within the covariate-adjusted regression model.

Fourthly, applying logarithmic transformations to $\tilde{Y}$ and $\tilde{X}$ to change the effect of the distortion functions $\psi(\cdot)$ and $\phi(\cdot)$ from multiplicative to additive also fails in the framework of the covariate-adjusted regression model, as it destroys the linearity of the model. Problems encountered when transforming multiplicative error regression models have been studied in Eagleson & Müller (1997).

Our proposed covariate-adjusted regression model also has similarities with multiplicative measurement error models where the error affects both the predictors and the response. Hwang (1986) derived consistent estimators for the regression coefficients under multiplicative measurement error in the predictors. Other estimation methods in this setting have been proposed by Iturria et al. (1999). However, the case of multiplicative measurement errors that affect both the predictors and the response has not been considered previously to our knowledge, and, furthermore, in the covariate-adjusted regression model the multiplicative errors affecting predictors and response are functions of an observed covariate U.

1.4 A motivating example

Assume the following simple linear regression model:

$$Y = 3 + X + e, \tag{6}$$

where $e \sim \mathcal{N}(0, 0.25)$ and $X \sim \mathcal{N}(1, 0.81)$. Assume that the distortion variable is $U \sim \mathrm{Un}(1, 7)$ and the distortion functions are $\psi(U) = (U - 0.5)^2/15.25$ and $\phi(U) = (U + 1)^2/28$, which satisfy the identifiability conditions. Then 1000 samples of $\tilde{Y}$ and $\tilde{X}$ were simulated from the specified distributions with sample size 400. For each sample, γ_1 was estimated using covariate-adjusted regression by applying estimators (11) and (12) from §3. In addition, the three simple adjustment methods introduced in §1.3 were applied, namely using an estimator of the slope α_1 from the regression models $\tilde{Y} = \alpha_0 + \alpha_1\tilde{X} + e$, $\tilde{Y} = \alpha_0 + \alpha_1\tilde{X} + \alpha_2 U + e$ and $\tilde{e}_{\tilde{Y}|U} = \alpha_0 + \alpha_1\tilde{e}_{\tilde{X}|U} + e$ as a substitute for the estimator of γ_1. The estimated biases were 0.0006, 1.1450, 0.1335 and 0.0850 for the estimators using covariate-adjusted regression and the three simple adjustment methods, respectively.

Note that the target bias values for the three adjustment methods are $|\xi_1 - \gamma_1| = 1.1460$, $|\xi_3 - \gamma_1| = 0.0841$ and $|\xi_2 - \gamma_1| \doteq |\xi_3 - \gamma_1|$, since $\psi(\cdot)$ and $\phi(\cdot)$ are close to linear in the interval $(1, 7)$. These results, along with the plots of the original variables Y versus X and the observed variables $\tilde{Y}$ versus $\tilde{X}$, given in Fig. 1, demonstrate that the distortion fundamentally changes the relationship between Y and X and that simple adjustment methods are not feasible.

2. Connection with varying-coefficient models

Consider the multiple regression model

$$Y = \gamma_0 + \sum_{r=1}^{p} \gamma_r X_r + e, \tag{7}$$

with predictors $X_1, \ldots, X_p$, response Y and error e. The observed variables that one has for (7) are U and

$$\tilde{X}_r(U) = \phi_r(U)X_r \qquad \text{and} \qquad \tilde{Y}(U) = \psi(U)Y,$$

for $r = 1, \ldots, p$, where ϕ_r and ψ are unknown smooth functions of U. If we write $\tilde{X} =$

$(\tilde{X}_1, \ldots, \tilde{X}_p)^{\mathrm{T}}$, the regression of the observed response on the observed predictors leads to

$$\begin{aligned} E(\tilde{Y}|\tilde{X}, U) &= E\{Y\psi(U)|\phi_1(U)X_1, \ldots, \phi_p(U)X_p, U\} \\ &= \psi(U)E\left\{\gamma_0 + \sum \gamma_r X_r + e|\phi_1(U)X_1, \ldots, \phi_p(U)X_p, U\right\}. \end{aligned}$$

If we assume that $E(e) = 0$ and that $(e,\, U,\, X_r)$ are mutually independent, for $r = 1, \ldots, p$, this reduces to

$$\begin{aligned} E(\tilde{Y}|\tilde{X}, U) &= \psi(U)\gamma_0 + \psi(U)\sum \gamma_r \frac{\phi_r(U)X_r}{\phi_r(U)} \\ &= \beta_0(U) + \sum \beta_r(U)\tilde{X}_r, \end{aligned} \tag{8}$$

where

$$\beta_0(u) = \psi(u)\gamma_0, \qquad \beta_r(u) = \gamma_r \frac{\psi(u)}{\phi_r(u)}. \tag{9}$$

Therefore,

$$\tilde{Y} = \beta_0(U) + \sum \beta_r(U)\tilde{X}_r + \psi(U)e, \tag{10}$$

which is a multiple varying-coefficient model; that is an extension of regression models where the coefficients are allowed to vary as a smooth function of a third variable (Hastie & Tibshirani, 1993). A unique feature is that in model (10) both response and predictors depend on the covariate U.

For varying-coefficient models, Hoover et al.(1998) have proposed smoothing methods based on local least squares and smoothing splines, and recent approaches include a componentwise kernel method (Wu & Chiang, 2000), a componentwise spline method (Chiang et al., 2001) and local maximum likelihood estimators (Cai et al., 2000). Wu & Yu (2002) provide a review of recent developments. We develop a consistent estimation method that is tailored to the special features of our model.

3. Estimation and consistency

The available data are of the form $(U_i, \tilde{X}_i, \tilde{Y}_i)$, $i = 1, \ldots, n$, for a sample of size n, where $\tilde{X}_i = (\tilde{X}_{1i}, \ldots, \tilde{X}_{pi})^{\mathrm{T}}$ are the p-dimensional predictors. To estimate the smooth varying-coefficient functions $\beta_0(\cdot), \ldots, \beta_p(\cdot)$ in (10), we use local smoothing methods based on an initial binning step. The binning is motivated by similar developments for longitudinal

data in Fan & Zhang (2000), who use the data collected at each fixed time point to fit a linear regression, obtaining the raw estimators for the smooth varying-coefficient functions. Generalising this idea to our independent and identically distributed data scheme, we partition the support of U into a number of bins, within which the covariate U has nearly constant levels. We then use the observed data $(\tilde{X}_i, \tilde{Y}_i)$ within each bin to fit linear regressions and to obtain raw estimators of the smooth varying-coefficient functions that contain the targeted regression parameters γ. Averaging these raw estimators over the bins with a special weighing scheme eliminates the influence of the contaminating functions of U, due to the identifiability conditions, leading to the targeted regression parameters γ.

We assume that the covariate U is bounded below and above, $-\infty < a \leq U \leq b < \infty$, for real numbers $a < b$, and divide the interval $[a, b]$ into m equidistant intervals denoted by $B_1, \ldots, B_m$, and referred to as bins. Given m, the B_j, $j = 1, \ldots, m$ are fixed, but the number of U_i's falling into B_j is random and is denoted by L_j. Let $\{(U'_{jk}, \tilde{X}'_{rjk}, \tilde{Y}'_{jk}),\ k = 1, \ldots, L_j, r = 1, \ldots, p\} = \{(U_i, \tilde{X}_{ri}, \tilde{Y}_i), i = 1, \ldots, n, r = 1, \ldots, p : U_i \in B_j\}$ denote the data for which $U_i \in B_j$, where we refer to $(U'_{jk}, \tilde{X}'_{rjk}, \tilde{Y}'_{jk})$ as the kth element in the jth bin B_j. Further define $(U'_j, \tilde{X}'_j, \tilde{Y}'_j)$ to be the data matrix belonging to the jth bin, where $U'_j = (U'_{j1}, \ldots, U'_{jL_j})^{\mathrm{T}}$, $\tilde{Y}'_j = (\tilde{Y}'_{j1}, \ldots, \tilde{Y}'_{jL_j})^{\mathrm{T}}$ and $\tilde{X}'_{jk} = (1, \tilde{X}'_{1jk}, \ldots, \tilde{X}'_{pjk})^{\mathrm{T}}$ for $k = 1, \ldots, L_j$ contains p components of the kth element in bin B_j, and $\tilde{X}'_j = (\tilde{X}'_{j1}, \ldots, \tilde{X}'_{jL_j})^{\mathrm{T}}_{L_j \times (p+1)}$.

After binning the data, we fit a linear regression of $\tilde{Y}'_j$ on $\tilde{X}'_j$ within each bin B_j, $j = 1, \ldots, m$. The least squares estimators of the multiple regression of the data in the jth bin are

$$\hat{\beta}_j = (\hat{\beta}_{0j}, \ldots, \hat{\beta}_{pj})^{\mathrm{T}} = (\tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j \tilde{Y}'_j.$$

Our proposed estimators of γ_0 and γ_r, for $r = 1, \ldots, p$, are then obtained as weighted averages of the $\hat{\beta}_j$'s, weighted according to the number of data L_j in the jth bin,

$$\hat{\gamma}_0 = \sum_{j=1}^{m} \frac{L_j}{n} \hat{\beta}_{0j} \tag{11}$$

$$\hat{\gamma}_r = \frac{1}{\hat{\mu}_{\tilde{X}_r}} \sum_{j=1}^{m} \frac{L_j}{n} \hat{\beta}_{rj} \hat{\mu}_{\tilde{X}'_{rj}}, \tag{12}$$

where $\hat{\mu}_{\tilde{X}_r} = n^{-1}\sum_{i=1}^{n}\tilde{X}_{ri}$ and $\hat{\mu}_{\tilde{X}'_{rj}} = L_j^{-1}\sum_{k=1}^{L_j}\tilde{X}'_{rjk}$. These estimators are motivated by $E\{\beta_0(U)\} = \gamma_0$ and $E\{\beta_r(U)\tilde{X}_r\} = \gamma_r E\{\psi(U)X_r\} = \gamma_r E(X_r) = \gamma_r E(\tilde{X}_r)$; see (3) and (9).

The $\hat{\beta}_j$ are the raw estimators for $\beta(U_j^M) = \{\beta_0(U_j^M), \ldots, \beta_p(U_j^M)\}^{\mathrm{T}}$, for midpoints $U_j^M = a + (2j-1)\{(b-a)/(2m)\}$ of the bins B_j. Since these raw estimators are not necessarily smooth, smooth estimators of the coefficient functions $\beta_r(\cdot)$, $r = 0, \ldots, p$ (10) are obtained by smoothing the scatterplot $\{(U_j^M, \hat{\beta}_{rj}), j = 1, \ldots, m\}$ for each component r, $1 \le r \le p$. If a linear smoother with weight functions $w_j(\cdot)$ is used, we obtain the linear smooth estimator

$$\tilde{\beta}_r(u) = \sum_{j=1}^{m}\hat{\beta}_{rj}w_j(u) \tag{13}$$

for $\beta_r(\cdot)$. The smooth estimators $\tilde{\beta}_r(\cdot)$ are used in the bootstrap test proposed in §4.

Next, we show the consistency of estimators $\hat{\gamma}_0$ and $\hat{\gamma}_r$ for γ_0 and γ_r in model (7), when the number of subjects n tends to infinity. As is typical for smoothing, the number of bins $m = m(n)$ is required to satisfy $m \to \infty$ and $n/(m\log n) \to \infty$ as $n \to \infty$.

For the estimators given in (11) and (12) to be well defined, the least squares estimator $\hat{\beta}_j$ must exist for each bin B_j. This requires that the inverse of $\tilde{X}_j'^{\mathrm{T}}\tilde{X}'_j$ be well defined, i.e. $\det(\tilde{X}_j'^{\mathrm{T}}\tilde{X}'_j) \neq 0$. Define the event

$$A = \{\omega \in \Omega : \inf_j |\det(\tilde{X}_j'^{\mathrm{T}}\tilde{X}'_j)| > 0\}, \tag{14}$$

where $(\Omega, \mathcal{F}, P)$ is the underlying probability space. On event A, $\hat{\gamma}_0$ and $\hat{\gamma}_r$ are well defined. It is shown in the Appendix that $\mathrm{pr}(A) \to 1$ as $n \to \infty$.

Theorem 1. *Under the technical conditions given in the Appendix, given event A,*

$$\hat{\gamma}_r \;=\; \gamma_r + O_p(n^{-1/2}) + O(m^{-1}), \qquad\qquad r = 0, \ldots, p.$$

4. Bootstrap test

It is often of interest to test for the significance of the regression coefficients. Equation (9) shows that $\gamma_r = 0$ is equivalent to $\beta_r(\cdot) = 0$, for $r = 0, \ldots p$, whenever $\psi(\cdot)$ and $\phi_r(\cdot)$ satisfy the identifiability conditions. Thus, testing $H_0 : \beta_r(\cdot) = 0$ is equivalent to testing $H_0 : \gamma_r = 0$. Testing $H_0 : \beta_r(\cdot) = 0$ is a special case of testing the 'no-effect' hypothesis, i.e. testing $H_0 : \beta_r(U) = c$ for a real c (Hart, 1997, p. 140).

Under the null hypothesis, the smooth estimator $\tilde{\beta}_r(\cdot)$ in (13) of $\beta_r(\cdot)$ is expected to be close to a horizontal line through zero. Reasonable test statistics quantify departures of the smooth estimator from this line. Similarly to the statistics proposed by Hart, we adopt, as a measure of departure,

$$R_n = \frac{1}{m}\sum_{j=1}^{m} |\tilde{\beta}_r(U_j^M; h_T)|,$$

where $\tilde{\beta}_r(U_j^M; h_T)$ is the linear smooth, fitted using the bandwidth h_T, evaluated at U_j^M.

For an automatic data-based choice of the bandwidth parameter h_T, we define

$$h_T = \arg\min_h\{T(h)\} = \arg\min_h\left\{\frac{(1/m)RSS(h)}{1-2tr(W_h)/m}\right\}, \tag{15}$$

(Rice, 1984), where W_h is an $m \times m$ matrix with (ℓ, j)th element $w_j(U_\ell^M; h)$ and $RSS(h) = \|\hat{\beta}_r - \tilde{\beta}_r\|^2$ for $\hat{\beta}_r = (\hat{\beta}_{r1}, \ldots, \hat{\beta}_{rm})^{\mathrm{T}}$, $\tilde{\beta}_r = (\tilde{\beta}_r(U_1^M), \ldots, \tilde{\beta}_r(U_m^M))^{\mathrm{T}}$. This criterion allows for fast implementation and led to good results.

The raw estimators $\hat{\beta}_{rj}$ are heteroscedastic, since the density function of U is in general not uniform. For this reason, the sampling distribution of R_n is approximated using the wild bootstrap. The bootstrap samples are obtained under the null hypothesis, and have the form $[\{U_1^M, (\hat{\beta}_{r1} - \hat{\mu}_{\hat{\beta}_r})V_1\}, \ldots, \{U_m^M, (\hat{\beta}_{rm} - \hat{\mu}_{\hat{\beta}_r})V_m\}]$, where $\hat{\mu}_{\hat{\beta}_r} = m^{-1}\sum_j \hat{\beta}_{rj}$ and V_j is sampled from the two-point distribution attaching masses $(\sqrt{5}+1)/2\sqrt{5}$ and $(\sqrt{5}-1)/2\sqrt{5}$ to the points $-(\sqrt{5}-1)/2$ and $(\sqrt{5}+1)/2$ (Davison & Hinkley, 1997, p. 272). The variables V_j defined in this way have means equal to zero, and variances and third moments equal to one. Variables $\{(\hat{\beta}_{rj} - \hat{\mu}_{\hat{\beta}_r})V_j\}$ have means zero, and crudely approximate the variance and skewness of the underlying distribution. The distribution of R_n^* computed from the bootstrap samples is used as an approximation to the distribution of R_n.

5. Application

Fibrinogen is a risk factor for cardiovascular disease, and its plasma concentration increases with inflammation. It is of particular interest to elucidate the relationship between this acute phase protein and other plasma proteins such as transferrin, ceruloplasmin and acid glycoprotein for haemodialysis patients. This motivated a study of seventy haemodialysis patients (Kaysen et al., 2003) where the main tool was linear regression of plasma

fibrinogen concentration, *FIB*, against various predictors, which included the serum transferrin level, *TRF*. A simple linear regression model would be $FIB = \gamma_0 + \gamma_1 TRF + e$, where e is the error term. Body mass index, BMI= weight/height^2, was considered to be a major confounding factor for both response and predictor. We applied the covariate-adjusted regression model, (8), (9), using body mass index as the confounder U.

The parameters γ_0 and γ_1 were estimated by the covariate-adjusted regression algorithm and the results were compared to the estimators obtained from the least squares regression of the observed *FIB* on observed *TRF*. One outlier was removed before the analysis. The estimates and p-values for the significance of the parameters for both methods are given in Table 1. The p-values for covariate-adjusted regression estimates were obtained from the bootstrap test proposed above, using the empirical percentiles of R_n^* from 1000 bootstrap samples.

Estimated coefficient functions $\tilde{\beta}_0(\cdot)$ and $\tilde{\beta}_1(\cdot)$, obtained by local linear smoothing using bandwidth choices given in (15), are displayed in Fig. 2 together with the raw estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. We tested whether or not the covariate-adjusted regression model is more appropriate for the data than the additive effects model by testing whether or not $\beta_1(\cdot)$ was equal to a constant, as discussed below in §6. The p-value of 0.07 from this test, and the increasing pattern of $\tilde{\beta}_1(\cdot)$, provide evidence that $\beta_1(\cdot)$ is not constant for these data, so that the covariate-adjusted regression model is preferred. Bin widths were chosen such that the average number of points falling in each bin is $p + 1$, enough to fit the linear regression, where p is the number of parameters of the regression model. Bins with fewer than $p + 1$ elements were merged with neighbouring bins.

For least squares regression, TRF was close to being significant, $p = 0.101$, while with covariate-adjusted regression it became highly significant, $p = 0.002$, with an increasing trend in $\tilde{\beta}_1(\cdot)$. As BMI increases, the negative slope of serum transferrin level as predictor for plasma fibrinogen level approaches zero, while the intercept declines. The effects of BMI are thus masking the true overall negative effect that TRF has on FIB in the unadjusted regression. It is believed that high fibrinogen levels are caused by inflammation and stimulation of albumin synthesis (Kaysen et al., 2003). While in least squares

regression modelling transferrin was not among the factors that had significant effects on fibrinogen levels, our analysis with covariate-adjusted regression indicates that there is a strong negative association if BMI is taken into account.

6. Model diagnostics and simulation study

Consider the three alternative distortion models that were discussed in §1.3. We note that the general adjustment method provided by covariate-adjusted regression works for these models as well, and in fact one of the attractions of the proposed adjustment is that the specific nature of the distortion of the variables need not be known. Nevertheless, in applications it may be of interest to investigate whether any of these models approximates the data sufficiently well, in which case the corresponding simpler adjustment could be implemented to obtain consistent estimation of the regression coefficients in (1).

If we focus on the simple linear regression case, in the additive effects model, $\tilde{Y}$ is related to $\tilde{X}$ and U through a partial linear model (Heckman, 1986), $\tilde{Y} = \gamma_0 + \gamma_1\tilde{X} + \upsilon(U) + e$, where $\upsilon(U) = \psi_a(U) - \gamma_1\phi_a(U)$. This partial linear model is a special case of the varying-coefficient model associated with the covariate-adjusted regression model, $\tilde{Y} = \beta_0(U) + \beta_1(U)\tilde{X} + \psi(U)e$, where the smooth coefficient function $\beta_1(U)$ is constant, $\beta_1(U) = \gamma_1$, and $\beta_0(U) = \upsilon(U) + \gamma_0$. If $\beta_1(U)$ is not constant, then this implies that covariate-adjusted regression is more appropriate for the data than the additive distortion model. Otherwise, if $\beta_1(U)$ is constant, this implies by (9) and the identifiability conditions that $\psi(U) = \phi(U)$ in the covariate-adjusted regression model. In this case, the adjustment method proposed in §1.3 for the additive distortion model can be used for consistent estimation regardless of which model is providing the best fit, since ξ_1 in (4) equals γ_1. Thus, one way of testing if the covariate-adjusted regression model is more appropriate for the data than the additive model is to test whether or not $\beta_1(\cdot)$ is equal to a constant, which is the 'no effect' test mentioned in §4. This test is equivalent to testing the null hypothesis $\beta_1(\cdot) = 0$ after the sample is centred around zero, and one could carry out the bootstrap test for $H_0 : \beta_1(\cdot) = 0$ with the data $\hat{\beta}_1 - \hat{\mu}_{\hat{\beta}_1}$ rather than $\hat{\beta}_1$.

To check if the additive model reduces to the model with linear distortion functions $\psi_a(\cdot)$ and $\phi_a(\cdot)$, leading to the multiple regression relationship $\tilde{Y} = \alpha_0 + \alpha_1 U + \gamma_1\tilde{X} + e$,

it is enough to test if $\beta_0(\cdot)$ is equal to a linear function; see Hart (1997) for suitable test statistics. For checking whether or not the model further reduces to the no effect case, $\tilde{Y} = \alpha_0 + \alpha_1 \tilde{X} + e$, one would check if $\beta_0(\cdot)$ is equal to a constant; see §5.

We carried out a simulation study to show the efficacy of the proposed adjustment method. The confounding covariate U was simulated from $\mathcal{N}(6, 1)$, truncated at two standard deviations. The underlying unobserved multiple regression model was

$$Y = 1 + 0.1X_1 + 2X_2 - 0.2X_3 + e,$$

where $X_1 \sim \mathcal{N}(2, 1.44)$, $X_2 \sim \mathcal{N}(0.5, 0.25)$, $X_3 \sim \mathcal{N}(1, 1)$, and $e \sim \mathcal{N}(0, 0.25)$. The distortion functions were chosen as $\psi(U) = (U + 3)^2/81.8090$, $\phi_1(U) = (U + 10)/16$, $\phi_2(U) = (U + 1)^2/49.8015$ and $\phi_3(U) = (U + 3)/9$, satisfying the identifiability conditions. We conducted 1000 Monte Carlo runs with sample sizes 70, 200, 400. The estimated mean squared errors for the estimators of γ_0, γ_1, γ_2 and γ_3 are (0.1192, 0.0493, 0.0235), (0.0165, 0.0067, 0.0031), (0.1029, 0.0295, 0.0170) and (0.0273, 0.0154, 0.0060), respectively, for the sample sizes n =(70, 200, 400) and number of bins $(11, 25, 50)$. These values are obtained after removing two outliers for each sample size. The cross-sectional means of the 1000 estimates $\tilde{\beta}_0(\cdot)$, $\tilde{\beta}_1(\cdot)$, $\tilde{\beta}_2(\cdot)$ and $\tilde{\beta}_3(\cdot)$ of the smooth coefficient functions $\beta_0(\cdot)$, $\beta_1(\cdot)$, $\beta_2(\cdot)$ and $\beta_3(\cdot)$ are shown in Fig. 3 for sample sizes $n = 70$ and $n = 400$. While the estimation is seen to work well in the interior, there is evidence of boundary bias near the endpoints of the range of U.

To examine the power of the proposed bootstrap test, we assume the model given in (6) with distortion functions $\psi(U) = (U + 3)^2/81.8090$ and $\phi(U) = (U + 3)/9$, satisfying the identifiability conditions when $U \sim \mathcal{N}(6, 1)$, truncated as above. The null hypothesis is $H_0 : \gamma_1 = 0$. Fig. 4 depicts power functions for the significance levels 0.05, 0.10 and 0.20, based on 1000 Monte Carlo runs with sample sizes $n = 70$ and $n = 400$. The observed type I errors at $\gamma_1 = 0$, for the above mentioned significance levels are, 0.055, 0.121 and 0.223 for $n = 70$, and 0.046, 0.089 and 0.197 for $n = 400$. The levels of the bootstrap test move closer to the target values and the power functions increase more rapidly as γ_1 moves away from zero when the sample size increases.

Acknowledgement

We are extremely grateful to an anonymous referee and the editor for many helpful remarks that improved the exposition of the paper. This research was supported in part by the National Science Foundation.

Appendix

Technical details

Technical conditions. The following assumptions are made.

Condition 1. The covariate U is bounded below and above: $-\infty < a \leq U \leq b < \infty$, for real numbers $a < b$. The density $f(u)$ of U satisfies $\inf_{a\leq u\leq b} f(u) > 0$, $\sup_{a\leq u\leq b} f(u) < \infty$, and is uniformly Lipschitz; that is there exists a real number M such that $\sup_{a\leq u\leq b} |f(u+c) - f(u)| \leq M|c|$ for any real number c.

Condition 2. The variables (e, U, X_r) are mutually independent for $r = 1, \ldots, p$.

Condition 3. For the predictors, $\sup_{1\leq i\leq n, 1\leq r\leq p} |X_{ri}| \leq B$ for some bound $B \in \mathbb{R}$.

Condition 4. Contamination functions $\psi(\cdot)$ and $\phi_r(\cdot)$, $1 \leq r \leq p$, are twice continuously differentiable, satisfying $E\psi(U) = 1, \quad E\phi_r(U) = 1, \quad \phi_r(\cdot) > 0 \quad 1 \leq r \leq p$.

Condition 5. As $n \to \infty$, $\frac{1}{n} X^{\mathrm{T}} X \to \mathcal{X}$ in probability, where the limiting $(p+1) \times (p+1)$ matrix $\mathcal{X}$ is nonsingular.

Condition 6. The function $h(u) = \int x g(x, u) dx$ is uniformly Lipschitz, where $g(\cdot, \cdot)$ is the joint density function of $\tilde{X}$ and U.

These are mild conditions that are satisfied in most practical situations. Bounded covariates are standard in asymptotic theory for least squares regression, as are Conditions 2 and 5 (Lai et al., 1979). The identifiability conditions, Condition 4, are equivalent to

$$E(\tilde{Y}|X) = E(Y|X), \qquad E(\tilde{X}_r|X_r) = X_r.$$

This means that the confounding of Y by U does not change the mean regression function, and the distorting effects of the confounding variable U average out to 0.

Proof that $\mathrm{pr}(A) \to 1$. For the event A as defined in (14), the following result leads to $\mathrm{pr}(\inf_j d_j > 0) \to 1$ as $n \to \infty$ for $d_j = \det(L_j^{-1} \tilde{X}_j'^{\mathrm{T}} \tilde{X}_j')$, which further implies that $\mathrm{pr}(A) \to 1$ as $n \to \infty$; $\mathcal{X}$ and U_j^M are as defined in Condition 5 and §3 respectively.

Lemma 1. *For a sequence r_n such that $r_n = O_p[\sqrt{\{(m \log n)/n\}}]$,*

$$\sup_j \left| d_j - \phi_1^2(U_j^M) \dots \phi_p^2(U_j^M) \det(\mathcal{X}) \right| = O_p(r_n).$$

Furthermore, $\inf_j \phi_1^2(U_j^M) \dots \phi_p^2(U_j^M) \det(\mathcal{X}) > 0$.

Proof. Define $\tilde{X}_{rj}^{\prime(\ell)} = L_j^{-1} \sum_{k=1}^{L_j} \tilde{X}_{rjk}^{\prime\ell}$, $(\tilde{X}_{rj}' \tilde{X}_{sj}')^{(\ell)} = L_j^{-1} \sum_{k=1}^{L_j} (\tilde{X}_{rjk}' \tilde{X}_{sjk}')^{\ell}$ and analogously for $X_{rj}^{\prime(\ell)}$ and $(X_{rj}' X_{sj}')^{(\ell)}$. Note that $L_j^{-1} \tilde{X}_j^{\prime \mathrm{T}} \tilde{X}_j' = (\chi_{rs})_{(p+1)\times(p+1)}$ for $r, s = 0, \dots, p$, where $\chi_{rs} = (\tilde{X}_{rj}' \tilde{X}_{sj}')^{(1)}$ and $\tilde{X}_{0j}' = 1$. Thus, $d_j = \sum (-1)^{\mathrm{sign}(\tau)} (L_j^{-1} \tilde{X}_j^{\prime \mathrm{T}} \tilde{X}_j')_{1\tau(1)} \dots$
$(L_j^{-1} \tilde{X}_j^{\prime \mathrm{T}} \tilde{X}_j')_{(p+1),\tau(p+1)}$, where the sum is taken over all permutations τ of $(1, \dots, p+1)$, and $\mathrm{sign}(\tau)$ equals $+1$ or -1, depending on whether τ can be written as the product of even or odd number of transpositions. The terms in the above sum have the general form

$$\tilde{X}_{r_1 j}^{\prime(1)} (\tilde{X}_{1j}' \tilde{X}_{r_2 j}')^{(1)} \dots (\tilde{X}_{pj}' \tilde{X}_{r_{p+1} j}')^{(1)}, \tag{A1}$$

where $(r_1, \dots, r_{p+1})$ is a permutation of $(0, \dots, p)$. Considering the definition of the Nadaraya-Watson estimator (Fan & Gijbels, 1996), we note that an arbitrary term in (A1) has the form $(\tilde{X}_{sj}' \tilde{X}_{r_{s+1} j}')^{(1)} = \hat{m}_{s r_{s+1}}(U_j^M)$ for $0 \le s \le p+1$, $K(\cdot) = (1/2)\mathrm{I}([-1, 1])$, $h = (b-a)/m$, and U_j^M as defined in Lemma 1 . Uniform consistency of Nadaraya-Watson estimators with kernels of compact support has been shown in Härdle et al.(1988),

$$\sup_{a \le u \le b} |\hat{m}_{s r_{s+1}}(u) - m_{s r_{s+1}}(u)| = O_p(r_n), \tag{A2}$$

where $m_{s r_{s+1}}(u) = E(\tilde{X}_s \tilde{X}_{r_{s+1}} | U = u) = \phi_s(u) \phi_{r_{s+1}}(u) E(X_s X_{r_{s+1}})$, and r_n is as defined in Lemma 1. Then (A2) implies that $\sup_j |\hat{m}_{s r_{s+1}}(U_j^M) - m_{s r_{s+1}}(U_j^M)| = O_p(r_n)$ and $\sup_j |(\tilde{X}_{sj}' \tilde{X}_{r_{s+1} j}')^{(1)} - \phi_s(U_j^M) \phi_{r_{s+1}}(U_j^M) E(X_s X_{r_{s+1}})| = O_p(r_n)$. Hence the uniform consistency of (A1) follows, where the limit of (A1) is $\phi_1^2(U_j^M) \dots \phi_p^2(U_j^M)\ E(X_{r_1}) E(X_1 X_{r_2}) \dots$ $E(X_p X_{r_{p+1}})$, and Lemma 1 holds.

Proof of Theorem 1. As $\tilde{X}_r$ is bounded, since $X_r = O(1)$, U has compact support and $\phi_r(\cdot)$ is continuous for $1 \le r \le p$, $\tilde{X}_{rj}'$ is also bounded for $1 \le r \le p$. Thus $\sup_{1 \le j \le m} |L_j^{-1} \tilde{X}_j^{\prime \mathrm{T}} \tilde{X}_j'| = O(1) 1_{(p+1)\times(p+1)}$, where $1_{(p+1)\times(p+1)}$ is a $(p+1) \times (p+1)$-dimensional matrix of ones. On event A, $\sup_{1 \le j \le m} |(L_j^{-1} \tilde{X}_j^{\prime \mathrm{T}} \tilde{X}_j')^{-1}| = O(1) 1_{(p+1)\times(p+1)}$ and $\sup_{1 \le j \le m} |(L_j^{-1} \tilde{X}_j^{\prime \mathrm{T}} \tilde{X}_j')^{-1} \tilde{X}_j'| = O(1) 1_{(p+1)\times 1}$. Defining $\delta_{0jk} = \psi(U_{jk}') - \psi(U_j^{\prime *})$, $\delta_{rjk} = \phi_r(U_{jk}') - \phi_r(U_j^{\prime *})$ and $\delta_{rjk}' = \psi(U_{jk}')/\phi_r(U_{jk}') - \psi(U_j^{\prime *})/\phi_r(U_j^{\prime *})$ for $1 \le k \le L_j$ and $1 \le r \le p$, where $U_j^{\prime *} = L_j^{-1} \sum_{k=1}^{L_j} U_{jk}'$, is the average of the U's in B_j, we obtain the

following results for $1 \leq r, s \leq p$, using Taylor expansions and boundedness considerations:
(a) $\sup_{k,j} |U'_{jk} - U'^{*}_{j}| \leq (b-a)/m$;
(b) $\sup_{k,j} |\delta_{0jk}| = O(m^{-1})$;
(c) $\sup_{k,j} |\delta_{rjk}| = O(m^{-1})$;
(d) $\sup_{k,j} |\delta'_{rjk}| = O(m^{-1})$;
(e) $\sup_{j} |L_j^{-1} \sum_{k=1}^{L_j} \{(L_j^{-1} \tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j\}_{rk} \delta_{0jk}| = O(m^{-1})$;
(f) $\sup_{j} |L_j^{-1} \sum_{k=1}^{L_j} \{(L_j^{-1} \tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j\}_{sk} \delta'_{rjk} \tilde{X}'_{rjk}| = O(m^{-1})$;
(g) $\sup_{j} |L_j^{-1} \sum_{k=1}^{L_j} X'^{\ell}_{rjk} \delta_{rjk}| = O(m^{-1})$, $1 \leq \ell \leq 2$;
(h) $\sup_{j} |L_j^{-1} \sum_{k=1}^{L_j} X'_{rjk} \delta_{0jk}| = O(m^{-1})$.

On event A, least squares estimators $\hat{\beta}_j$ are well defined:

$$\hat{\beta}_j^{\mathrm{T}} = (\tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j \begin{bmatrix} \beta_0(U'_{j1}) + \beta_1(U'_{j1}) \tilde{X}'_{1j1} + \ldots + \beta_p(U'_{j1}) \tilde{X}'_{pj1} + \epsilon(U'_{j1}) \\ \vdots \\ \beta_0(U'_{jL_j}) + \beta_1(U'_{j1L_j}) \tilde{X}'_{1jL_j} + \ldots + \beta_p(U'_{jL_j}) \tilde{X}'_{pj1L_j} + \epsilon(U'_{jL_j}) \end{bmatrix}$$

$$= (\tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j \left\{ \gamma_0 \psi(U'^{*}_j), \gamma_1 \frac{\psi(U'^{*}_j)}{\phi_1(U'^{*}_j)}, \ldots, \gamma_p \frac{\psi(U'^{*}_j)}{\phi_p(U'^{*}_j)} \right\}^{\mathrm{T}}$$

$$+ (\tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j (\delta_{0j1}\gamma_0 + \gamma_1 \delta'_{1j1} \tilde{X}'_{1j1} + \ldots + \gamma_p \delta'_{pj1} \tilde{X}'_{pj1},$$
$$\ldots, \delta_{0jL_j}\gamma_0 + \gamma_1 \delta'_{1jL_j} \tilde{X}'_{1jL_j} + \ldots + \gamma_p \delta'_{pjL_j} \tilde{X}'_{pjL_j})^{\mathrm{T}}$$

$$+ (\tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j \{\psi(U'_{j1}) e'_{j1}, \ldots, \psi(U'_{jL_j}) e'_{jL_j}\}^{\mathrm{T}}. \tag{A3}$$

If we substitute $L_j^{-1}(L_j^{-1} \tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j$ in place of $(\tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j$, and use (A3), $\hat{\gamma}_r$ becomes

$$\frac{\gamma_r}{\hat{\mu}_{\tilde{X}_r}} \sum_{j=1}^{m} \frac{L_j}{n} \hat{\mu}_{\tilde{X}'_{rj}} \frac{\psi(U'^{*}_j)}{\phi_r(U'^{*}_j)}$$

$$+ \frac{1}{\hat{\mu}_{\tilde{X}_r}} \sum_{j=1}^{m} \frac{L_j}{n} \hat{\mu}_{\tilde{X}'_{rj}} L_j^{-1} \sum_{k=1}^{L_j} \{(L_j^{-1} \tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j\}_{rk} (\gamma_0 \delta_{0jk} + \gamma_1 \delta'_{1jk} + \ldots + \gamma_p \delta'_{pjk} \tilde{X}'_{pjk})$$

$$+ \frac{1}{\hat{\mu}_{\tilde{X}_r}} \sum_{j=1}^{m} \frac{L_j}{n} \hat{\mu}_{\tilde{X}'_{rj}} L_j^{-1} \sum_{k=1}^{L_j} \{(L_j^{-1} \tilde{X}'^{\mathrm{T}}_j \tilde{X}'_j)^{-1} \tilde{X}'^{\mathrm{T}}_j)\}_{rk} \psi(U'_{jk}) e'_{jk} = T_1 + T_2 + T_3.$$

By property (g) above, $\hat{\mu}_{\tilde{X}'_{rj}} = \phi_r(U'^{*}_j) \bar{X}'_{rj} + L_j^{-1} \sum_{k=1}^{L_j} X'_{rjk} \delta_{rjk} = \phi_r(U'^{*}_j) \bar{X}'_{rj} + O(m^{-1})$,

and, by property (h), T_1 becomes

$$\frac{\gamma_r}{\hat{\mu}_{\tilde{X}_r}} \sum_{j=1}^{m} \frac{L_j}{n} \psi(U_j'^*) \bar{X}'_{rj} + O(m^{-1}) = \frac{\gamma_r}{\hat{\mu}_{\tilde{X}_r}} \sum_{j=1}^{m} \frac{1}{n} \sum_{k=1}^{L_j} X'_{rjk} \{\psi(U'_{jk}) - \delta_{0jk}\} + O(m^{-1})$$

$$= \frac{\gamma_r}{\hat{\mu}_{\tilde{X}_r}} \frac{1}{n} \sum_{i=1}^{n} \psi(U_i) X_{ri} + O(m^{-1}) = \gamma_r + O_p(n^{-1/2}) + O(m^{-1}).$$

By properties (e) and (f), and the fact that $(L_j^{-1} \tilde{X}_j'^{\mathrm{T}} \tilde{X}'_j)^{-1} \tilde{X}_j'^{\mathrm{T}}$ is bounded uniformly over j on event A, $T_2 = O(m^{-1})$. Note that, on event A, $E(T_3|U, \tilde{X}, L_j, X) = 0$ and

$$\mathrm{var}(T_3|U, \tilde{X}, L_j, X) = \frac{\sigma^2}{n} \sum_{j=1}^{m} \frac{\hat{\mu}^2_{\tilde{X}'_{rj}}}{n \hat{\mu}^2_{\tilde{X}_r}} \sum_{k=1}^{L_j} \{(L_j^{-1} \tilde{X}_j'^{\mathrm{T}} \tilde{X}'_j)^{-1} \tilde{X}_j'^{\mathrm{T}}\}^2_{rk} \psi^2(U'_{jk}) = O(n^{-1}).$$

Thus, $E(T_3) = 0$ and $\mathrm{var}(T_3) = O(n^{-1})$, implying that $T_3 = O_p(n^{-1/2})$ on A.

It follows that, on A,

$$\begin{aligned}
\hat{\gamma}_r &= \gamma_r + O_p(n^{-1/2}) + O(m^{-1}), \quad 1 \le r \le p, \\
\hat{\gamma}_0 &= \sum_{j=1}^{m} \frac{L_j}{n} \hat{\beta}_{0j} = \sum_{j=1}^{m} \frac{L_j}{n} \left(\frac{1}{L_j} \sum_{k=1}^{L_j} \tilde{Y}'_{jk} - \hat{\beta}_{1j} \hat{\mu}_{\tilde{X}'_{1j}} - \ldots - \hat{\beta}_{pj} \hat{\mu}_{\tilde{X}'_{pj}} \right) \\
&= \frac{1}{n} \sum_{j=1}^{m} \sum_{k=1}^{L_j} \tilde{Y}'_{jk} - \hat{\gamma}_1 \hat{\mu}_{\tilde{X}_1} - \ldots - \hat{\gamma}_p \hat{\mu}_{\tilde{X}_p} = E\tilde{Y} - \sum_{r=1}^{p} \gamma_r E X_r + O_p(n^{-1/2}) + O(m^{-1}) \\
&= \gamma_0 + O_p(n^{-1/2}) + O(m^{-1}).
\end{aligned}$$

Analysis of ξ_1 defined in (4). Assuming Conditions 1-6, we estimate γ_1 by the slope obtained from the least squares regression of $\tilde{e}_{\tilde{Y}|U}$ on $\tilde{e}_{\tilde{X}|U}$, where $\tilde{e}_{\tilde{Y}|U}$ and $\tilde{e}_{\tilde{X}|U}$ are the residuals from the nonparametric regression models $\tilde{Y} = E(\tilde{Y}|U) + e_{\tilde{Y}|U}$ and $\tilde{X} = E(\tilde{X}|U) + e_{\tilde{X}|U}$, respectively. Thus, $e_{\tilde{Y}|U} = \tilde{Y} - E(\tilde{Y}|U) = \tilde{Y} - \psi(U)\{\gamma_0 + \gamma_1 E(X)\}$ and $e_{\tilde{X}|U} = \tilde{X} - E(\tilde{X}|U) = \tilde{X} - \phi(U)E(X)$. Therefore, using the population normal equations for regression, we have

$$\xi_1 = \frac{E(e_{\tilde{Y}|U} e_{\tilde{X}|U}) - E(e_{\tilde{Y}|U}) E(e_{\tilde{X}|U})}{\mathrm{var}(e_{\tilde{X}|U})} = \gamma_1 \Delta = \xi_1,$$

where Δ as defined in §1.3 is equal to $[\gamma_1 E\{\psi(U)\phi(U)\}]/E\phi^2(U)$.

Next, we show that ξ_1 can assume any real value under suitable conditions. Let $\{\rho_1, \rho_2, \rho_3, \ldots\}$ be an orthogonal basis of the inner-product space $\mathcal{C}[a, b]$, which is the space of continuous functions on $[a, b]$, using the inner product

$$< g_1, g_2 >= \int_a^b g_1(u) g_2(u) f(u) du,$$

where $f(\cdot)$ represents the density function of U and we choose $\rho_1 \equiv 1$. Then ψ and ϕ can be expanded as $\psi = \sum_i \mu_i \rho_i$ and $\phi = \sum_i \eta_i \rho_i$, for sets of real numbers μ_i and η_i. The identifiability conditions imply that $\mu_1 = \eta_1 = 1$. Assume without loss of generality that for a given set of η_i, $i \geq 2$, $\mu_i = \lambda \eta_i$ for an arbitrary real number λ, and that $\sum_{i \geq 2} \eta_i^2 = 1$, i.e. $< \phi, \phi >= 2$. Hence, $\Delta = (1 + \lambda)/2$, which along with ξ_1 may assume any real value, since λ was arbitrary.

Analysis of ξ_2. In the regression model $\tilde{Y} = \alpha_0 + \alpha_1 \tilde{X} + \alpha_2 U + e$, α_1 is equivalent to the slope when regressing $e_{\tilde{Y}|U}$ on $e_{\tilde{X}|U}$, where $e_{\tilde{Y}|U}$ and $e_{\tilde{X}|U}$ are the residuals from the regression models $\tilde{Y} = a_0 + a_1 U + e_{\tilde{Y}|U}$ and $\tilde{X} = b_0 + b_1 U + e_{\tilde{X}|U}$, respectively. Assuming that $\psi(U) = c_0 + c_1 U$ and $\phi(U) = d_0 + d_1 U$, for some real numbers c_0, c_1, d_0 and d_1, we can evaluate $e_{\tilde{Y}|U}$ and $e_{\tilde{X}|U}$, and thus α_1. Using the population normal equations for regression, we find that $a_1 = \{E(\tilde{Y}U) - E(\tilde{Y})E(U)\}/\mathrm{var}(U) = c_1\{\gamma_0 + \gamma_1 E(X)\}$, $a_0 = E(\tilde{Y}) - a_1 E(U) = c_0\{\gamma_0 + \gamma_1 E(X)\}$, $b_1 = \{E(\tilde{X}U) - E(\tilde{X})E(U)\}/\mathrm{var}(U) = d_1 E(X)$, and $b_0 = E(\tilde{X}) - b_1 E(U) = d_0 E(X)$. Therefore, $e_{\tilde{Y}|U} = \tilde{Y} - \{\gamma_0 + \gamma_1 E(X)\}(c_0 + c_1 U)$, $e_{\tilde{X}|U} = \tilde{X} - E(X)(d_0 + d_1 U)$, and

$$\alpha_1 = \frac{E(e_{\tilde{Y}|U} e_{\tilde{X}|U}) - E(e_{\tilde{Y}|U})E(e_{\tilde{X}|U})}{\mathrm{var}(e_{\tilde{X}|U})} = \gamma_1 \Delta = \xi_2.$$

Analysis of ξ_3 *in (5).* Consider the regression model $\tilde{Y} = \alpha_0 + \alpha_1 \tilde{X} + e$. Applying the population normal equation for the regression slope, and simplifying terms, we find that

$$\alpha_1 = \frac{E(\tilde{Y}\tilde{X}) - E(\tilde{Y})E(\tilde{X})}{\mathrm{var}(\tilde{X})} = \xi_3.$$

Expanding ψ and ϕ in the same way as in the above analysis of ξ_1, and also assuming that $E(\tilde{X}) = 1$, which implies that $E(X) = 1$ under the identifiability conditions, we see that $\xi_3 = [(1 + \lambda)\{\gamma_0 + \gamma_1 E(X^2)\} - \gamma_0 - \gamma_1]/\{2E(X^2) - 1\}$ can assume any real value under minimal conditions.

References

Cai, Z., Fan, J. & Li, R. (2000). Efficient estimation and inferences for varying coefficient models. *J. Am. Statist. Assoc.* **95,** 888-902.

Chiang, C., Rice, J. A. & Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Am. Statist. Assoc.* **96,** 605-17.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. New York: Cambridge University Press.

Eagleson, G.K. & Müller, H.G. (1997). Transformations for smooth regression models with multiplicative errors. *J. R. Statist. Soc.* B **59,** 173-89.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.

Fan, J. & Zhang J. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Statist. Soc.* B **62,** 303-22.

Härdle, W., Janssen, P. & Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16,** 1428-49.

Hart, J. (1997). *Nonparametric Smoothing and Lack of Fit Tests*. New York: Springer-Verlag.

Hastie, T. & Tibshirani, R. (1993). Varying coefficient models. *J. R. Statist. Soc.* B **55,** 757-96.

Heckman, N. E. (1986). Spline smoothing in a partly linear model. *J. R. Statist. Soc.* B **48,** 244-8.

Hoover, D. R., Rice, J. A., Wu, C. O. & Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85,** 809-22.

Hwang, J.T. (1986). Multiplicative errors-in-variables models with applications to recent data released by the U.S. Department of Energy. *J. Am. Statist. Assoc.* **81,** 680-8.

Iturria, S., Carroll, R. J. & Firth, D. (1999). Polynomial regression and estimating functions in the presence of multiplicative measurement error. *J. R. Statist. Soc.* B **61,** 547-61.

Kaysen, G. A., Dubin, J. A., Müller, H. G., Mitch, W. E., Rosales, L. M., Levin, N. W. & the Hemo Study Group (2003). Relationship among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients. *Kidney Int.* **61,** 2240-9.

Lai, T. L., Robbins, H. & Wei, C. Z. (1979). Strong consistency of least-squares estimates in multiple regression II. *J. Mult. Anal.* **9,** 343-61.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12,** 1215-

30.

Wu, C. O. & Chiang, C. T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statist. Sinica* **10,** 433-56.

Wu, C. O. & Yu, K. F. (2002). Nonparametric varying coefficient models for the analysis of longitudinal data. *Int. Statist. Rev.* **70,** 373-93.

Table 1: Parameter estimates for the regression model $FIB = \gamma_0 + \gamma_1 TRF + e$, calculated by least squares regression of $\tilde{Y}$ on $\tilde{X}$ and by covariate-adjusted regression, for $n = 69$ haemodialysis patients. The p-values were obtained from t-tests and the proposed bootstrap test, respectively.

	Least sq. Reg.		Covariate Adj. Reg.	
Coefficients	Estimate	p-value	Estimate	p-value
Intercept	675.987	0.000	701.163	0.000
TRF	-0.704	0.101	-0.844	0.002

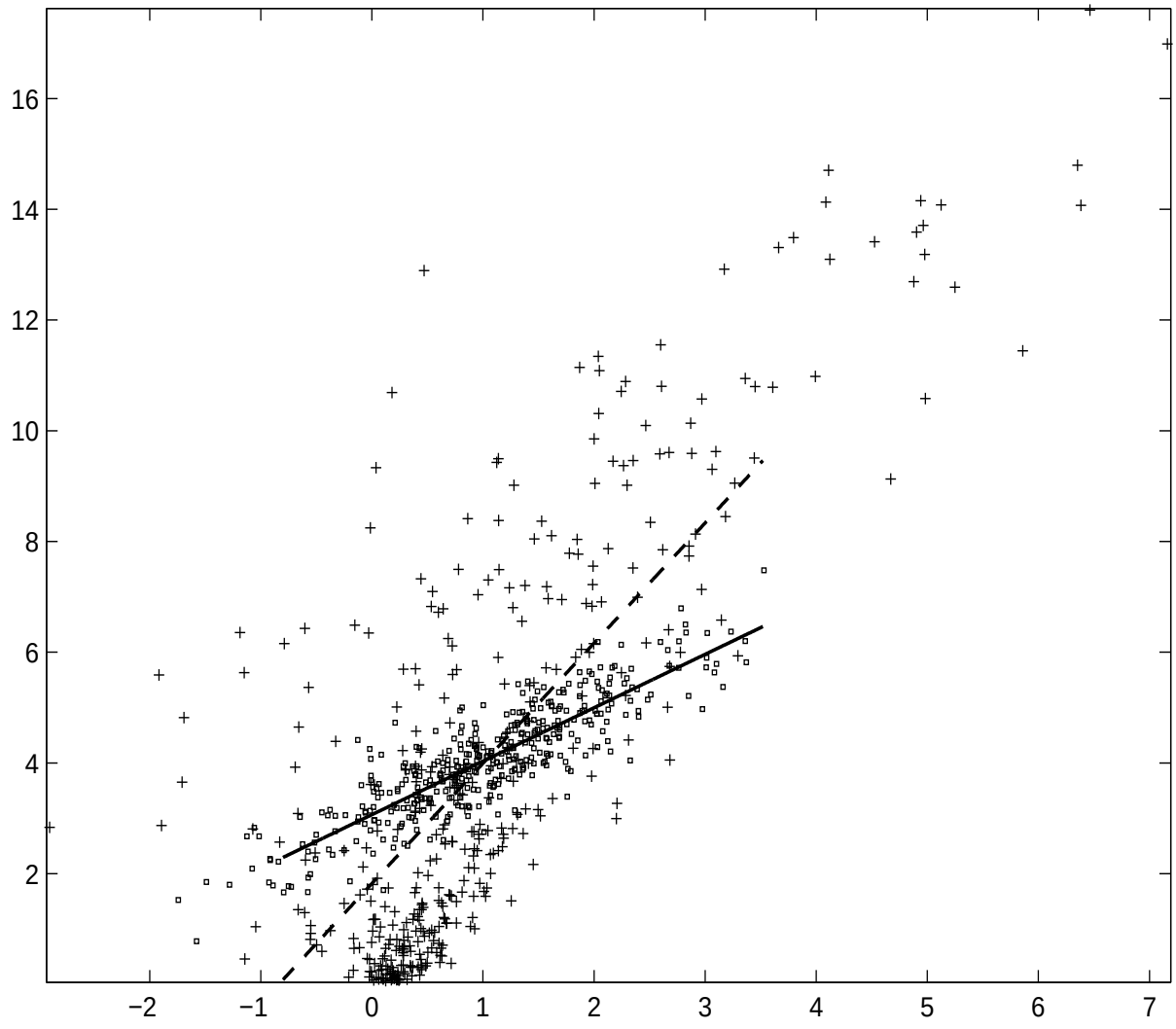

Figure 1: Data (X_i, Y_i) (squares), $i = 1, \ldots, 400$, generated from the underlying regression model in (6), along with the distorted data $(\tilde{X}_i, \tilde{Y}_i)$ (crosses). Least squares linear fits for distorted data, $\tilde{y} = 1.8231 + 2.1706\tilde{x}$ (dashed) ($\hat{r}^2 = 0.7763$), and for original data, $y = 3.0665 + 0.9652x$ (solid) ($\hat{r}^2 = 0.8748$) are also shown.

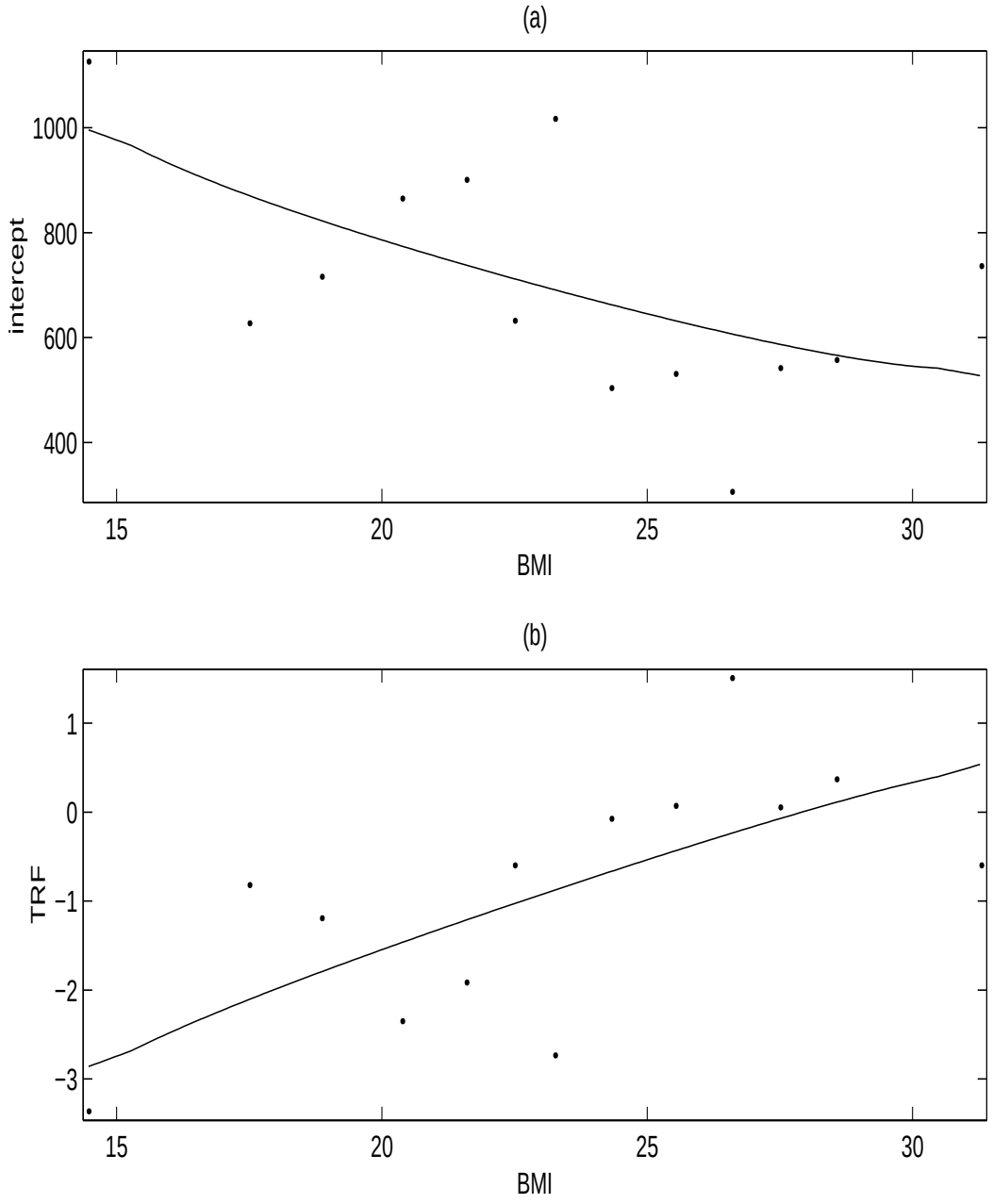

Figure 2: Plots of the estimated smooth coefficient functions (a) $\tilde{\beta}_0(\cdot)$ and (b) $\tilde{\beta}_1(\cdot)$ for the covariate-adjusted regression model $FIB = \beta_0(BMI) + \beta_1(BMI)TRF + \epsilon(BMI)$, estimated with local linear smoothing with smoothing parameter choices of $h = 16$ for each curve, obtained by applying (15). Sample size is 69, the number of bins formed is 13, and BMI =body mass index, FIB =plasma fibrinogen concentration and TRF = plasma transferrin level.

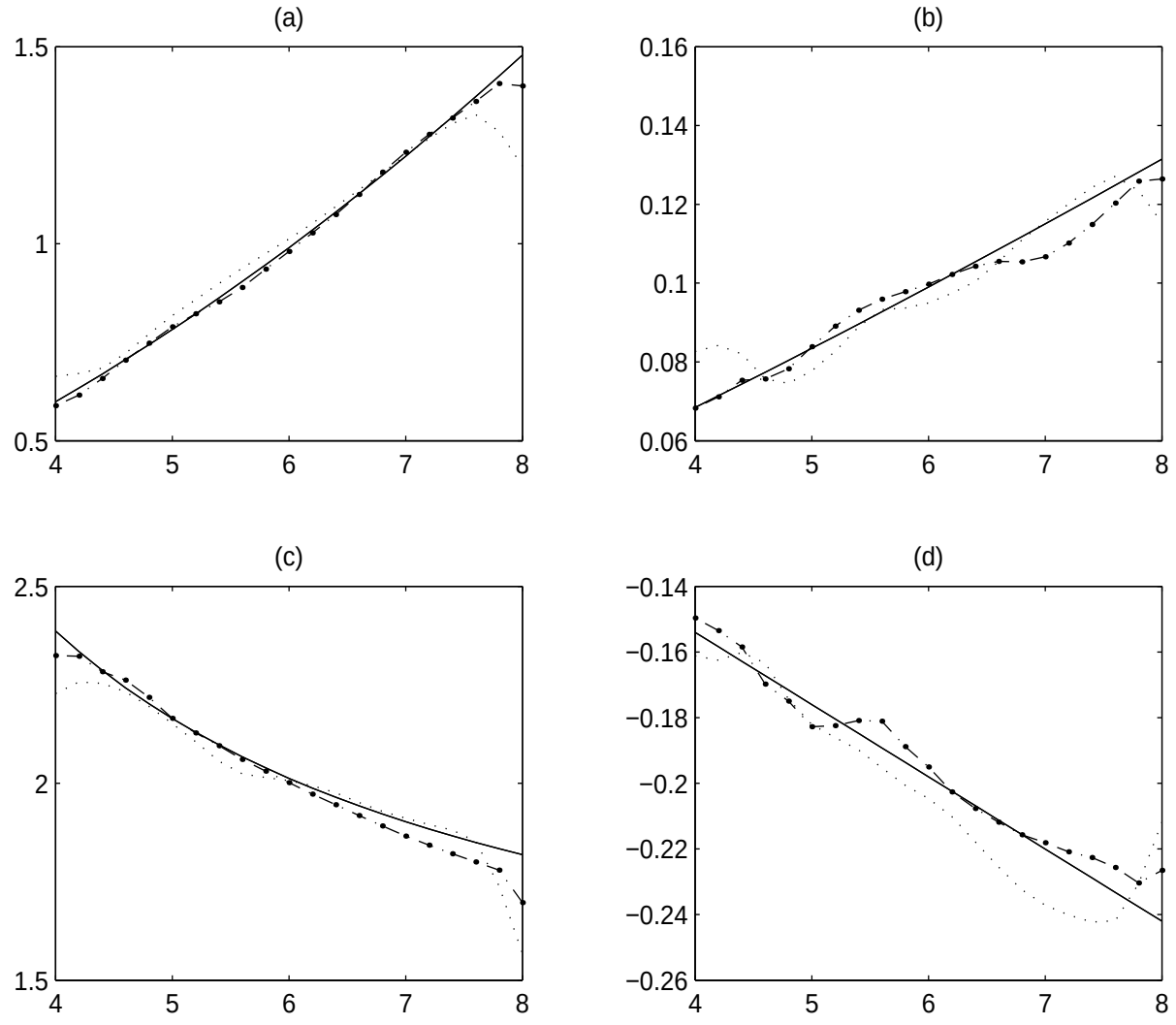

Figure 3: Cross-sectional means of the 1000 estimates of (a) $\tilde{\beta}_0(\cdot)$, (b) $\tilde{\beta}_1(\cdot)$, (c) $\tilde{\beta}_2(\cdot)$ and (d) $\tilde{\beta}_3(\cdot)$ of the smooth coefficient functions $\beta_0(\cdot)$, $\beta_1(\cdot)$, $\beta_2(\cdot)$, $\beta_3(\cdot)$ of the model in §6, fitted using local linear smoothing. On average three curves considered as outliers have been removed for each plot. The solid, dotted and dash-dotted lines correspond, respectively, to the target coefficient functions and the cross-sectional means for sample sizes of $n = 70$ and $n = 400$.

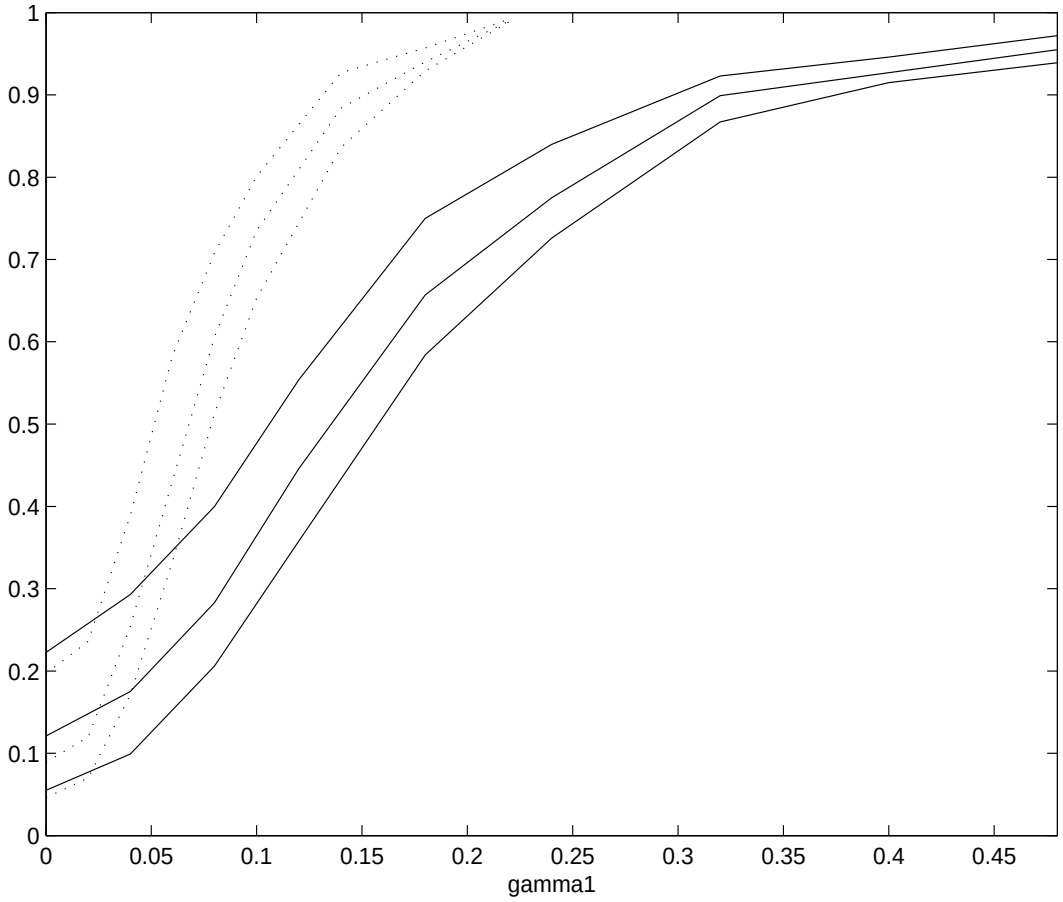

Figure 4: Power functions of the proposed bootstrap test for γ_1 in (6) for the simulation as described in §6, at the three significance levels 0.05, bottom curve, 0.10, middle curve, and 0.20, top curve. Solid lines correspond to $n = 70$, and dotted lines to $n = 400$.