INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

2011/23

ESTIMATION OF THE ERROR DENSITY IN A SEMIPARAMETRIC TRANSFORMATION MODEL

SAMB, R., HEUCHENNE, C. and I. VAN KEILEGOM

Estimation of the Error Density in a Semiparametric Transformation Model

Rawane Samb
Université catholique de Louvain *

Cédric Heuchenne
University of Liège and Université catholique de Louvain †

Ingrid Van Keilegom
Université catholique de Louvain ‡

September 27, 2011

Abstract

Consider the semiparametric transformation model $\Lambda_{\theta_o}(Y) = m(X) + \varepsilon$, where θ_o is an unknown finite dimensional parameter, the functions Λ_{θ_o} and m are smooth, ε is independent of X, and $\mathbb{E}(\varepsilon) = 0$. We propose a kernel-type estimator of the density of the error ε, and prove its asymptotic normality. The estimated errors, which lie at the basis of this estimator, are obtained from a profile likelihood estimator of θ_o and a nonparametric kernel estimator of m. The practical performance of the proposed density estimator is evaluated in a simulation study.

Key Words: Density estimation; Kernel smoothing; Nonparametric regression; Profile likelihood; Transformation model.

*R. Samb acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy).

†C. Heuchenne acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

‡I. Van Keilegom acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy), from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

1 Introduction

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent replicates of the random vector (X, Y), where Y is a univariate dependent variable and X is a one-dimensional covariate. We assume that Y and X are related via the semiparametric transformation model

$$\Lambda_{\theta_o}(Y) = m(X) + \varepsilon, \tag{1.1}$$

where ε is independent of X and has mean zero. We assume that $\{\Lambda_\theta : \theta \in \Theta\}$ (with $\Theta \subset \mathbb{R}^p$ compact) is a parametric family of strictly increasing functions defined on an unbounded subset $\mathcal{D}$ in $\mathbb{R}$, while m is the unknown regression function, belonging to an infinite dimensional parameter set $\mathcal{M}$. We assume that $\mathcal{M}$ is a space of functions endowed with the norm $\|\cdot\|_{\mathcal{M}} = \|\cdot\|_\infty$. We denote $\theta_o \in \Theta$ and $m \in \mathcal{M}$ for the true unknown finite and infinite dimensional parameters. Define the regression function

$$m_\theta(x) = \mathbb{E}[\Lambda_\theta(Y)|X = x],$$

for each $\theta \in \Theta$, and let $\varepsilon_\theta = \varepsilon(\theta) = \Lambda_\theta(Y) - m_\theta(X)$.

In this paper, we are interested in the estimation of the probability density function (p.d.f.) f_ε of the residual term $\varepsilon = \Lambda_{\theta_o}(Y) - m(X)$. To this end, we first obtain the estimators $\widehat{\theta}$ and $\widehat{m}_\theta$ of the parameter θ_o and the function m_θ, and second, form the semiparametric regression residuals $\widehat{\varepsilon}_i(\widehat{\theta}) = \Lambda_{\widehat{\theta}}(Y_i) - \widehat{m}_{\widehat{\theta}}(X_i)$. To estimate θ_o we use a profile likelihood (PL) approach, developed in Linton, Sperlich and Van Keilegom (2008), whereas $\widehat{m}_\theta$ is estimated by means of a Nadaraya-Watson-type estimator (Nadaraya, 1964, Watson, 1964). To our knowledge, the estimation of the density of ε in model (1.1) has not yet been investigated in the statistical literature. This estimation may be very useful in various regression problems. Indeed, taking transformations of the data may induce normality and error variance homogeneity in the transformed model. So the estimation of the error density in the transformed model may be used for testing these hypotheses.

Taking transformations of the data has been an important part of statistical practice for many years. A major contribution to this methodology was made by Box and Cox (1964), who proposed a parametric power family of transformations that includes the logarithm and the identity. They suggested that the power transformation, when applied to the dependent variable in a linear regression model, might induce normality and homoscedasticity. Lots of effort has been devoted to the investigation of the Box-Cox transformation since its introduction. See, for example, Amemiya (1985), Horowitz (1998), Chen, Lockhart and Stephens (2002), Shin (2008), and Fitzenberger, Wilke and Zhang (2010). Other dependent variable transformations have been suggested, for example, the Zellner and Revankar (1969) transform and the Bickel and Doksum

(1981) transform. The transformation methodology has been quite successful and a large literature exists on this topic for parametric models. See Carroll and Ruppert (1988) and Sakia (1992) and references therein.

The estimation of (functionals of) the error distribution and density under simplified versions of model (1.1) has received considerable attention in the statistical literature in recent years. Consider e.g. model (1.1) but with $\Lambda_{\theta_o} \equiv id$, i.e. the response is not transformed. Under this model, Escanciano and Jacho-Chavez (2010) considered the estimation of the (marginal) density of the response Y via the estimation of the error density. Akritas and Van Keilegom (2001) estimated the cumulative distribution function of the regression error in a heteroscedastic model with univariate covariates. The estimator they proposed is based on nonparametrically estimated regression residuals. The weak convergence of their estimator was proved. The results obtained by Akritas and Van Keilegom (2001) have been generalized by Neumeyer and Van Keilegom (2010) to the case of multivariate covariates. Müller, Schick and Wefelmeyer (2004) investigated linear functionals of the error distribution in nonparametric regression. Cheng (2005) established the asymptotic normality of an estimator of the error density based on estimated residuals. The estimator he proposed is constructed by splitting the sample into two parts: the first part is used for the estimation of the residuals, while the second part of the sample is used for the construction of the error density estimator. Efromovich (2005) proposed an adaptive estimator of the error density, based on a density estimator proposed by Pinsker (1980). Finally, Samb (2010) also considered the estimation of the error density, but his approach is more closely related to the one in Akritas and Van Keilegom (2001).

In order to achieve the objective of this paper, namely the estimation of the error density under model (1.1), we first need to estimate the transformation parameter θ_o. To this end, we make use of the results in Linton, Sperlich and Van Keilegom (2008). In the latter paper, the authors first discuss the nonparametric identification of model (1.1), and second, estimate the transformation parameter θ_o under the considered model. For the estimation of this parameter, they propose two approaches. The first approach uses a semiparametric profile likelihood (PL) estimator, while the second is based on a 'mean squared distance from independence-estimator (MD) using the estimated distributions of X, ε and (X, ε). Linton, Sperlich and Van Keilegom (2008) derived the asymptotic distributions of their estimators under certain regularity conditions, and proved that both estimators of θ_o are root-n consistent. The authors also showed that, in practice, the performance of the PL method is better than that of the MD approach. For this reason, the PL method will be considered in this paper for the estimation of θ_o.

The rest of the paper is organized as follows. Section 2 presents our estimator of the error density and groups some notations and technical assumptions. Section 3 describes the asymptotic results of the paper.

A simulation study is given in Section 4, while Section 5 is devoted to some general conclusions. Finally, the proofs of the asymptotic results are collected in Section 6.

2 Definitions and assumptions

2.1 Construction of the estimators

The approach proposed here for the estimation of f_ε is based on a two-steps procedure. In a first step, we estimate the finite dimensional parameter θ_o. This parameter is estimated by the profile likelihood (PL) method, developed in Linton, Sperlich and Van Keilegom (2008). The basic idea of this method is to replace all unknown expressions in the likelihood function by their nonparametric kernel estimates. Under model (1.1), we have

$$\mathbb{P}\left(Y \leq y|X\right) = \mathbb{P}\left(\Lambda_{\theta_o}(Y) \leq \Lambda_{\theta_o}(y)|X\right) = \mathbb{P}\left(\varepsilon_{\theta_o} \leq \Lambda_{\theta_o}(y) - m_{\theta_o}(X)|X\right) = F_\varepsilon\left(\Lambda_{\theta_o}(y) - m_{\theta_o}(X)\right).$$

Here, $F_\varepsilon(t) = \mathbb{P}(\varepsilon \leq t)$, and so

$$f_{Y|X}(y|x) = f_\varepsilon\left(\Lambda_{\theta_o}(y) - m_{\theta_o}(x)\right)\Lambda'_{\theta_o}(y),$$

where f_ε and $f_{Y|X}$ are the densities of ε, and of Y given X, respectively. Then, the log likelihood function is

$$\sum_{i=1}^{n}\left\{\log f_{\varepsilon_\theta}(\Lambda_\theta(Y_i) - m_\theta(X_i)) + \log \Lambda'_\theta(Y_i)\right\}, \quad \theta \in \Theta,$$

where f_{ε_θ} is the density function of ε_θ. Now, let

$$\widehat{m}_\theta(x) = \frac{\sum_{j=1}^n \Lambda_\theta(Y_j) K_1\left(\frac{X_j - x}{h}\right)}{\sum_{j=1}^n K_1\left(\frac{X_j - x}{h}\right)} \tag{2.1}$$

be the Nadaraya-Watson estimator of $m_\theta(x)$, and let

$$\widehat{f}_{\varepsilon_\theta}(t) = \frac{1}{ng}\sum_{i=1}^{n} K_2\left(\frac{\widehat{\varepsilon}_i(\theta) - t}{g}\right). \tag{2.2}$$

where $\widehat{\varepsilon}_i(\theta) = \Lambda_\theta(Y_i) - \widehat{m}_\theta(X_i)$. Here, K_1 and K_2 are kernel functions and h and g are bandwidth sequences. Then, the PL estimator of θ_o is defined by

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n}\left[\log \widehat{f}_{\varepsilon_\theta}(\Lambda_\theta(Y_i) - \widehat{m}_\theta(X_i)) + \log \Lambda'_\theta(Y_i)\right]. \tag{2.3}$$

Recall that $\widehat{m}_\theta(X_i)$ converges to $m_\theta(X_i)$ at a slower rate for those X_i which are close to the boundary of the support $\mathcal{X}$ of the covariate X. That is why we assume implicitly that the proposed estimator (2.3) of θ_o

trims the observations X_i outside a subset $\mathcal{X}_0$ of $\mathcal{X}$. Note that we keep the root-n consistency of $\widehat{\theta}$ proved in Linton, Sperlich and Van Keilegom (2008) by trimming the covariates outside $\mathcal{X}_0$. But in this case, the resulting asymptotic variance is different to the one obtained in the latter paper.

In a second step, we use the above estimator $\widehat{\theta}$ to build the estimated residuals $\widehat{\varepsilon}_i(\widehat{\theta}) = \Lambda_{\widehat{\theta}}(Y_i) - \widehat{m}_{\widehat{\theta}}(X_i)$. Then, our proposed estimator $\widehat{f}_{\widehat{\varepsilon}}(t)$ of $f_\varepsilon(t)$ is defined by

$$\widehat{f}_{\widehat{\varepsilon}}(t) = \frac{1}{nb} \sum_{i=1}^{n} K_3 \left(\frac{\widehat{\varepsilon}_i(\widehat{\theta}) - t}{b} \right), \tag{2.4}$$

where K_3 is a kernel function and b is a bandwidth sequence, not necessarily the same as the kernel K_2 and the bandwidth g used in (2.2). Observe that this estimator is a feasible estimator in the sense that it does not depend on any unknown quantity, as is desirable in practice. This contrasts with the unfeasible ideal kernel estimator

$$\widetilde{f}_{\varepsilon}(t) = \frac{1}{nb} \sum_{i=1}^{n} K_3 \left(\frac{\varepsilon_i - t}{b} \right), \tag{2.5}$$

which depends in particular on the unknown regression errors $\varepsilon_i = \varepsilon_i(\theta_o) = \Lambda_{\theta_o}(Y_i) - m(X_i)$. It is however intuitively clear that $\widehat{f}_{\widehat{\varepsilon}}(t)$ and $\widetilde{f}_{\varepsilon}(t)$ will be very close for n large enough, as will be illustrated by the results given in Section 3.

2.2 Notations

When there is no ambiguity, we use ε and m to indicate ε_{θ_o} and m_{θ_o}. Moreover, $\mathcal{N}(\theta_o)$ represents a neighborhood of θ_o. For the kernel K_j $(j = 1, 2, 3)$, let $\mu(K_j) = \int v^2 K_j(v)dv$ and let $K_j^{(p)}$ be the pth derivative of K_j. For any function $\varphi_\theta(y)$, denote $\dot{\varphi}_\theta(y) = \partial\varphi_\theta(y)/\partial\theta = (\partial\varphi_\theta(y)/\partial\theta_1, \ldots, \partial\varphi_\theta(y)/\partial\theta_p)^t$ and $\varphi'_\theta(y) = \partial\varphi_\theta(y)/\partial y$. Also, let $\|A\| = (A^t A)^{1/2}$ be the Euclidean norm of any vector A.

For any functions $\widetilde{m}$, r, f, φ and q, and any $\theta \in \Theta$, let $s = (\widetilde{m}, r, f, \varphi, q)$, $s_\theta = (m_\theta, \dot{m}_\theta, f_{\varepsilon_\theta}, f'_{\varepsilon_\theta}, \dot{f}_{\varepsilon_\theta})$, $\varepsilon_i(\theta, \widetilde{m}) = \Lambda_\theta(Y_i) - \widetilde{m}(X_i)$, and define

$$G_n(\theta, s) = n^{-1} \sum_{i=1}^{n} \left\{ \frac{1}{f\{\varepsilon_i(\theta, \widetilde{m})\}} \left[\varphi\{\varepsilon_i(\theta, \widetilde{m})\}\{\dot{\Lambda}_\theta(Y_i) - r(X_i)\} + q\{\varepsilon_i(\theta, \widetilde{m})\} \right] + \frac{\dot{\Lambda}'_\theta(Y_i)}{\Lambda'_\theta(Y_i)} \right\},$$

$G(\theta, s) = \mathbb{E}[G_n(\theta, s)]$ and $\mathcal{G}(\theta_o, s_{\theta_o}) = \frac{\partial}{\partial\theta} G(\theta, s_\theta)\downarrow_{\theta=\theta_o}$.

2.3 Technical assumptions

The assumptions we need for the asymptotic results are listed below for convenient reference.

(**A1**) The function K_j $(j = 1, 2, 3)$ is symmetric, has compact support, $\int v^k K_j(v)dv = 0$ for $k = 1, \ldots, q_j - 1$ and $\int v^{q_j} K_j(v)dv \neq 0$ for some $q_j \geq 4$, K_j is twice continuously differentiable, and $\int K_3^{(1)}(v)dv = 0$.

(**A2**) The bandwidth sequences h, g and b satisfy $nh^{2q_1} = o(1)$, $ng^{2q_2} = o(1)$ (where q_1 and q_2 are defined in (A1)), $(nb^5)^{-1} = O(1)$, $nb^3h^2(\log h^{-1})^{-2} \to \infty$ and $ng^6(\log g^{-1})^{-2} \to \infty$.

(**A3**) (i) The support $\mathcal{X}$ of the covariate X is a compact subset of $\mathbb{R}$, and $\mathcal{X}_0$ is a subset with non empty interior, whose closure is in the interior of $\mathcal{X}$.
(ii) The density f_X is bounded away from zero and infinity on $\mathcal{X}$, and has continuous second order partial derivatives on $\mathcal{X}$.

(**A4**) The function $m_\theta(x)$ is twice continuously differentiable with respect to θ on $\mathcal{X} \times \mathcal{N}(\theta_0)$, and the functions $m_\theta(x)$ and $\dot{m}_\theta(x)$ are q_1 times continuously differentiable with respect to x on $\mathcal{X} \times \mathcal{N}(\theta_0)$. All these derivatives are bounded, uniformly in $(x, \theta) \in \mathcal{X} \times \mathcal{N}(\theta_o)$.

(**A5**) The error $\varepsilon = \Lambda_{\theta_o}(Y) - m(X)$ has finite fourth moment and is independent of X.

(**A6**) The distribution $F_{\varepsilon_\theta}(t)$ is $q_3 + 1$ (respectively three) times continuously differentiable with respect to t (respectively θ), and

$$\sup_{\theta,t} \left\| \frac{\partial^{k+\ell}}{\partial t^k \partial\theta_1^{\ell_1} \ldots \partial\theta_p^{\ell_p}} F_{\varepsilon_\theta}(t) \right\| < \infty$$

for all k and ℓ such that $0 \leq k + \ell \leq 2$, where $\ell = \ell_1 + \ldots + \ell_p$ and $\theta = (\theta_1, \ldots, \theta_p)^t$.

(**A7**) The transformation $\Lambda_\theta(y)$ is three times continuously differentiable with respect to both θ and y, and there exists a $\alpha > 0$ such that

$$\mathbb{E}\left[\sup_{\theta' : \|\theta' - \theta\| \leq \alpha} \left\| \frac{\partial^{k+\ell}}{\partial y^k \partial\theta_1^{\ell_1} \ldots \partial\theta_p^{\ell_p}} \Lambda_{\theta'}(Y) \right\| \right] < \infty$$

for all $\theta \in \Theta$, and for all k and ℓ such that $0 \leq k + \ell \leq 3$, where $\ell = \ell_1 + \ldots + \ell_p$ and $\theta = (\theta_1, \ldots, \theta_p)^t$. Moreover, $\sup_{x \in \mathcal{X}} \mathbb{E}[\dot{\Lambda}^4_{\theta_o}(Y)|X = x] < \infty$.

(**A8**) For all $\eta > 0$, there exists $\epsilon(\eta) > 0$ such that

$$\inf_{\|\theta - \theta_o\| > \eta} \|G(\theta, s_\theta)\| \geq \epsilon(\eta) > 0.$$

Moreover, the matrix $\mathcal{G}(\theta_o, s_{\theta_o})$ is non-singular.

(**A9**) (i) $\mathbb{E}(\Lambda_{\theta_o}(Y)) = 1$, $\Lambda_{\theta_o}(0) = 0$ and the set $\{x \in \mathcal{X}_0 : m'(x) \neq 0\}$ has nonempty interior.

(ii) Assume that $\phi(x,t) = \dot\Lambda_{\theta_o}(\Lambda_{\theta_o}^{-1}(m(x)+t))f_\varepsilon(t)$ is continuously differentiable with respect to t for all x and that

$$\sup_{s:|t-s|\le\delta} \mathbb{E}\left|\frac{\partial\phi}{\partial s}(X,s)\right| < \infty. \tag{2.6}$$

for all $t\in\mathbb{R}$ and for some $\delta>0$.

Assumptions (A1), part of (A2), (A3)(ii), (A4) and (A6), (A7) and (A8) are used by Linton, Sperlich and Van Keilegom (2008) to show that the PL estimator $\widehat\theta$ of θ_o is root n-consistent. The differentiability of K_j up to second order imposed in assumption (A1) is used to expand the two-steps kernel estimator $\widehat f_{\widehat\varepsilon}(t)$ in (2.4) around the unfeasible one $\widetilde f_\varepsilon(t)$. Assumptions (A3)(ii) and (A4) impose that all the functions to be estimated have bounded derivatives. The last assumption in (A2) is useful for obtaining the uniform convergence of the Nadaraya-Watson estimator of m_{θ_o} in (2.1) (see for instance Einmahl and Mason, 2005). This assumption is also needed in the study of the difference between the feasible estimator $\widehat f_{\widehat\varepsilon}(t)$ and the unfeasible estimator $\widetilde f_\varepsilon(t)$. Finally, (A9)(i) is needed for identifying the model (see Vanhems and Van Keilegom (2011)).

3 Asymptotic results

In this section we are interested in the asymptotic behavior of the estimator $\widehat f_{\widehat\varepsilon}(t)$. To this end, we first investigate its asymptotic representation, which will be needed to show its asymptotic normality.

Theorem 3.1. *Assume* (A1)-(A9). *Then,*

$$\widehat f_{\widehat\varepsilon}(t) - f_\varepsilon(t) = \frac{1}{nb}\sum_{i=1}^n K_3\left(\frac{\varepsilon_i - t}{b}\right) - f_\varepsilon(t) + R_n(t),$$

where $R_n(t) = o_{\mathbb{P}}\left((nb)^{-1/2}\right)$ *for all* $t\in\mathbb{R}$.

This result is important, since it shows that, provided the bias term is negligible, the estimation of θ_o and $m(\cdot)$ has asymptotically no effect on the behavior of the estimator $\widehat f_{\widehat\varepsilon}(t)$. Therefore, this estimator is asymptotically equivalent to the unfeasible estimator $\widetilde f_\varepsilon(t)$, based on the unknown true errors $\varepsilon_1,\ldots,\varepsilon_n$.

Our next result gives the asymptotic normality of the estimator $\widehat f_{\widehat\varepsilon}(t)$.

Theorem 3.2. *Assume* (A1)-(A9). *In addition, assume that* $nb^{2q_3+1} = O(1)$. *Then,*

$$\sqrt{nb}\left(\widehat f_{\widehat\varepsilon}(t) - \overline{f}_\varepsilon(t)\right) \xrightarrow{d} N\left(0, f_\varepsilon(t)\int K_3^2(v)dv\right),$$

where

$$\overline{f}_\varepsilon(t) = f_\varepsilon(t) + \frac{b^{q_3}}{q_3!} f_\varepsilon^{(q_3)}(t) \int v^{q_3} K_3(v) dv.$$

The proofs of Theorems 3.1 and 3.2 are given in Section 6.

4 Simulations

In this section, we investigate the performance of our method for different models and different sample sizes. Consider

$$\Lambda_{\theta_o}(Y) = b_0 + b_1 X^2 + b_2 \sin(\pi X) + \sigma_e \varepsilon, \tag{4.1}$$

where Λ_θ is the Manly (1976) transformation

$$\Lambda_\theta(y) = \begin{cases} \frac{e^{\theta y}-1}{\theta}, & \theta \neq 0, \\ y, & \theta = 0, \end{cases}$$

$\theta_o \in [-0.5, 1.5]$, X is uniformly distributed on the interval $[-0.5, 0.5]$, and ε is independent of X and has a standard normal distribution but restricted to the interval $[-3, 3]$. We study three different model settings. For each of them, $b_0 = 3\sigma_e + b_2$. The other parameters are chosen as follows:

Model 1: $b_1 = 5$, $b_2 = 2$, $\sigma_e = 1.5$;
Model 2: $b_1 = 3.5$, $b_2 = 1.5$, $\sigma_e = 1$;
Model 3: $b_1 = 2.5$, $b_2 = 1$, $\sigma_e = 0.5$.

The parameters and the error distribution have been chosen in such a way that the transformation $\Lambda_{\theta_o}(Y)$ is positive, to avoid problems when generating the variable Y. Our simulations are done for $\theta_o = 0$, 0.5 or 1. The estimator of θ_o is chosen from a grid on the interval $[-0.5, 1.5]$ with step size 0.0625. We used the kernel $K(x) = \frac{15}{16}\left(1 - x^2\right)^2 \mathbb{1}\left(|x| \leq 1\right)$ for both the regression function and the density estimators. The results are based on 100 random samples of size $n = 50$ or $n = 100$, and we worked with the bandwidths $h = 0.3 \times n^{-1/5}$ and $b = g = r_n$, where $r_n = 1.06 \times \text{std}(\widehat{\varepsilon}) \times n^{-1/5}$, which is Silverman's (1986) rule of thumb bandwidth for univariate density estimation. Here $\text{std}(\widehat{\varepsilon})$ is the average of the standard deviations of $\widehat{\varepsilon}$ over the 100 samples.

Table 1 shows the values of the mean, standard deviation and mean squared error of $\widehat{\theta}$ for the considered models, sample sizes and values of θ_o. We observe that the results for the different models are quite similar, and as expected, the results are better for $n = 100$ than for $n = 50$.

Table 2 shows the mean squared error (MSE) of the estimator $\widehat{f}_{\widetilde{\varepsilon}}(t)$ of the standardized (pseudo-estimated) error $\widetilde{\varepsilon} = (\Lambda_{\widehat{\theta}}(Y) - \widehat{m}_{\widehat{\theta}}(X))/\sigma_e$, for sample sizes $n = 50$ and $n = 100$ and for $t = -1$, 0 and 1. Results for $\widehat{f}_{\widehat{\varepsilon}}(t)$ have also been obtained, but are not reported here. Indeed, Figure 1, displaying $\widehat{f}_{\widetilde{\varepsilon}}(t)$, shows that, even though residuals are standardized for each simulation (with known σ_e), better behavior is observed for models with smaller σ_e. Moreover, we observe that for $\theta_o = 0$ there is very little difference between the curve of $\widehat{f}_{\widetilde{\varepsilon}}$ and the one of the standard normal density. On the other hand for $\theta_o = 0.5$ and $\theta_o = 1$, we notice an important difference between the two curves under Model 1 and 2, but the difference is less important under Model 3.

n	θ_o	mean($\widehat{\theta}$)			std($\widehat{\theta}$)			MSE($\widehat{\theta}$)		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
50	0	0.0063	0.0065	0.0071	0.0116	0.0161	0.0239	0.0064	0.0124	0.0277
	0.5	0.3787	0.3754	0.3907	0.0417	0.0438	0.0486	0.0783	0.0867	0.1140
	1	0.8197	0.8449	0.8658	0.0792	0.0796	0.0798	0.3506	0.3492	0.3409
100	0	0.0055	0.0148	0.0170	0.0057	0.0078	0.0116	0.0032	0.0059	0.0132
	0.5	0.4596	0.4621	0.4728	0.0246	0.254	0.0270	0.0634	0.0676	0.0752
	1	0.9196	0.9545	0.9749	0.0401	0.0437	0.0438	0.2092	0.1999	0.1637

Table 1: Approximation of the mean, the standard deviation and the mean squared error of $\widehat{\theta}$ for the three regression models. All numbers are calculated based on 100 random samples.

5 Conclusions

In this paper we have studied the estimation of the density of the error in a semiparametric transformation model. The regression function in this model is unspecified (except for some smoothness assumptions), whereas the transformation (of the dependent variable in the model) is supposed to belong to a parametric family of monotone transformations. The proposed estimator is a kernel-type estimator, and we have shown its asymptotic normality. The finite sample performance of the estimator is illustrated by means of a simulation study.

It would be interesting to explore various possible applications of the results in this paper. For example, one could use the results on the estimation of the error density to test hypotheses concerning e.g. the normality of the errors, the homoscedasticity of the error variance, or the linearity of the regression function, all of which are important features in the context of transformation models.

6 Proofs

Proof of Theorem 3.1. Write

$$\widehat{f}_{\widehat{\varepsilon}}(t) - f_\varepsilon(t) = [\widehat{f}_\varepsilon(t) - f_\varepsilon(t)] + [\widehat{f}_{\widehat{\varepsilon}}(t) - \widehat{f}_\varepsilon(t)],$$

where

$$\widehat{f}_\varepsilon(t) = \frac{1}{nb}\sum_{i=1}^n K_3\left(\frac{\widehat{\varepsilon}_i - t}{b}\right)$$

and $\widehat{\varepsilon}_i = \Lambda_{\theta_o}(Y_i) - \widehat{m}_{\theta_o}(X_i)$, $i = 1, \dots, n$. In a completely similar way as was done for Lemma A.1 in Linton, Sperlich and Van Keilegom (2008), it can be shown that

$$\widehat{f}_\varepsilon(t) - f_\varepsilon(t) = \frac{1}{nb}\sum_{i=1}^n K_3\left(\frac{\varepsilon_i - t}{b}\right) - f_\varepsilon(t) + o_{\mathbb{P}}((nb)^{-1/2}) \tag{6.1}$$

for all $t \in \mathbb{R}$. Note that the remainder term in Lemma A.1 in the above paper equals a sum of i.i.d. terms of mean zero, plus a $o_{\mathbb{P}}(n^{-1/2})$ term. Hence, the remainder term in that paper is $O_{\mathbb{P}}(n^{-1/2})$, whereas we write $o_{\mathbb{P}}((nb)^{-1/2})$ in (6.1). Therefore, the result of the theorem follows if we prove that $\widehat{f}_{\widehat{\varepsilon}}(t) - \widehat{f}_\varepsilon(t) = o_{\mathbb{P}}((nb)^{-1/2})$. To this end, write

$$\begin{aligned}
&\widehat{f}_{\widehat{\varepsilon}}(t) - \widehat{f}_\varepsilon(t) \\
&= \frac{1}{nb^2}\sum_{i=1}^n (\widehat{\varepsilon}_i(\widehat{\theta}) - \widehat{\varepsilon}_i(\theta_o)) K_3^{(1)}\left(\frac{\widehat{\varepsilon}_i(\theta_o) - t}{b}\right) \\
&\quad + \frac{1}{2nb^3}\sum_{i=1}^n (\widehat{\varepsilon}_i(\widehat{\theta}) - \widehat{\varepsilon}_i(\theta_o))^2 K_3^{(2)}\left(\frac{\widehat{\varepsilon}_i(\theta_o) + \beta(\widehat{\varepsilon}_i(\widehat{\theta}) - \widehat{\varepsilon}_i(\theta_o)) - t}{b}\right),
\end{aligned}$$

for some $\beta \in (0, 1)$. In what follows, we will show that each of the terms above is $o_{\mathbb{P}}((nb)^{-1/2})$. First consider the last term of (6.2). Since $\Lambda_\theta(y)$ and $\widehat{m}_\theta(x)$ are both twice continuously differentiable with respect to θ, the second order Taylor expansion gives, for some θ_1 between θ_o and $\widehat{\theta}$ (to simplify the notations, we assume here that $p = \dim(\theta) = 1$),

$$\begin{aligned}
&\widehat{\varepsilon}_i(\widehat{\theta}) - \widehat{\varepsilon}_i(\theta_o) \\
&= \Lambda_{\widehat{\theta}}(Y_i) - \Lambda_{\theta_o}(Y_i) - \left(\widehat{m}_{\widehat{\theta}}(X_i) - \widehat{m}_{\theta_o}(X_i)\right) \\
&= (\widehat{\theta} - \theta_o)(\dot{\Lambda}_{\theta_o}(Y_i) - \dot{\widehat{m}}_{\theta_o}(X_i)) + \frac{1}{2}(\widehat{\theta} - \theta_o)^2(\ddot{\Lambda}_{\theta_1}(Y_i) - \ddot{\widehat{m}}_{\theta_1}(X_i)).
\end{aligned}$$

Therefore, since $\widehat{\theta} - \theta_o = o_{\mathbb{P}}((nb)^{-1/2})$ by Theorem 4.1 in Linton, Sperlich and Van Keilegom (2008) (as before, we work with a slower rate than what is shown in the latter paper, since this leads to weaker conditions on the bandwidths), assumptions (A2) and (A7) imply that

$$\frac{1}{nb^3}\sum_{i=1}^{n}(\widehat{\varepsilon}_i(\widehat{\theta}) - \widehat{\varepsilon}_i(\theta_o))^2 K_3^{(2)}\left(\frac{\widehat{\varepsilon}_i(\theta_o) + \beta(\widehat{\varepsilon}_i(\widehat{\theta}) - \widehat{\varepsilon}_i(\theta_o)) - t}{b}\right) = O_{\mathbb{P}}\left((nb^3)^{-1}\right),$$

which is $o_{\mathbb{P}}((nb)^{-1/2})$, since $(nb^5)^{-1} = O(1)$ under (A2). For the first term of (6.2), the decomposition of $\widehat{\varepsilon}_i(\widehat{\theta}) - \widehat{\varepsilon}_i(\theta_o)$ given above yields

$$\begin{aligned}
&\frac{1}{nb^2}\sum_{i=1}^{n}(\widehat{\varepsilon}_i(\widehat{\theta}) - \widehat{\varepsilon}_i(\theta_o))K_3^{(1)}\left(\frac{\widehat{\varepsilon}_i(\theta_o) - t}{b}\right) \\
&= \frac{(\widehat{\theta} - \theta_o)}{nb^2}\sum_{i=1}^{n}(\dot{\Lambda}_{\theta_o}(Y_i) - \dot{\widehat{m}}_{\theta_o}(X_i))K_3^{(1)}\left(\frac{\widehat{\varepsilon}_i(\theta_o) - t}{b}\right) + o_{\mathbb{P}}((nb)^{-1/2}) \\
&= \frac{(\widehat{\theta} - \theta_o)}{nb^2}\sum_{i=1}^{n}(\dot{\Lambda}_{\theta_o}(Y_i) - \dot{m}_{\theta_o}(X_i))K_3^{(1)}\left(\frac{\varepsilon_i - t}{b}\right) + o_{\mathbb{P}}((nb)^{-1/2}), \qquad (6.2)
\end{aligned}$$

where the last equality follows from a Taylor expansion applied to $K_3^{(1)}$, the fact that

$$\dot{\widehat{m}}_{\theta_o}(x) - \dot{m}_{\theta_o}(x) = O_{\mathbb{P}}((nh)^{-1/2}(\log h^{-1})^{1/2}),$$

uniformly in $x \in \mathcal{X}_0$ by Lemma 6.1, and the fact that $nhb^3(\log h^{-1})^{-1} \to \infty$ under (A2). Further, write

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n}(\dot{\Lambda}_{\theta_o}(Y_i) - \dot{m}_{\theta_o}(X_i))K_3^{(1)}\left(\frac{\varepsilon_i - t}{b}\right)\right] \\
&= \sum_{i=1}^{n}\mathbb{E}\left[\dot{\Lambda}_{\theta_o}(Y_i)K_3^{(1)}\left(\frac{\varepsilon_i - t}{b}\right)\right] - \sum_{i=1}^{n}\mathbb{E}\left[\dot{m}_{\theta_o}(X_i)\right]\mathbb{E}\left[K_3^{(1)}\left(\frac{\varepsilon_i - t}{b}\right)\right] \\
&= A_n - B_n.
\end{aligned}$$

We will only show that the first term above is $O(nb^2)$ for any $t \in \mathbb{R}$. The proof for the other term is similar. Let $\varphi(x,t) = \dot{\Lambda}_{\theta_o}(\Lambda_{\theta_o}^{-1}(m(x)+t))$ and set $\phi(x,t) = \varphi(x,t)f_\varepsilon(t)$. Then, applying a Taylor expansion to $\phi(x,\cdot)$, it follows that (for some $\beta \in (0,1)$)

$$\begin{aligned}
A_n &= \sum_{i=1}^{n}\mathbb{E}\left[\dot{\Lambda}_{\theta_o}\left(\Lambda_{\theta_o}^{-1}(m(X_i) + \varepsilon_i)\right)K_3^{(1)}\left(\frac{\varepsilon_i - t}{b}\right)\right] \\
&= n\int\int \phi(x,e)K_3^{(1)}\left(\frac{e - t}{b}\right)f_X(x)dxde \\
&= nb\int\int \phi(x,t+bv)K_3^{(1)}(v)f_X(x)dxdv \\
&= nb\int\int \left[\phi(x,t) + bv\frac{\partial\phi}{\partial t}(x,t+\beta bv)\right]K_3^{(1)}(v)f_X(x)dxdv \\
&= nb^2\int\int v\frac{\partial\phi}{\partial t}(x,t+\beta bv)K_3^{(1)}(v)f_X(x)dxdv,
\end{aligned}$$

since $\int K_3^{(1)}(v)dv = 0$, and this is bounded by $Knb^2 \sup_{s:|t-s|\le\delta} \mathbb{E}|\frac{\partial\phi}{\partial s}(X,s)| = O(nb^2)$ by assumption (A9)(ii). Hence, Tchebychev's inequality ensures that

$$\begin{aligned}
&\frac{(\widehat{\theta}-\theta_o)}{b^2}\sum_{i=1}^n(\dot{\Lambda}_{\theta_o}(Y_i)-\dot{m}_{\theta_o}(X_i))K_3^{(1)}\left(\frac{\varepsilon_i-t}{b}\right)\\
&= \frac{(\widehat{\theta}-\theta_o)}{nb^2}O_{\mathbb{P}}(nb^2+(nb)^{1/2}) = o_{\mathbb{P}}((nb)^{-1/2}),
\end{aligned}$$

since $nb^{3/2}\to\infty$ by (A2). Substituting this in (6.2), yields

$$\frac{1}{nb^2}\sum_{i=1}^n(\widehat{\varepsilon}_i(\widehat{\theta})-\widehat{\varepsilon}_i(\theta_o))K_3^{(1)}\left(\frac{\widehat{\varepsilon}_i(\theta_o)-t}{b}\right) = o_{\mathbb{P}}((nb)^{-1/2}),$$

for any $t\in\mathbb{R}$. This completes the proof. □

Proof of Theorem 3.2. It follows from Theorem 3.1 that

$$\widehat{f_{\widehat{\varepsilon}}}(t)-f_\varepsilon(t) = [\widetilde{f}_\varepsilon(t)-\mathbb{E}\widetilde{f}_\varepsilon(t)]+[\mathbb{E}\widetilde{f}_\varepsilon(t)-f_\varepsilon(t)]+o_{\mathbb{P}}((nb)^{-1/2}). \tag{6.3}$$

The first term on the right hand side of (6.3) is treated by Lyapounov's Central Limit Theorem (LCT) for triangular arrays (see e.g. Billingsley 1968, Theorem 7.3). To this end, let

$$\widetilde{f}_{in}(t) = \frac{1}{b}K_3\left(\frac{\varepsilon_i-t}{b}\right).$$

Then, under (A1), (A2) and (A5) it can be easily shown that

$$\frac{\sum_{i=1}^n \mathbb{E}\left|\widetilde{f}_{in}(t)-\mathbb{E}\widetilde{f}_{in}(t)\right|^3}{\left(\sum_{i=1}^n \mathrm{Var}\widetilde{f}_{in}(t)\right)^{3/2}} \le \frac{Cnb^{-2}f_\varepsilon(t)\int|K_3(v)|^3\,dv+o\left(nb^{-2}\right)}{\left(nb^{-1}f_\varepsilon(t)\int K_3^2(v)dv+o\left(nb^{-1}\right)\right)^{3/2}} = O((nb)^{-1/2}) = o(1),$$

for some $C>0$. Hence, the LCT ensures that

$$\frac{\widetilde{f}_\varepsilon(t)-\mathbb{E}\widetilde{f}_\varepsilon(t)}{\sqrt{\mathrm{Var}\widetilde{f}_\varepsilon(t)}} = \frac{\widetilde{f}_\varepsilon(t)-\mathbb{E}\widetilde{f}_\varepsilon(t)}{\sqrt{\frac{\mathrm{Var}\widetilde{f}_{1n}(t)}{n}}} \xrightarrow{d} N(0,1).$$

This gives

$$\sqrt{nb}\left(\widetilde{f}_\varepsilon(t)-\mathbb{E}\widetilde{f}_\varepsilon(t)\right) \xrightarrow{d} N\left(0, f_\varepsilon(t)\int K_3^2(v)dv\right). \tag{6.4}$$

For the second term of (6.3), straightforward calculations show that

$$\mathbb{E}\widetilde{f}_\varepsilon(t)-f_\varepsilon(t) = \frac{b^{q_3}}{q_3!}f_\varepsilon^{(q_3)}(t)\int v^{q_3}K_3(v)dv+o(b^{q_3}).$$

Combining this with (6.4) and (6.3), we obtain the desired result. □

Lemma 6.1. *Assume* (A1)-(A5) *and* (A7). *Then,*

$$\begin{aligned}\sup_{x\in\mathcal{X}_0}|\widehat{m}_{\theta_o}(x)-m_{\theta_o}(x)| &= O_{\mathbb{P}}((nh)^{-1/2}(\log h^{-1})^{1/2}),\\ \sup_{x\in\mathcal{X}_0}|\dot{\widehat{m}}_{\theta_o}(x)-\dot{m}_{\theta_o}(x)| &= O_{\mathbb{P}}((nh)^{-1/2}(\log h^{-1})^{1/2}).\end{aligned}$$

Proof. We will only show the proof for $\dot{\widehat{m}}_{\theta_o}(x)-\dot{m}_{\theta_o}(x)$, the proof for $\widehat{m}_{\theta_o}(x)-m_{\theta_o}(x)$ being very similar. Let $c_n=(nh)^{-1/2}(\log h^{-1})^{1/2}$, and define

$$\dot{\widehat{r}}_{\theta_o}(x)=\frac{1}{nh}\sum_{j=1}^n\dot{\Lambda}_{\theta_o}(Y_j)K_1\left(\frac{X_j-x}{h}\right),\quad \dot{\overline{r}}_{\theta_o}(x)=\mathbb{E}[\dot{\widehat{r}}_{\theta_o}(x)],\quad \overline{f}_X(x)=\mathbb{E}[\widehat{f}_X(x)],$$

where $\widehat{f}_X(x)=(nh)^{-1}\sum_{j=1}^n K_1(\frac{X_j-x}{h})$. Then,

$$\sup_{x\in\mathcal{X}_0}|\dot{\widehat{m}}_{\theta_o}(x)-\dot{m}_{\theta_o}(x)|\le\sup_{x\in\mathcal{X}_0}\left|\dot{\widehat{m}}_{\theta_o}(x)-\frac{\dot{\overline{r}}_{\theta_o}(x)}{\overline{f}_X(x)}\right|+\sup_{x\in\mathcal{X}_0}\frac{1}{\overline{f}_X(x)}\left|\dot{\overline{r}}_{\theta_o}(x)-\overline{f}_X(x)\dot{m}_{\theta_o}(x)\right|. \tag{6.5}$$

Since $\mathbb{E}[\dot{\Lambda}^4_{\theta_o}(Y)|X=x]<\infty$ uniformly in $x\in\mathcal{X}$ by assumption (A7), a similar proof as was given for Theorem 2 in Einmahl and Mason (2005) ensures that

$$\sup_{x\in\mathcal{X}_0}\left|\dot{\widehat{m}}_{\theta_o}(x)-\frac{\dot{\overline{r}}_{\theta_o}(x)}{\overline{f}_X(x)}\right|=O_{\mathbb{P}}(c_n).$$

Consider now the second term of (6.5). Since $\mathbb{E}[\dot{\varepsilon}(\theta_o)|X]=0$, where $\dot{\varepsilon}(\theta_o)=\frac{d}{d\theta}(\Lambda_\theta(Y)-m_\theta(X))|_{\theta=\theta_o}$, we have

$$\begin{aligned}\dot{\overline{r}}_{\theta_o}(x) &= h^{-1}\mathbb{E}\left[\{\dot{m}_{\theta_o}(X)+\dot{\varepsilon}(\theta_o)\}K_1\left(\frac{X-x}{h}\right)\right]\\ &= h^{-1}\mathbb{E}\left[\dot{m}_{\theta_o}(X)K_1\left(\frac{X-x}{h}\right)\right]\\ &= \int\dot{m}_{\theta_o}(x+hv)K_1(v)f_X(x+hv)dv,\end{aligned}$$

from which it follows that

$$\dot{\overline{r}}_{\theta_o}(x)-\overline{f}_X(x)\dot{m}_{\theta_o}(x)=\int[\dot{m}_{\theta_o}(x+hv)-\dot{m}_{\theta_o}(x)]K_1(v)f_X(x+hv)dv.$$

Hence, a Taylor expansion applied to $\dot{m}_{\theta_o}(\cdot)$ yields

$$\sup_{x\in\mathcal{X}_0}\left|\dot{\overline{r}}_{\theta_o}(x)-\overline{f}_X(x)\dot{m}_{\theta_o}(x)\right|=O(h^{q_1})=O(c_n),$$

since $nh^{2q_1+1}(\log h^{-1})^{-1}=O(1)$ by (A2). This proves that the second term of (6.5) is $O(c_n)$, since it can be easily shown that $\overline{f}_X(x)$ is bounded away from 0 and infinity, uniformly in $x\in\mathcal{X}_0$, using (A3)(ii). □

References

[1] Akritas, M.G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Statist.*, **28**, 549–567.

[2] Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press, Cambridge.

[3] Bickel, P.J. and Doksum, K. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.,* **76**, 296–311.

[4] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley.

[5] Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. - Ser. B*, **26**, 211–252.

[6] Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, New York.

[7] Chen, G., Lockhart, R.A. and Stephens, A. (2002). Box-Cox transformations in linear models: Large sample theory and tests of normality (with discussion). *Canad. J. Statist,* **30**, 177–234.

[8] Cheng, F. (2005). Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *J. Statist. Plann. Infer.,* **128**, 327–349.

[9] Efromovich, S. (2005). Estimation of the density of the regression errors. *Ann. Statist.,* **33**, 2194–2227.

[10] Einmahl, U. and Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.,* **33**, 1380–1403.

[11] Escanciano, J.C. and Jacho-Chavez, D. (2010). $\sqrt{n}$-uniformly consistent density estimation in nonparametric regression models (submitted).

[12] Fitzenberger, B., Wilke, R.A. and Zhang, X. (2010). Implementing Box-Cox quantile regression. *Econometric Rev.,* **29**, 158–181.

[13] Horowitz, J.L. (1998). *Semiparametric Methods in Economics.* Springer-Verlag, New York.

[14] Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation of a semiparametric transformation model. *Ann. Statist.,* **36**, 686–718.

[15] Manly, B. F. (1976). Exponential data transformation. *The Statistician,* **25**, 37–42.

[16] Müller, U.U., Schick, A. and Wefelmeyer, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *J. Statist. Plann. Infer.,* **119**, 75–93.

[17] Nadaraya, E.A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen,* **9**, 157–159.

[18] Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multiv. Anal.,* **101**, 1067–1078.

[19] Pinsker, M.S. (1980). Optimal filtering of a square integrable signal in Gaussian white noise. *Problems Inform. Transmission*, **16**, 52–68.

[20] Sakia, R.M. (1992). The Box-Cox transformation technique: a review. *The Statistician,* **41**, 169–178.

[21] Samb, R. (2010). *Contribution à l'estimation nonparamétrique de la densité des erreurs de régression.* Doctoral Thesis. Avalaible at $http://arxiv.org/PS\_cache/arxiv/pdf/1011/1011.0674v1.pdf$.

[22] Shin, Y. (2008). Semiparametric estimation of the Box-Cox transformation model. *Econometrics J.,* **11**, 517–537.

[23] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Monographs on Statistics and Applied Probability, Chapman and Hall, London.

[24] Vanhems, A. and Van Keilegom, I. (2011). Semiparametric transformation model with endogeneity: a control function approach (submitted).

[25] Watson, G.S. (1964). Smooth regression analysis. *Sankhyā - Ser. A*, **26**, 359–372.

[26] Zellner, A. and Revankar, N.S. (1969). Generalized production functions. *Rev. Economic Studies,* **36**, 241–250.

n	θ_o	t	Mean Squared Error of $\widehat{f}_{\widehat{\varepsilon}}(t)$		
			Model 1	Model 2	Model 3
50	0	-1	0.0026	0.0025	0.0019
		0	0.0040	0.0037	0.0028
		1	0.0026	0.0023	0.0017
	0.5	-1	0.0063	0.0048	0.0025
		0	0.0527	0.0372	0.0147
		1	0.0062	0.0046	0.0020
	1	-1	0.0078	0.0048	0.0024
		0	0.0564	0.0314	0.0133
		1	0.0049	0.0030	0.0019
100	0	-1	0.0012	0.0011	0.0008
		0	0.0039	0.0035	0.0026
		1	0.0017	0.0015	0.0012
	0.5	-1	0.0015	0.0014	0.0011
		0	0.0075	0.0057	0.0031
		1	0.0021	0.0018	0.0012
	1	-1	0.0024	0.0018	0.0012
		0	0.0110	0.0052	0.0270
		1	0.0019	0.0016	0.0012

Table 2: Mean squared error of $\widehat{f}_{\widehat{\varepsilon}}(t)$ for three regression models. All numbers are calculated based on 100 random samples.

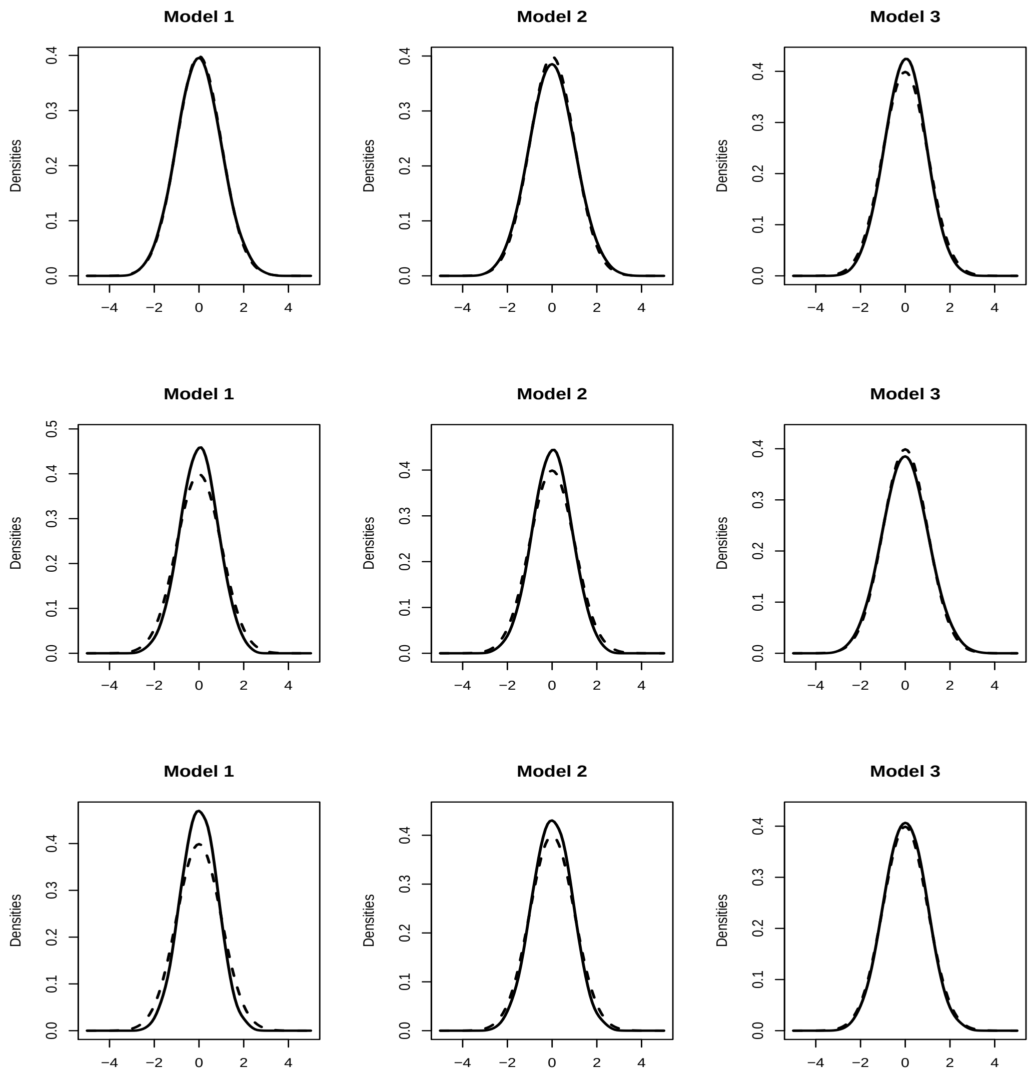

Figure 1: Curves of the pointwise average of $\widehat{f_{\widehat{\varepsilon}}}$ over 100 random samples of size $n = 100$ (solid curve) and of the standard normal density (dashed curve) for $\theta_o = 0$ (first row), $\theta_o = 0.5$ (second row) and $\theta_o = 1$ (third row).