Heteroscedastic semiparametric transformation models: estimation and testing for validity

Natalie Neumeyer[*] Hohsuk Noh[†] Ingrid Van Keilegom[‡]

November 28, 2015

Supplementary Material

Abstract

This supplementary file provides the proof of Theorem 2.1 and the proof of some auxiliary results used to prove Theorem 3.1.

[*]Department of Mathematics, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany, E-mail: neumeyer@math.uni-hamburg.de

[†]Department of Statistics, Sookmyung Women's University, 100 Cheongpa-ro 47-gil, Yongsan-gu, Seoul, South Korea 140-742, E-mail: word5810@gmail.com

[‡]Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, E-mail: ingrid.vankeilegom@uclouvain.be

S1 Proof of Theorem 2.1

We follow the different steps of the proof of Theorem 4.1 in Linton *et al.* (2008), which shows the asymptotic normality of $\hat{\vartheta}$ in the homoscedastic case. However, for brevity we focus on the differences with respect to that proof. The proof in Linton *et al.* (2008) consists of 11 lemmas from which the result follows. The lemmas that need closer attention are Lemmas A.1, A.2, A.3 and A.11. The other lemmas can be extended to the heteroscedastic case in a straightforward way. We start with the extension of Lemma A.1 to the heteroscedastic case. This lemma develops an i.i.d. expansion for $\hat{f}_{\hat{\varepsilon}(\vartheta_0)}(y) - f_{\varepsilon(\vartheta_0)}(y)$. For this, first note that

$$\hat{m}_{\vartheta_0}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} W_{x,n}\Big(\frac{x - X_i}{h}\Big)\Lambda_{\vartheta_0}(Y_i), \tag{S1.1}$$

where $W_{x,n}(u) = K^*(u)/f_X(x)(1 + o_P(1))$ uniformly in $u \in [-1, 1]^d$ and $x \in R_X$, and $(nh^d)^{-1} \sum_{i=1}^n W_{x,n}((x - X_i)/h) = 1$. The kernel $K^*(\cdot)$ is the so-called equivalent kernel and is a linear combination of functions of the form $\prod_{i=1}^d k(u_i)u_i^{j_i}$ with $(j_1, \ldots, j_d) \in \mathbb{N}_0^d$, $0 \le \sum_{i=1}^d j_i \le p$. This can be deduced from representation (3.25) in combination with (3.30), (3.9) and (3.19) in Gu, Li and Yang (2014); see also Masry (1996a, 1996b) and Fan and Gijbels (1996), p. 63–64, for the case $d = 1$. In a similar way we can also write

$$\begin{aligned}\hat{\sigma}_{\vartheta_0}(x) - \sigma_{\vartheta_0}(x) \;&=\; \frac{1}{2\sigma_{\vartheta_0}(x)} \frac{1}{nh^d} \sum_{i=1}^{n} W_{x,n}\Big(\frac{x - X_i}{h}\Big)\Big[(\Lambda_{\vartheta_0}(Y_i) - m_{\vartheta_0}(x))^2 - \sigma^2_{\vartheta_0}(x)\Big] \\ &\quad + o_P(n^{-1/2}). \end{aligned} \tag{S1.2}$$

It follows that we can write

$$\begin{aligned}
&\hat{f}_{\hat{\varepsilon}(\vartheta_0)}(y) - f_{\varepsilon(\vartheta_0)}(y) \\
&= \frac{1}{ng} \sum_{i=1}^{n} \ell_g'(\varepsilon_i - y)(\hat{\varepsilon}_i(\vartheta_0) - \varepsilon_i) + \frac{1}{n} \sum_{i=1}^{n} \ell_g(\varepsilon_i - y) - f_\varepsilon(y) + o_P(n^{-1/2}) \\
&= -\frac{1}{ng} \sum_{i=1}^{n} \frac{\ell_g'(\varepsilon_i - y)}{\sigma(X_i)} \Big\{[\hat{m}_{\vartheta_0}(X_i) - m_{\vartheta_0}(X_i)] + \varepsilon_i[\hat{\sigma}_{\vartheta_0}(X_i) - \sigma_{\vartheta_0}(X_i)]\Big\} \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \ell_g(\varepsilon_i - y) - f_\varepsilon(y) + o_P(n^{-1/2}) \\
&= (T_1 + T_2)(y) + o_P(n^{-1/2}) \qquad \text{(say)}.
\end{aligned}$$

Using decompositions (S1.1) and (S1.2), we have

$$\begin{aligned}
T_1(y) &= -\frac{1}{ng}\sum_{i=1}^n \frac{\ell_g'(\varepsilon_i - y)}{\sigma(X_i)} \frac{1}{nh^d}\sum_{j=1}^n W_{X_i,n}\Big(\frac{X_i - X_j}{h}\Big)(\Lambda_{\vartheta_0}(Y_j) - m_{\vartheta_0}(X_i)) \\
&\quad -\frac{1}{ng}\sum_{i=1}^n \frac{\ell_g'(\varepsilon_i - y)}{\sigma^2(X_i)} \frac{\varepsilon_i}{2nh^d}\sum_{j=1}^n W_{X_i,n}\Big(\frac{X_i - X_j}{h}\Big)\Big((\Lambda_{\vartheta_0}(Y_j) - m_{\vartheta_0}(X_i))^2 - \sigma^2_{\vartheta_0}(X_i)\Big) \\
&\quad + o_P(n^{-1/2}) \\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n A_{nij}(\varepsilon_j + \frac{\varepsilon_i}{2}(\varepsilon_j^2 - 1)) + o_P(n^{-1/2}),
\end{aligned}$$

where $A_{nij} = -(gh^d)^{-1}\ell_g'(\varepsilon_i - y)W_{X_i,n}((X_i - X_j)/h)$. Using similar arguments as in Linton *et al.* (2008) and Colling and Van Keilegom (2015), the last expression can be written as

$$f'_{\varepsilon(\vartheta_0)}(y)\frac{1}{n}\sum_{i=1}^n \varepsilon_i + \Big(y f'_{\varepsilon(\vartheta_0)}(y) + f_{\varepsilon(\vartheta_0)}(y)\Big)\frac{1}{2n}\sum_{i=1}^n(\varepsilon_i^2 - 1) + o_P(n^{-1/2}).$$

In a similar way i.i.d. expansions for $\dot{\hat{f}}_{\hat\varepsilon(\vartheta_0)}(y) - \dot{f}_{\varepsilon(\vartheta_0)}(y)$ and $\hat{f}'_{\hat\varepsilon(\vartheta_0)}(y) - f'_{\varepsilon(\vartheta_0)}(y)$ can be obtained, which then extend Lemmas A.2 and A.3 in Linton *et al.* (2008) to the heteroscedastic case.

These three i.i.d. expansions all come together when we develop the i.i.d. expansion for $\hat\vartheta - \vartheta_0$. For the homoscedastic case this is done in Lemma A.11 in Linton *et al.* (2008), and it is shown there that all terms that come from the estimation of m, $\dot m$, f_ε, f'_ε and $\dot f_\varepsilon$ cancel and one therefore obtains the same expansion as in the case where all these functions would be known. In our heteroscedastic model a similar development can be done by using the above expansions for $\hat{f}_{\hat\varepsilon(\vartheta_0)}$, $\dot{\hat{f}}_{\hat\varepsilon(\vartheta_0)}$ and $\hat{f}'_{\hat\varepsilon(\vartheta_0)}$. We find in a similar way as in the homoscedastic case that all these expansions cancel out, and hence we get asymptotically the same i.i.d. expansion as in the case where these functions would be known. This shows the first part of Theorem 2.1. The second part follows immediately from the central limit theorem, together with the fact that $E[g_{\vartheta_0}(X,Y)] = G(\vartheta_0) = 0$. □

S2 Some auxiliary results

For $k = (k_1, \ldots, k_d) \in \mathbb{N}_0^d$, let $k. = \sum_{j=1}^d k_j$, $D^k = \partial^{k.}/\partial x_1^{k_1}\ldots\partial x_d^{k_d}$, and

$$\|f\|_{d+\alpha} = \max_{k.\le d}\sup_{x\in R_X}|D^k f(x)| + \max_{k.=d}\sup_{x,x'\in R_X}\frac{|D^k f(x) - D^k f(x')|}{\|x - x'\|^\alpha},$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$. Further, let $\mathcal{G}_1 = C_1^{d+\alpha}(R_X)$ be the class of d times differentiable functions f defined on R_X such that $\|f\|_{d+\alpha} \leq 1$, and $\mathcal{G}_2 = \tilde{C}_2^{d+\alpha}(R_X)$ be the class of d times differentiable functions f defined on R_X such that $\|f\|_{d+\alpha} \leq 2$ and $\inf_{x \in R_X} f(x) \geq 1/2$.

Proposition S2.1 *Let $\mathcal{F} = \{\varphi_{\vartheta,g_1,g_2,y} \mid \vartheta \in \Theta, g_1 \in \mathcal{G}_1, g_2 \in \mathcal{G}_2, y \in \mathbb{R}\}$, where*

$$\varphi_{\vartheta,g_1,g_2,y}(X,Y) = I\Big\{\frac{\Lambda_\vartheta(Y) - m(X)}{\sigma(X)} \leq y g_2(X) + g_1(X)\Big\} - I\Big\{\frac{\Lambda_{\vartheta_0}(Y) - m(X)}{\sigma(X)} \leq y\Big\}$$

is a function from $R_X \times \mathbb{R}$ to $\mathbb{R}$ and $\mathcal{G}_1$, $\mathcal{G}_2$ are defined above. Then $\mathcal{F}$ is Donsker.

Proof of Proposition S2.1 In Lemma 1 in Heuchenne *et al.* (2015) the special case of univariate X and $\sigma \equiv 1$ (i. e. homoscedasticity) is considered. For the subclass of $\mathcal{F}$ obtained by setting $g_2 \equiv 1$ the assertion is proved. On the other hand Lemma A.3 in Neumeyer and Van Keilegom (2010) shows the assertion for the function class defined analogously to $\mathcal{F}$, but replacing Λ_ϑ by the identity (for multivariate X). A detailed proof combining the arguments of both proofs is omitted for the sake of brevity. □

Proposition S2.2 *For the estimators $\hat{m}$ and $\hat{\sigma}$ defined in Section 2 and the function classes $\mathcal{G}_1$, $\mathcal{G}_2$ defined above we have under the assumptions of Theorem 3.1 that $P((\hat{m} - m)/\sigma \in \mathcal{G}_1) \to 1$ and $P(\hat{\sigma}/\sigma \in \mathcal{G}_2) \to 1$ for $n \to \infty$.*

Proof of Proposition S2.2 Note that the assertion follows from $\|\hat{m} - m\|_{d+\alpha} = o_P(1)$ and $\|\hat{\sigma} - \sigma\|_{d+\alpha} = o_P(1)$. Further note that

$$\hat{m} - m = (\hat{m}_{\vartheta_0} - m) + (\hat{m}_{\hat{\vartheta}} - \hat{m}_{\vartheta_0}), \quad \hat{\sigma} - \sigma = (\hat{\sigma}_{\vartheta_0} - \sigma) + (\hat{\sigma}_{\hat{\vartheta}} - \hat{\sigma}_{\vartheta_0})$$

and that $\|\hat{m}_{\vartheta_0} - m\|_{d+\alpha} = o_P(1)$, $\|\hat{\sigma}_{\vartheta_0} - \sigma\|_{d+\alpha} = o_P(1)$ was shown in Lemma A.1 in Neumeyer and Van Keilegom (2010) under assumptions (a1), (a2), (A1)–(A3). We will apply Taylor expansions for the remainder terms. To this end due to $\hat{\vartheta} = \vartheta_0 + o_P(1)$ (see assumption (A5)) we may assume that $\|\hat{\vartheta} - \vartheta_0\| \leq \eta$ for η from assumption (A7). Denote by $\widehat{\tilde{m}}_{\vartheta_0}$ a local polynomial estimator defined analogously to $\hat{m}_{\vartheta_0}$, but based on the sample $(X_i, \dot{\Lambda}_{\vartheta_0}(Y_i))$, $i = 1, \ldots, n$. Let, by slight abuse of notation,

$$d^k V_{x,n}(z) = \frac{\partial^{k\cdot}(W_{x,n}(\frac{x-z}{h}))}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}}$$

for $k = (k_1, \dots, k_d) \in \mathbb{N}_0^d$, with $W_{x,n}$ from (S1.1). Then we obtain from (B.1) that

$$
\begin{aligned}
& \|\hat{m}_{\hat{\vartheta}} - \hat{m}_{\vartheta_0}\|_{d+\alpha} \\
\leq\ & \|\hat{\vartheta} - \vartheta_0\| \|\widehat{\tilde{m}}_{\vartheta_0}\|_{d+\alpha} && \text{(S2.1)} \\
& + \frac{1}{2}\|\hat{\vartheta} - \vartheta_0\|^2 \max_{k. \leq d} \frac{1}{nh^d} \sum_{i=1}^n \sup_{x \in R_X} |d^k V_{x,n}(X_i)| \sup_{\|\vartheta - \vartheta_0\| \leq \eta} \|\ddot{\Lambda}_\vartheta(Y_i)\| && \text{(S2.2)} \\
& + \frac{1}{2}\|\hat{\vartheta} - \vartheta_0\|^2 \max_{k. = d} \frac{1}{nh^d} \sum_{i=1}^n \sup_{x, x' \in R_X} \frac{|d^k V_{x,n}(X_i) - d^k V_{x',n}(X_i)|}{\|x - x'\|^\alpha} \sup_{\|\vartheta - \vartheta_0\| \leq \eta} \|\ddot{\Lambda}_\vartheta(Y_i)\|. && \text{(S2.3)}
\end{aligned}
$$

Under assumptions (a1), (a2), (A1) and (A8) we have that $\|\widehat{\tilde{m}}_{\vartheta_0}\|_{d+\alpha}$ converges to $\|\tilde{m}_{\vartheta_0}\|_{d+\alpha}$ in probability, where $\tilde{m}_{\vartheta_0}(\cdot) = E[\dot{\Lambda}_{\vartheta_0}(Y)|X = \cdot]$. Thus (S2.1) is negligible since $\|\hat{\vartheta} - \vartheta_0\| = O_P(n^{-1/2})$. Under assumptions (a1) and (a2), from the representations of the multivariate local polynomial estimator in Masry (1996a, 1996b) one can deduce that $h^d \sup_{x,z} |d^k V_{x,n}(z)|$ is bounded (for $k. \leq d$). Thus applying the law of large numbers to $\sup_{\|\vartheta - \vartheta_0\| \leq \eta} \|\ddot{\Lambda}_\vartheta(Y_i)\|$ (compare to assumption (A7)) for (S2.2) we obtain the order $O_P(\|\hat{\vartheta} - \vartheta_0\|^2 h^{-2d}) = o_P(1)$ by assumption (a2). Further, by considering the cases $\|x - x'\| \geq h$ and $\|x - x'\| < h$ one obtains

$$
\sup_{x, x' \in R_X} \frac{|d^k V_{x,n}(X_i) - d^k V_{x',n}(X_i)|}{\|x - x'\|^\alpha} \leq 2 \sup_{x,z} |d^k V_{x,n}(z)| \frac{1}{h^\alpha} + \sum_{j=1}^d \sup_{x,z} \left| \frac{\partial d^k V_{x,n}(z)}{\partial x_j} \right| h^{1-\alpha}.
$$

All partial derivatives of order one of $h^{d+1} d^k V_{x,n}(z)$ in x-direction are bounded in x, z. Thus for (S2.3) one obtains the rate $O_P(\|\hat{\vartheta} - \vartheta_0\|^2 (h^{-(2d+\alpha)} + h^{-(2d+1-(1-\alpha))})) = o_P(1)$ by assumption (a2). Similar arguments hold for $\hat{\sigma}_{\hat{\vartheta}} - \hat{\sigma}_{\vartheta_0}$. $\square$

Proposition S2.3 *With the definitions in Proposition S2.1 we have under the assumptions of Theorem 3.1 that* $E[(\varphi_{\hat{\vartheta}, (\hat{m}-m)/\sigma, \hat{\sigma}/\sigma, y}(X, Y) - \varphi_{\vartheta_0, 0, 1, y}(X, Y))^2 \mid \mathcal{Y}_n] = o_P(\delta_n^2)$ *uniformly with respect to* $y \in \mathbb{R}$ *with some* $\delta_n \searrow 0$ *for* $n \to \infty$, *where* $\mathcal{Y}_n = \{(X_i, Y_i) : i = 1, \dots, n\}$.

Proof of Proposition S2.3 Note that $\varphi_{\vartheta_0, 0, 1, y} \equiv 0$. The expectation in the assertion can be bounded by the sum

$$
\begin{aligned}
& 2E[(\varphi_{\hat{\vartheta}, (\hat{m}-m)/\sigma, \hat{\sigma}/\sigma, y}(X, Y) - \varphi_{\vartheta_0, (\hat{m}-m)/\sigma, \hat{\sigma}/\sigma, y}(X, Y))^2 \mid \mathcal{Y}_n] && \text{(S2.4)} \\
& + 2E[(\varphi_{\vartheta_0, (\hat{m}-m)/\sigma, \hat{\sigma}/\sigma, y}(X, Y))^2 \mid \mathcal{Y}_n]. && \text{(S2.5)}
\end{aligned}
$$

We first consider (S2.4) which equals

$$\begin{aligned}
& E[(I\{\Lambda_{\hat\vartheta}(Y) \leq y\hat\sigma(X) + \hat m(X)\} - I\{\Lambda_{\vartheta_0}(Y) \leq y\hat\sigma(X) + \hat m(X)\})^2 \mid \mathcal{Y}_n] \\
\leq & \int |F_{Y|X}(V_{\hat\vartheta}(y\hat\sigma(x) + \hat m(x))|x) - F_{Y|X}(V_{\vartheta_0}(y\hat\sigma(x) + \hat m(x))|x)| \, dF_X(x)
\end{aligned}$$

with the notations from the proof of Theorem 3.1. Note that this term is very similar to A_n in that proof, only that an absolute value is added inside the integral. With the same methods as there the rate $O_P(n^{-1/2})$ can be shown.

Next we consider (S2.5) which equals

$$\begin{aligned}
& E\Big[\Big(I\Big\{\varepsilon \leq y\frac{\hat\sigma(X)}{\sigma(X)} + \frac{\hat m(X) - m(X)}{\sigma(X)}\Big\} - I\{\varepsilon \leq y\}\Big)^2 \mid \mathcal{Y}_n\Big] \\
\leq & \int \Big| F_\varepsilon\Big(y\frac{\hat\sigma(x)}{\sigma(x)} + \frac{\hat m(x) - m(x)}{\sigma(x)}\Big) - F_\varepsilon(y)\Big| \, dF_X(x) \\
\leq & \sup_{y\in\mathbb{R}} |f_\varepsilon(\xi_n(y))| \int \Big|\frac{\hat m(x) - m(x)}{\sigma(x)}\Big| \, dF_X(x) + \sup_{y\in\mathbb{R}} |y f_\varepsilon(\xi_n(y))| \int \Big|\frac{\hat\sigma(x) - \sigma(x)}{\sigma(x)}\Big| \, dF_X(x)
\end{aligned}$$

where $\xi_n(y)$ converges to y in probability. Hence the supremum terms are bounded thanks to assumption (A3). Further, using the decomposition $\hat m - m = (\hat m_{\vartheta_0} - m) + (\hat m_{\hat\vartheta} - \hat m_{\vartheta_0})$ as in the proof of Proposition S2.2 (and similar for $\hat\sigma$) one can show the rate $O_P((nh^d/\log n)^{-1/2}) + O_P(n^{-1/2})$. This proves the assertion. □

References

Colling, B. and Van Keilegom, I. (2015). Goodness-of-fit tests in semiparametric transformation models. *TEST*, to appear.

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

Gu, J., Li, Q. and Yang, J-C. (2014). Multivariate local polynomial kernel estimators: Leading bias and asymptotic distribution. *Econometric Reviews*, to appear.

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

Masry, E. (1996a). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Ser. Anal.* **17**, 571–599.

Masry, E. (1996b). Multivariate regression estimation – local polynomial fitting for time series. *Stochastic Process. Appl.* **65**, 81–101.

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