Estimation of the Boundary of a Variable observed with Symmetric Error

Jean-Pierre Florens [*]
jean-pierre.florens@tse-fr.eu

Léopold Simar[*, §]
leopold.simar@uclouvain.be

Ingrid Van Keilegom [**]
ingrid.vankeilegom@kuleuven.be

September 17, 2018

Abstract

Consider the model $Y = X + \varepsilon$ with $X = \tau + Z$, where τ is an unknown constant (the boundary of X), Z is a random variable defined on $\mathbb{R}^+$, ε is a symmetric error, and ε and Z are independent. Based on a iid sample of Y we aim at identifying and estimating the boundary τ when the law of ε is unknown (apart from symmetry) and in particular its variance is unknown. We propose an estimation procedure based on a minimal distance approach and by making use of Laguerre polynomials. Asymptotic results as well as finite sample simulations are shown. The paper also proposes an extension to stochastic frontier analysis, where the model is conditional to observed variables. The model becomes $Y = \tau(w_1, w_2) + Z + \varepsilon$, where Y is a cost, w_1 are the observed outputs and w_2 represents the observed values of other conditioning variables, so Z is the cost inefficiency. Some simulations illustrate again how the approach works in finite samples, and the proposed procedure is illustrated with data coming from post offices in France.

Key Words: Characteristic function; cumulant function; flexible parametric family; frontier estimation; Laguerre polynomials.

[*]Toulouse School of Economics, Université Toulouse Capitole.
[§]ISBA, Université Catholique de Louvain.
[**]ORSTAT, KU Leuven. Financial support from the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409), and from IAP Research Network P7/06 of the Belgian State, are gratefully acknowledged.

1 Introduction

The objective of this paper is to identify and estimate the lower endpoint of the support of a random variable that is bounded from below and that is observed with error when 'minimal' assumptions are imposed on the error term, in the sense that the error is only assumed to be symmetric around zero with no distributional assumptions. The problem of estimating the minimum of a random variable that is observed with noise, has received a lot of attention in the literature. Consider the relation $Y = X+\varepsilon = \tau+Z+\varepsilon$, where Y is the observed variable, X is the variable of interest, τ is the lower endpoint of the support of X, $Z \geq 0$, ε is the noise, and Z and ε are independent. When the distribution of ε is known, a large body of papers exists on the estimation of the support of the variable X, most of which are based on deconvolution techniques. See e.g. Goldenshluger and Tsybakov (2004), Delaigle and Gijbels (2006), Meister (2006a), and Aarts, Groeneboom and Jongbloed (2007), among others. The assumption that the distribution of ε is given, is a strong hypothesis (see Daouia et al., 2018). However, it has been shown that thanks to the positivity of Z the variance of ε is identified if ε is normally distributed (see Schwarz and Van Bellegem, 2010), and in that case the minimum τ may be estimated nonparametrically (see Hall and Simar, 2002, and Kneip et al., 2015). Also note that a different but related literature is on the identification and estimation of the density of X under certain smoothness assumptions on the density of X and ε. This has been studied in Butucea and Matias (2005), Meister (2006b, 2007), Butucea et al. (2008), Schwarz and Van Bellegem (2010) and Delaigle and Hall (2016), among others.

This paper goes one step further in the sense that we weaken the hypotheses on the stochastic error ε: we only assume that ε is symmetric around zero. Our approach is in the line of Delaigle and Hall (2016), who estimate the density of X using the empirical phase function. See also Butucea and Vandekerkhove (2014) and Butucea et al. (2017) for somewhat related papers, in which ε has a symmetric density and X is discrete. The price of this weak assumption on ε is that the distribution of X is no longer identified. As we will show below, only the odd cumulants of the distribution of X are identified, which is not sufficient to identify in a unique way the distribution. However, a precise parametric but flexible approximation of the density of X may be estimated from the odd cumulants.

This model may be extended to conditional models. Our main motivation for such an extension is to analyze the so-called stochastic frontier models. Let us assume that X is a cost generated conditionally on variables $W = (W_1, W_2)$. The distribution of X is conditional to $W_1 \geq w_1$ in $\mathbb{R}^{p_1}$ and to $W_2 = w_2$ in $\mathbb{R}^{p_2}$. In econometric applications W_1 represents the outputs and W_2 describes the conditions of the production process. We are interested in the

minimal cost given that $W_1 \geq w_1$ and $W_2 = w_2$, which we denote by $\tau(w_1, w_2)$. If we define $Z = X - \tau(w_1, w_2)$, the usual expression of the stochastic frontier model is

$$Y = \tau(w_1, w_2) + Z + \varepsilon,$$

where Z is positive, the error ε has a symmetric but unknown density, and ε and Z are independent conditionally on $W_1 \geq w_1$ and $W_2 = w_2$. This model has been studied extensively in the case where Z, ε and τ have a parametric form (for example, a Cobb-Douglas model for τ, an exponential distribution for Z, and a normal distribution for ε), see Aigner et al. (1977), Meeusen and van den Broek (1977). Some semiparametric approaches have also been proposed where for instance the frontier function is unspecified but the stochastic part is fully parametric (see e.g. Fan et al., 1996 or Kuosmanen and Kortelainen, 2012). We extend this analysis to the case where the density of ε is only assumed to be symmetric.

The paper is organized as follows. In Section 2 we discuss the identifiability of the marginal model. Estimation in the case without conditioning variables is treated in Section 3. A tractable flexible parametric family of distributions based on Laguerre polynomials is considered in Section 4, together with finite sample simulations. Section 5 presents the extension to frontier estimation, with some simulations and an illustration based on data coming from post offices in France. Conclusions and further research are given in Section 6. The proofs of the main results are gathered in Appendix A, whereas Appendix B collects some properties of Laguerre polynomials. Finally, Appendix C gives some details about the practical implementation of our estimators.

2 Identifiability of the model

We consider the following model. We observe the random variable Y defined as

$$Y = X + \varepsilon = \tau + Z + \varepsilon, \tag{2.1}$$

where Z is an unobservable random variable with support starting at 0 and with density g, τ is an unknown constant corresponding to the lower endpoint of the support of the random variable $X = \tau + Z$, Y has density f, the error ε has a symmetric density h (around zero), and ε and Z are independent. The model parameters are τ, g and h, but we are mostly interested in estimating τ. Note that we do not specify the distribution of ε, nor its variance, we only assume symmetry. We suppose throughout the paper that all variables are continuous and have an analytic characteristic function, and in particular that all moments

exist (see Lukacs, 1960, Chapter 7). The characteristic function of Y (and similarly for all other variables) will sometimes be written as ψ_Y and sometimes as ψ_f, depending on the context.

Let $\theta_0 = (\tau_0, g_0, h_0)$ be the true parameter values, and let $\Theta = \mathbb{R} \times \mathcal{G} \times \mathcal{H}$, where $\mathcal{G}$ is the class of densities living on $[0, \infty)$, and $\mathcal{H}$ is the class of symmetric densities around zero. We suppose that $\theta_0 \in \Theta$.

We know that the true density f of Y satisfies

$$f(y) = \int_{-\infty}^{+\infty} g_0(y - \tau_0 - e) h_0(e)\, de \tag{2.2}$$

for all y. The question we like to answer is how to identify τ, g and h from (2.2) given that g lives on $[0, \infty)$ and that h is symmetric, or in other words under which conditions does the following hold true : if for some $(\tau, g, h) \in \Theta$ we have that $f(y) = \int g(y - \tau - e)h(e)\, de$ for all y, then necessarily $\tau = \tau_0, g = g_0$ and $h = h_0$?

It is clear that additional assumptions are needed to identify the model. Consider for example model (2.1) with $\tau = 2$, $Z \sim \text{Exp}(1)$ and ε distributed according to the independent sum of a $\text{Un}[-1, 1]$ variable and a standard normal variable. Then, exactly the same density f will be obtained if $\tau = 1$, Z is distributed according to the independent sum of a $\text{Exp}(1)$ and a $\text{Un}[0, 2]$ variable, and ε is standard normal. More generally, if in model (2.1) we take

$$\tau = \tilde{\tau} - c, \quad Z = c + \tilde{Z} + U, \quad \varepsilon = \tilde{\varepsilon} - U,$$

with $c \geq 0$, U independent of $(\tilde{Z}, \tilde{\varepsilon})$, $\tilde{Z}$ and $\tilde{\varepsilon}$ independent, the density of U symmetric around zero and the support of U equal to $[-c, c]$, then we can equally well write Y as $Y = \tilde{\tau} + \tilde{Z} + \tilde{\varepsilon}$. Note that this shows that the constraints on Z to make the model identifiable are related to the concept of indecomposability of Z, which means that Z cannot be decomposed as the sum of two independent non-constant random variables (see Lukacs, 1960, 1983 for more details about indecomposability of a distribution). Here it is necessary that such a decomposition is not possible if one of the two variables is required to be symmetric, which we call *symmetric indecomposability*. The following lemma gives sufficient conditions on a positive random variable Z to be symmetric indecomposable.

Lemma 2.1 *Let ψ_Z be the characteristic function of a positive random variable Z, and suppose that ψ_Z has at most a finite number of zeros. Then, Z is symmetric indecomposable.*

Proof. Suppose that Z would be symmetric decomposable and could be written as $Z = \tilde{Z} + U$ with U symmetric around zero and $\tilde{Z}$ independent of U. Since $Z > 0$, we necessarily

have that $|U| \leq a$ for some finite $a > 0$ and $\tilde{Z} \geq a$, because $\min Z = \min \tilde{Z} + \min U = 0$. Then, by Theorem 7.2.3 in Lukacs (1960), the characteristic function ψ_U of U has an infinite number of zeros. Since $\psi_Z = \psi_{\tilde{Z}}\psi_U$, this contradicts the assumption that ψ_Z has at most a finite number of zeros. □

So, for example, the exponential, the half-normal and the truncated normal distribution are symmetric indecomposable, since $\psi_Z(t)$ has no zeros in these cases. The same is true for the gamma distribution or for finite linear combinations of these, since $\psi_Z(t)$ is in these cases a polynomial of finite order in t with at most a finite number of zeros. So the above lemma indicates that the symmetric indecomposability of a distribution defined on the positive real line is rather common. But still this is not a universal property as the simple example above reminds us.

In order to formulate conditions that guarantee the identifiability of our model, we follow a somewhat similar approach as Delaigle and Hall (2016), who studied the identifiability of the model $Y = X + \varepsilon$, where ε has a symmetric density (as in our case), but they do not assume that the support of X has a finite lower bound. Assumption (2.3) in their paper is a crucial assumption to identify the law of X, and states that the distribution of X uniquely has least variance among all distributions sharing the phase function $\psi_Y/\|\psi_Y\|$, where $\|\cdot\|$ is the modulus. Note that this implies that X and hence Z must be symmetric indecomposable. [1] One way to identify our model would therefore be to impose their assumption (2.3) restricted to the class $\mathcal{F}$ of densities living on $[\tau, \infty)$ with τ unknown, since τ is identified as soon as the law of X is identified. However, we are in a more particular framework than Delaigle and Hall (2016) and we feel that it is more natural to disentangle Z and τ in our model. Therefore, we will look for alternative assumptions to identify the density g of Z, since once g is identified we know τ via the relation

$$E(Y) = \tau + E(Z).$$

Our approach is based on cumulants, where for $\ell = 1, 2, \ldots$, the ℓ-th cumulant of X is defined as $\kappa_\ell(f) = K_f^{(\ell)}(0)$, and the cumulant generating function $K_f(t)$ is given by $K_f(t) = \log E[\exp(tY)]$. Using standard properties of cumulants, we have that

$$\kappa_1(f) = E(Y) = \tau + E(Z) + E(\varepsilon) = \tau + \kappa_1(g).$$

[1] Note that the opposite is not true : if X is symmetric indecomposable, it does not necessarily imply that X is identified.

On the other hand, for $p \geq 1$,

$$\kappa_{2p+1}(f) = \kappa_{2p+1}(\delta_\tau) + \kappa_{2p+1}(g) + \kappa_{2p+1}(h) = \kappa_{2p+1}(g),$$

with δ_τ the Dirac measure at τ, since $\kappa_{2p+1}(\delta_\tau) = 0$ and since odd cumulants of symmetric densities are zero (because of formula (2.3) below, and the fact that the characteristic function of a symmetric function is real). So the odd cumulants of g of order 3 or higher are identifiable. We also know that again for $p \geq 1$,

$$\kappa_{2p}(f) = \kappa_{2p}(g) + \kappa_{2p}(h),$$

and since the even cumulants of h do not vanish in general, the even cumulants of g are not identifiable. We therefore introduce a class of densities of Z denoted by $\mathcal{G}$, that satisfies the following hypothesis :

(A1) The density g_0 belongs to $\mathcal{G}$, and if $\kappa_{2p+1}(g) = \kappa_{2p+1}(g_0)$ for $p \geq 1$ and for some $g \in \mathcal{G}$, then $g = g_0$.

Then as $\kappa_{2p+1}(f) = \kappa_{2p+1}(g_0)$, the model restricted to the class $\mathcal{G}$ becomes identified.

In particular we will consider cases where $\mathcal{G}$ is parametrized by a vector of parameters $\lambda \in \Lambda$ of finite dimension :

$$\mathcal{G} = \{g(\cdot|\lambda) : \lambda \in \Lambda \subset \mathbb{R}^m, \Lambda \text{ compact}\}.$$

In Delaigle and Hall (2016) identification is obtained by the selection of the distribution of smallest variance in the class of equivalent densities (i.e. defining the same f).

In order to formulate assumption (A1) in a different way, we introduce now the function η_f (and similarly for η_g) :

$$\eta_f(t) = \frac{\partial}{\partial t}\Big[\Im \log \psi_f(t)\Big],$$

where $\Im(a + ib) = b$ for any complex number $a + ib$, and where for any complex number $z = \rho \exp(i\theta)$ with $\theta \in (-\pi, \pi]$, the logarithm of z is defined as $\log z = \log \rho + i\theta$. Note that

$$\log \psi_f(t) = \sum_{p=1}^{\infty} \frac{(it)^p}{p!}\kappa_p(f) = \sum_{p=1}^{\infty} \frac{(-t^2)^p}{(2p)!}\kappa_{2p}(f) + i\sum_{p=1}^{\infty} \frac{(-1)^{p+1}t^{2p-1}}{(2p-1)!}\kappa_{2p-1}(f), \tag{2.3}$$

and hence

$$\Im \log \psi_f(t) = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}t^{2p-1}}{(2p-1)!}\kappa_{2p-1}(f),$$

or

$$\eta_f(t) = \kappa_1(f) + \sum_{p=2}^{\infty} \frac{(-1)^{p+1}t^{2p-2}}{(2p-2)!}\kappa_{2p-1}(f) = \kappa_1(f) + \sum_{p=1}^{\infty} \frac{(-1)^p t^{2p}}{(2p)!}\kappa_{2p+1}(f).$$

This shows that we can write

$$\eta_f(t) - \eta_f(0) = \sum_{p=1}^{\infty} \frac{(-1)^p t^{2p}}{(2p)!} \kappa_{2p+1}(f),$$

which depends only on $\kappa_{2p+1}(f)$ for $p \geq 1$. The above derivations now lead to the following equivalent formulation of assumption (A1) :

(A1') The density g_0 belongs to $\mathcal{G}$, and if $\eta_g(t) - \eta_g(0) = \eta_{g_0}(t) - \eta_{g_0}(0)$ for all t and for some $g \in \mathcal{G}$, then $g = g_0$.

Note that the mean (first cumulant) $\kappa_1(f)$ is not identified, as it contains the unknown parameter τ. By considering the function $\eta_g(t) - \eta_g(0)$ instead of the function $\eta_g(t)$ the mean cancels out, and we end up with an identified function. If we denote by $T_\lambda(t)$ the function $\eta_g(t) - \eta_g(0)$ with $g \equiv g(\cdot|\lambda)$, the identification condition means that for all $\lambda \in \Lambda$,

$$T_\lambda \equiv T_{\lambda_0} \Rightarrow \lambda = \lambda_0.$$

In a non-linear context local identification is a useful concept and is defined as follows : there exists a neighborhood $\mathcal{V}$ of λ_0 such that

$$T_\lambda \equiv T_{\lambda_0} \text{ for some } \lambda \in \mathcal{V} \Rightarrow \lambda = \lambda_0.$$

In this paper we identify $\mathcal{V}$ without loss of generality with Λ. We denote now by $\kappa_\ell(\lambda)$ the ℓ-th cumulant of $g(\cdot|\lambda)$. Then, the equality $T_\lambda \equiv T_{\lambda_0}$ is equivalent to the property that if $\kappa_{2p+1}(\lambda) = \kappa_{2p+1}(\lambda_0)$ for all $p \geq 1$, then $\lambda = \lambda_0$. The implicit function theorem implies that if $\kappa_{2p+1}(\lambda_0)$ is differentiable with respect to λ, then there exists a $P \geq m$ such that the matrix

$$\Big(\frac{\partial}{\partial \lambda_j} \kappa_{2p+1}(\lambda) \Big|_{\lambda=0} \Big)_{\substack{p=1,\ldots,P \\ j=1,\ldots,m}}$$

has rank equal to m.

To conclude, if λ is identified, the distribution of Z is also identified, and so $E(Z)$ and $E(Z^2)$ are identified as well as a function of λ. Since $\tau_0 = E(Y) - E(Z)$ and $\sigma_0^2 = \mathrm{Var}(\varepsilon) = \mathrm{Var}(Y) - \mathrm{Var}(Z)$, the identification of τ_0 and σ_0^2 is guaranteed as well.

3 Estimation of the model

We start by explaining the methodology, followed by the asymptotic properties of the proposed estimators. Throughout this section the parametric family $\mathcal{G}$ can be any parametric class which is such that it contains the true density $g(\cdot|\lambda_0)$ and such that the cumulants of odd order starting from order 3 identify the densities in the class. In Section 4 we will work with one particular class of densities determined by Laguerre polynomials.

3.1 Methodology

Recall that the class $\mathcal{G}$ is a parametric class of densities $g(\cdot|\lambda)$ defined on $\mathbb{R}^+$ and indexed by $\lambda \in \Lambda \subset \mathbb{R}^m$. We equip Λ with the canonical inner product $\langle \lambda, \tilde{\lambda} \rangle_{\mathbb{R}^m} = \sum_{j=1}^m \lambda_j \tilde{\lambda}_j$ (we will suppress the subscript $\mathbb{R}^m$ in $\langle \cdot, \cdot \rangle_{\mathbb{R}^m}$ whenever there is no ambiguity). Let $\eta_\lambda(t) = \frac{\partial}{\partial t}\big[\Im \log \psi_\lambda(t)\big]$, where ψ_λ is the characteristic function corresponding to $g(\cdot|\lambda)$. We suppose that the function $T_\lambda(\cdot) = \eta_\lambda(\cdot) - \eta_\lambda(0)$ belongs to the space $\mathcal{E}$ of functions from $\mathbb{R}$ to $\mathbb{C}$ such that $\langle v_1, v_2 \rangle_{\mathcal{E}} = \int v_1(t)\overline{v_2(t)}w(t)\,dt < \infty$ for $v_1, v_2 \in \mathcal{E}$, for a certain measure w, and where $\overline{v}$ is the complex conjugate of v. (Note that $T_\lambda(\cdot)$ takes values in $\mathbb{R}$, which we embed in $\mathbb{C}$.) We suppose that the Fréchet derivative $T'_\lambda : \Lambda \to \mathcal{E}$ exists and that T_λ and T'_λ are injective. The linear operator T'_λ is bounded, and hence the dual operator T'^*_λ, defined by

$$\langle T'_\lambda(\lambda), v \rangle_{\mathcal{E}} = \langle \lambda, T'^*_\lambda(v) \rangle_{\mathbb{R}^m}$$

exists, for any $v \in \mathcal{E}$. Note that $T'^*_\lambda T'_\lambda : \mathbb{R}^m \to \mathbb{R}^m$ is regular (since T'_λ is injective), but often ill-conditioned, and hence we will use penalization techniques to estimate λ_0.

Define $\hat{T}(t) = \hat{\eta}(t) - \hat{\eta}(0)$, where $\hat{\eta}(t) = \frac{\partial}{\partial t}\big[\Im \log \hat{\psi}_Y(t)\big]$ and $\hat{\psi}_Y(t) = n^{-1}\sum_{j=1}^n \exp(itY_j)$, where $Y_1, \ldots, Y_n$ is an i.i.d. sample of Y. Then, the estimator of λ_0 is defined as

$$\begin{aligned} \hat{\lambda}_\alpha &= \operatorname{argmin}_{\lambda\in\Lambda}\Big\{ \int_{-\infty}^{+\infty} \big(\hat{T}(t) - T_\lambda(t)\big)^2 w(t)\,dt + \alpha P(\lambda)\Big\} \\ &= \operatorname{argmin}_{\lambda\in\Lambda}\Big\{ \|\hat{T} - T_\lambda\|^2_{\mathcal{E}} + \alpha\|\lambda\|^2_{\mathbb{R}^m}\Big\}, \end{aligned} \tag{3.1}$$

where $P(\lambda) = \|\lambda\|^2$ and $\alpha > 0$ is a regularization parameter. We could also work with other penalties like sparsity penalties ($P(\lambda) = \|\lambda\|$) or entropy penalties ($P(\lambda) = \sum_{j=1}^m \lambda_j \log|\lambda_j|$), but we do not consider them here for reasons of brevity. Note that instead of working with $P(\lambda) = \|\lambda\|^2$, which favors models for which λ is close to zero, we could also work with $P(\lambda) = \|\lambda - \lambda^*\|^2$, if we have an a priori guess of the value of λ. Properties of $\hat{\lambda}_\alpha$ are not affected by the choice of λ^*.

Remark 3.1 *We assume that there exists a parametric family Λ of dimension m such that $g = g(\cdot|\lambda_0)$ for some $\lambda_0 \in \Lambda$. This assumption is not restrictive if m may be large (but finite). In that case, the model remains identified but severely ill-conditioned (i.e. the ratio of the largest and smallest eigenvalue of $T'^*_{\lambda_0}T'_{\lambda_0}$ is very high). The penalization solves the numerical problems caused by this ill-conditioning.*

Note that for all $\lambda \in \Lambda$, $\hat{\lambda}_\alpha$ satisfies

$$-\langle T'_{\hat{\lambda}_\alpha}(\lambda), \hat{T} - T_{\hat{\lambda}_\alpha} \rangle + \alpha\langle \lambda, \hat{\lambda}_\alpha \rangle = -\langle \lambda, T'^*_{\hat{\lambda}_\alpha}(\hat{T} - T_{\hat{\lambda}_\alpha}) \rangle + \alpha\langle \lambda, \hat{\lambda}_\alpha \rangle = 0,$$

and hence

$$T'^{*}_{\hat{\lambda}_\alpha}(\hat{T} - T_{\hat{\lambda}_\alpha}) - \alpha\hat{\lambda}_\alpha = 0. \tag{3.2}$$

For what follows, we need to define the regularized version λ_α of λ_0 :

$$\lambda_\alpha = \text{argmin}_{\lambda\in\Lambda}\Big\{\|T_{\lambda_0} - T_\lambda\|^2_{\mathcal{E}} + \alpha\|\lambda\|^2_{\mathbb{R}^m}\Big\},$$

which satisfies (using a similar derivation as for $\hat{\lambda}_\alpha$)

$$T'^{*}_{\lambda_\alpha}(T_{\lambda_0} - T_{\lambda_\alpha}) - \alpha\lambda_\alpha = 0. \tag{3.3}$$

Finally, note that since $E(Z) = a(\lambda)$ and $\text{Var}(Z) = b(\lambda)$ for some known functions a and b, we may estimate τ and σ^2 by

$$\hat{\tau}_\alpha = \overline{Y} - a(\hat{\lambda}_\alpha) \qquad \text{and} \qquad \hat{\sigma}^2_\alpha = \hat{\sigma}^2_Y - b(\hat{\lambda}_\alpha), \tag{3.4}$$

where $\overline{Y}$ and $\hat{\sigma}^2_Y$ are the empirical average and variance of Y, respectively. Appendix C gives some details about the practical implementation of the whole method.

3.2 Asymptotic properties

The proofs of the results of this subsection are deferred to Appendix A, as well as the conditions under which these results are valid. We start with the weak convergence of the process $(\hat{T}(\cdot) - T_{\lambda_0}(\cdot))$ as a process in $\mathcal{E}$, equipped with $\langle\cdot,\cdot\rangle_{\mathcal{E}}$.

Theorem 3.1 *The process $n^{1/2}(\hat{T}(\cdot)-T_{\lambda_0}(\cdot))$ converges weakly in $\mathcal{E}$ to a mean-zero Gaussian process $W(\cdot)$ with covariance operator Γ given by $(\Gamma\psi)(t) = \int\gamma(s,t)\psi(s)w(s)ds$ with*

$$\gamma(s,t) = \sigma(s,t) - \sigma(s,0) - \sigma(0,t) + \sigma(0,0),$$

where

$$\sigma(s,t) = \frac{1}{4}\frac{\partial^2}{\partial s\partial t}\Big\{\frac{\psi_Y(t-s)}{\psi_Y(t)\overline{\psi_Y(s)}} + \frac{\overline{\psi_Y(t-s)}}{\overline{\psi_Y(t)}\psi_Y(s)} - \frac{\psi_Y(t+s)}{\psi_Y(t)\psi_Y(s)} - \frac{\overline{\psi_Y(t+s)}}{\overline{\psi_Y(t)}\,\overline{\psi_Y(s)}}\Big\}.$$

Next, we consider the weak consistency and asymptotic normality of $\hat{\lambda}_\alpha$ for fixed $\alpha > 0$. Note that the term $\nabla(T'^{*}_\lambda S)|_{S=T_{\lambda_\alpha}-T_{\lambda_0},\lambda=\lambda_\alpha}$ in the formula of the asymptotic variance is non-standard and is caused by the non-linearity of the operator T_λ, where ∇ denotes the gradient vector with respect to λ.

Theorem 3.2 *Assume (C1)-(C4). Then, for fixed $\alpha > 0$, $\hat{\lambda}_\alpha \to_P \lambda_\alpha$, and*

$$n^{1/2}(\hat{\lambda}_\alpha - \lambda_\alpha) \xrightarrow{d} N(0, \Omega),$$

where

$$\Omega = \Delta^{-1} T'^{*}_{\lambda_\alpha} \Gamma T'_{\lambda_\alpha} \Delta^{-1},$$

$$\Delta = \alpha I + T'^{*}_{\lambda_\alpha} T'_{\lambda_\alpha} + \nabla (T'^{*}_{\lambda} S)|_{S = T_{\lambda_\alpha} - T_{\lambda_0}, \lambda = \lambda_\alpha}.$$

The variance matrix Ω depends on T'_λ and T'^{*}_λ, which can be calculated analytically, although their formula might be complicated. Let us consider T'^{*}_λ. Define $\beta_{j\lambda} = T'_\lambda e_j$, where e_j is the vector in $\mathbb{R}^m$ with j-th component equal to one and all other components equal to zero. Then, for all $\tilde{\lambda} \in \Lambda$, $T'_\lambda \tilde{\lambda} = T'_\lambda \big(\sum_{j=1}^m \tilde{\lambda}_j e_j \big) = \sum_{j=1}^m \tilde{\lambda}_j \beta_{j\lambda}$, and therefore

$$\langle \tilde{\lambda}, T'^{*}_\lambda a \rangle_{\mathbb{R}^m} = \langle T'_\lambda \tilde{\lambda}, a \rangle_{\mathcal{E}} = \sum_{j=1}^m \tilde{\lambda}_j \langle \beta_{j\lambda}, a \rangle_{\mathcal{E}} = \langle \tilde{\lambda}, \langle \beta_\lambda, a \rangle_{L^2} \rangle_{\mathbb{R}^m}.$$

Hence, $T'^{*}_\lambda a = \langle \beta_\lambda, a \rangle$.

Note that the choice of α is an important but delicate issue. We refer to Theorem 10.7 in Engl et al. (1996) for a result based on balancing the variance and the squared bias of $\hat{\lambda}_\alpha$.

Next, we consider the asymptotic normality of $\hat{\tau}_\alpha$ and $\hat{\sigma}^2_\alpha$.

Theorem 3.3 *Assume (C1)-(C5). Then, for fixed $\alpha > 0$,*

$$n^{1/2}(\hat{\tau}_\alpha - \tau_\alpha) \xrightarrow{d} N(0, v_\tau) \quad \text{and} \quad n^{1/2}(\hat{\sigma}^2_\alpha - \sigma^2_\alpha) \xrightarrow{d} N(0, v_{\sigma^2}),$$

where $\tau_\alpha = E(Y) - a(\lambda_\alpha)$, $\sigma^2_\alpha = Var(Y) - b(\lambda_\alpha)$, and where v_τ and v_{σ^2} are given in the proof in the Appendix.

We finish this section with the weak consistency of $\hat{\lambda}_\alpha$, $\hat{\tau}_\alpha$ and $\hat{\sigma}^2_\alpha$ when α tends to zero.

Theorem 3.4 *Assume (C1)-(C4) and assume that $\alpha \to 0$ and $(n\alpha)^{-1} = O(1)$. Then, there exists a subsequence $\hat{\lambda}_{\alpha,ss}$ of $\hat{\lambda}_\alpha$ that converges in probability to λ_0. Similarly, there exist subsequences $\hat{\tau}_{\alpha,ss}$ of $\hat{\tau}_\alpha$ and $\hat{\sigma}^2_{\alpha,ss}$ of $\hat{\sigma}^2_\alpha$ that converge in probability to τ_0 and σ^2_0, respectively.*

4 A flexible parametric family of densities on $\mathbb{R}^+$

4.1 Laguerre approximation of densities

We now consider a particular parametric family of densities, that is rich and flexible and that is based on Laguerre polynomials. The idea is to extend the basic exponential density e^{-z} by

using polynomial expansions which are orthogonal to it. These polynomials are the Laguerre polynomials (see e.g. Rice, 1964). Similar ideas have already been used, e.g., by Geerdens et al. (2013) in a different context (frailty models) for one-parameter gamma densities, and by Bertrand et al. (2017) for approximating the density of a variable with compact support based on Bernstein polynomials. The extended-exponential model is fixed by the chosen order m of the polynomial expansions. The densities in this family are given by

$$f_m(z|\theta) = \frac{e^{-z}}{||\theta||^2} \left\{ \sum_{k=0}^{m} \theta_k v_k(z) \right\}^2, \tag{4.1}$$

where $\theta = (\theta_0, \theta_1, \ldots, \theta_m)^t$, $\|\theta\|$ is the Euclidean norm of θ, and v_k are polynomials orthonormal to the exponential density e^{-z}. We fix $\theta_0 = 1$ to fix the scale of θ. Note that if $m = 0$, we have $f_0(z|\theta) = e^{-z}$, and we can check that for any m, $f_m(z|\theta)$ is a density on $\mathbb{R}^+$. Note that in the notation of Section 3, the relation between θ and λ is given by

$$\theta^t = (1,\ \lambda^t).$$

The idea of orthogonal polynomials is as follows. The polynomials have to be orthogonal with respect to the inner product $\langle \varphi_1, \varphi_2 \rangle = \int_0^\infty \varphi_1(z)\varphi_2(z)e^{-z}dz$ and they can be obtained by the Gram-Schmidt orthogonalization procedure. Next, they are divided by their norm to get an orthonormal sequence. As pointed out in Geerdens et al. (2013), the interesting feature is that since $\int_0^\infty e^{cz}e^{-z}dz < \infty$ for any $0 < c < 1$, the system of polynomials $\{v_k : k = 0, 1, \ldots\}$ is closed to continuous functions φ on $\mathbb{R}^+$ that satisfy

$$\int_0^\infty \varphi^2(z)e^{-z}dz < \infty. \tag{4.2}$$

It follows thatby defining $c_k = \langle \varphi, v_k \rangle$,we have that $\lim_{m\to\infty} \int_0^\infty \left[\varphi(z) - \sum_{k=0}^m c_k v_k(z)\right]^2 e^{-z}dz = 0$ for any continuous function φ satisfying (4.2). Now consider any continuous density $f_Z(\cdot)$ of Z on $\mathbb{R}^+$. By defining $\varphi(z) = [f_Z(z)/e^{-z}]^{1/2}$, we see that (4.2) is verified, so with $c_k = \int_0^\infty [f_Z(z)e^{-z}]^{1/2} v_k(z)dz$ we obtain

$$\lim_{m\to\infty} \int_0^\infty \left[\sqrt{f_Z(z)} - \sum_{k=0}^{m} c_k \sqrt{e^{-z}} v_k(z) \right]^2 dz = 0,$$

and hence

$$\lim_{m\to\infty} \int_A \left[\sqrt{f_Z(z)} - \sqrt{f_m(z|\theta)} \right]^2 dz = 0,$$

where A is any set in $\mathbb{R}^+$ for which there exists a $\delta > 0$ such that $f_Z(z) > \delta$ for all $z \in A$, which indicates that the densities in (4.1) parametrized by the appropriate $\theta = (\theta_0, \theta_1, \ldots, \theta_m)^t$ can

potentially approximate any continuous density on $\mathbb{R}^+$ with respect to the Hellinger distance restricted to the set A.[2]

Appendix B explains how to construct Laguerre polynomials and illustrates by means of some simulated examples how flexible the family is and how well continuous functions can be approximated by functions of this family.

Another nice feature of the Laguerre family is that we can easily express the density in (4.1) as a linear combination of gamma densities. We have indeed the following result. The proof is given in Appendix B.

Lemma 4.1 *The density given in* (4.1) *can be written as*

$$f_m(z|\theta) = \sum_{j=0}^{2m} a_j f_\gamma(z|j+1,1), \tag{4.3}$$

where $\sum_{j=0}^{2m} a_j = 1$ *and* $f_\gamma(z|j+1,1) = \dfrac{z^j e^{-z}}{\Gamma(j+1)}$ *are the gamma densities with shape parameter* $j+1$ *and scale equal to 1. The* a_j *are real numbers defined as*

$$a_j = \begin{cases} ||\theta||^{-2}\Gamma(j+1)\sum_{k=0}^{j} c_k c_{j-k} & \text{for } j = 0,\ldots,m \\ ||\theta||^{-2}\Gamma(j+1)\sum_{k=j-m}^{m} c_k c_{j-k} & \text{for } j = m+1,\ldots,2m \end{cases}, \tag{4.4}$$

where the coefficients c_k *are given by*

$$c_k = \sum_{j=k}^{m} \binom{j}{k} \frac{(-1)^{j-k}\theta_j}{\Gamma(k+1)}, \text{ for } k = 0,\ldots,m. \tag{4.5}$$

Note that equation (4.3) does not define a mixture of Gamma densities, but a linear combination with coefficients a_j summing to one but having possibly positive or negative values, as shown by equation (4.5). We refer to Appendix B for the calculation of the coefficients a_j in a simulated example.

Thanks to this result we can provide explicit expressions for the moments and an analytical expression for $\psi_Z(t|\theta)$, its derivative and finally for the function $\eta_Z(t|\theta)$. For the moments we have

$$\mathbb{E}(Z^k|\theta) = \sum_{j=0}^{2m} a_j (j+1)_k, \text{ for } k = 1,2,\ldots, \tag{4.6}$$

[2]As pointed out in Geerdens et al. (2013), with an additional assumption on the function φ, i.e. $\int_0^\infty [\varphi'(z)]^2 z e^{-z} dz < \infty$, the convergence is uniform on any interval $[z_1, z_2] \subset \mathbb{R}^+$ (see Nikiforov and Uvarov, 1988).

where $(j+1)_k$ is the Pochhammer symbol representing $\Gamma(j+1+k)/\Gamma(j+1)$. In the same way we have

$$\psi_Z(t|\theta) = \sum_{j=0}^{2m} a_j \left(\frac{1}{1-it}\right)^{j+1} = \sum_{j=0}^{2m} a_j \left(\frac{1+it}{1+t^2}\right)^{j+1}, \tag{4.7}$$

therefore we can derive

$$\psi'_Z(t|\theta) = \frac{(-2t+i(1-t^2))}{(1+t^2)^2} \sum_{j=0}^{2m} a_j \frac{j+1}{(1-it)^j}. \tag{4.8}$$

Finally we can compute $\eta_Z(t|\theta) = \Im\Big[\psi'_Z(t|\theta)/\psi_Z(t|\theta)\Big]$.

4.2 Simulations

For exploring the behavior of our estimator in finite samples, we will compare its performance with that of the estimator suggested in Kneip et al. (2015) (hereafter denoted by KSVK), where an explicit density (a normal with unknown variance) is used to build the penalized likelihood function (see KSVK for details). For the distribution of Z, KSVK use an histogram based density with fixed number of bins. We will use the same scenarios as in their Examples 1 to 4 and we will compare the results in their Tables 2–7 with ours, where we do not use the information that the noise is normal and we use our approach for approximating the density of Z.

The scenarios can be summarized as follows. We have the model $Y = \tau + Z + \varepsilon$, where $\tau = 0$ and $\varepsilon \sim \mathrm{N}(0, \sigma^2)$ with $\sigma = \rho_{\mathrm{nts}}\sigma_Z$. Various scenarios will be considered for the density of Z, and we take samples of size $n = 50, 100, 500$ and noise-to-signal ratios equal to $\rho_{\mathrm{nts}} = 0, 0.10, 0.25, 0.50$ and 0.75. In Examples 1a and 1b in KSVK, $Z \sim \mathrm{Exp}(\beta)$ with $\mu_Z = \sigma_Z = \beta$ with $\beta = 2$ and $\beta = 1$. In their Examples 2a and 2b, $Z \sim \mathrm{N}^+(\beta_1, \beta_2^2)$ is a truncated normal, where first $\beta_1 = 0$ and $\beta_2 = 0.8$ giving an half-normal distribution with $\mu_Z = 0.6383$ and $\sigma_Z = 0.4822$, whereas in their Example 2b, $\beta_1 = 0.6$ and $\beta_2 = 0.6$ leading to a truncated-normal distribution with $\mu_Z = 0.7726$ and $\sigma_Z = 0.4761$, having its mode far from the frontier point. By analogy with KSVK to select the tuning penalty parameter, we select the values of m (the order of the polynomials) and of α (the penalty parameter) by selecting the pair (m, α) that minimizes the Monte-Carlo sum of the two Root Mean Squared Errors ($RMSE$): $RMSE(\widehat{\tau}_\alpha) + RMSE(\widehat{\sigma}_\alpha)$.

Our results are displayed in Tables 1 to 4. First of all, a quick scan through the tables confirms that the results are going in the expected directions: as the sample size n increases the $RMSE$'s decrease for a fixed level of ρ_{nts}. This is not uniform in all the cases: see,

e.g. Table 3, for $\rho_{\text{nts}} = 0.25$, $RMSE(\widehat{\tau}_\alpha)$ is the same for $n = 50$ and $n = 100$, but this is because the optimal pair (m, α) is selected for optimizing the sum of the two $RMSE$'s (for estimating τ and σ). But in most cases, the behavior is as expected. Second, for a fixed sample size n, the $RMSE$ increases with ρ_{nts} as expected. Finally we observe that in the 4 scenarios, the level of the accuracy is quite comparable through the different scenarios for the density of Z.

Now we will compare the results with the corresponding results from KSVK. We kept the same order for the tables and the same "presentation" to facilitate the comparison. Globally, at least for the 4 preceding cases where the normal assumption is correct, we expect to have less accurate performances, since we only use the symmetry of ε in our method. We summarize in what follows the global picture and the main differences.

Except for the no-noise case ($\rho_{\text{nts}} = 0$), globally our results in terms of the $RMSE$ of the estimators are of the same order as in KSVK. When the size of the noise increases, and when the sample size increases, our results are sometimes better than those obtained in KSVK, although the difference is small. But this is globally good news, because our method does not use the (correct) assumption of normality of ε. The case without noise is much harder to estimate with our approach, which only uses the information on the odd cumulants of order greater than 3. We should also stress that in Table 4, the truncated normal case, where the density of Z is bell-shaped with mode 0.7726 and truncated at zero, our approach based on Laguerre polynomial approximations, works much better than the histogram approach used in KSVK: here the $RMSE$'s are substantially smaller in our approach, in particular for the estimation of τ, in all the cases where $\rho_{\text{nts}} > 0$.

Next, in their Examples 3 and 4, KSVK illustrate how their approach is robust to the normality assumption, by generating samples with $\varepsilon \sim C_1 \times \text{Stud}(4)$, where $C_1 = \sigma/\sqrt{2}$ is a constant to adjust the same noise-to-signal ratios as in the preceding examples.[3] The signal Z is chosen as above as $\text{N}^+(0, 0.8^2)$. The second non-normal specification is $\varepsilon \sim C_2 \times \text{Laplace}$, where again $C_2 = \sigma/\sqrt{2}$ to tune correctly ρ_{nts}. Our results are displayed in Tables 5 and 6.

In the last two examples we may expect better results from our approach since our specification is correct, whereas the one in KSVK is wrong. The global picture coming from the comparisons with Tables 6 and 7 in KSVK is however very similar to the preceding cases: same order of magnitude for the $RMSE$ for the two cases, which confirms that KSVK is rather robust to these two deviations from normality, and that our approach is not quite

[3]Note the unfortunate typo in Kneip et al. (2015), where C_1 is wrongly defined as $\sigma/2$, but this is only a typo, the calculations have been done in KSVK with the correct value of C_1.

well adapted to situations where we have no noise.

5 Estimation of frontiers

We now extend the marginal analysis studied in the previous sections, to the conditional analysis of the boundary of X given a number of explanatory variables. The observable variables are now (Y, W), where the vector $W = (W_1^t, W_2^t)^t$ contains variables describing the level of outputs in a cost analysis (denoted by W_1) and/or environmental variables (denoted by W_2). We are interested in the distribution of X, and in particular in the frontier of X, conditional on $W_1 \geq w_1$ and/or $W_2 = w_2$. In this paper our approach is based on the estimation of τ for given values of w_1 and w_2 and the estimated $\hat{\tau}_\alpha(w_1, w_2)$ is the estimated frontier function. This means that we estimate $\tau(w_1, w_2)$ without any regularity constraint (smoothness, monotonicity, concavity,...) on this frontier function. If the estimator $\hat{\tau}(w_1, w_2)$ does not satisfy the required constraints, $\hat{\tau}(w_1, w_2)$ may in a second step be approximated by a function that verifies the constraints. Another approach not presented in this paper would be to estimate the parameters λ as a function of w_1 and w_2 by imposing constraints. For instance, λ may be estimated under smoothness constraints with respect to w_1 and w_2 in order to estimate the smooth frontier $\tau(w_1, w_2)$.

Let us now give the precise definition of $\hat{\tau}(w_1, w_2)$. Suppose we have a sample (Y_i, W_i), $i = 1, \ldots, n$, of i.i.d. observations having the same distribution as (Y, W). Conditioning on $W_1 \geq w_1$ is done in an elementary way : the sample is restricted to the observations verifying $W_{1j} \geq w_1$ (where the inequality is taken componentwise), and the relevant sample size is no longer n but $n(w_1) = \sum_{j=1}^n I\{W_{1j} \geq w_1\}$, the number of observations for which $W_{1j} \geq w_1$. In absence of the W_2 variable the construction is hence identical to our previous marginal construction and the asymptotic normality of $\hat{\lambda}_\alpha(w_1)$ is derived from the property that

$$\{n(w_1)\}^{1/2}\Big\{\hat{\psi}_Y(t|W_1 \geq w_1) - \psi_Y(t|W_1 \geq w_1)\Big\}$$

converges to a Gaussian process with covariance function $\psi_Y(s+t|W_1 \geq w_1) - \psi_Y(s|W_1 \geq w_1)\psi_Y(t|W_1 \geq w_1)$, where $\hat{\psi}_Y(t|W_1 \geq w_1) = n(w_1)^{-1} \sum_{j=1}^n I\{W_{1j} \geq w_1\} e^{itY_j}$.

Conditioning on $W_2 = w_2$ requires kernel smoothing or any other method for estimating a conditional expectation. Our marginal approach studied in the previous sections was based on the (penalized) minimization of the distance between $\hat{\eta}(\cdot) - \hat{\eta}(0)$ and $\eta_\lambda(\cdot) - \eta_\lambda(0)$. We now replace $\hat{\eta}(\cdot)$ by $\hat{\eta}(\cdot|W_2 = w_2)$ and the solution λ will be a function of w_2. Here, $\hat{\eta}(t|W_2 = w_2)$

equals $\frac{\partial}{\partial t}\Im \log \hat{\psi}_Y(t|W_2 = w_2)$, where $\hat{\psi}_Y(t|W_2 = w_2)$ is given by

$$\hat{\psi}_Y(t|W_2 = w_2) = \frac{1}{n}\sum_{j=1}^{n} \frac{K_h(w_2 - W_{2j})}{\hat{f}_{W_2}(w_2)} e^{itY_j},$$

where $K(u) = k(u_1) \times \cdots \times k(u_{p_2})$ for $u = (u_1, \ldots, u_{p_2})$ with p_2 the dimension of W_2, k is a univariate kernel function, $K_h(u) = h^{-p_2}k(u_1/h) \times \cdots \times k(u_{p_2}/h)$, h is bandwidth sequence, and $\hat{f}_{W_2}(w_2) = n^{-1}\sum_{j=1}^{n} K_h(w_2 - W_{2j})$ is a kernel estimator of the density f_{W_2} of W_2. Under usual regularity conditions, it may be proved that

$$(nh^{p_2})^{1/2}\big(\hat{\psi}_Y(\cdot|W_2 = w_2) - \psi_Y(\cdot|W_2 = w_2)\big)$$

converges weakly in $\mathcal{E}$ to a zero-mean Gaussian process with covariance operator characterized by the function

$$\frac{\int k^2(u)du}{f_{W_2}(w_2)}\big[\psi_Y(s + t|W_2 = w_2) - \psi_Y(s|W_2 = w_2)\psi_Y(t|W_2 = w_2)\big].$$

Apart from these modifications on the rate and on the covariance, the analysis is identical to the non-conditional case.

Finally, the combination of the conditions $W_1 \geq w_1$ and $W_2 = w_2$ is obvious : it suffices to replace n by $n(w_1)$, $f_{W_2}(w_2)$ by $f_{W_2}(w_2|W_1 \geq w_1)$ and the conditional characteristic function is now $\psi_Y(t|W_1 \geq w_1, W_2 = w_2)$.

Based on this result, the asymptotic normality of the estimators $\hat{\lambda}(w_1, w_2)$ and $\hat{\tau}(w_1, w_2)$ can now be derived in the same way as it has been done in Theorems 3.2 and 3.3 in the univariate case. In particular, the rates of convergence are the same as for the estimator $\hat{\psi}_Y$.

We conclude this section by some Monte-Carlo experiments to illustrate how the methods work and for allowing the comparison with the Kneip et al. (2015) (KSVK) approach. We thus follow the KSVK scenario and their strategy for the estimation of the frontier function. The only difference is that they use the assumption of normality of the noise in their approach, whereas we only use the information that the noise is symmetric, as developed above. The Monte-Carlo scenario in KSVK (which was already used in Hall and Simar, 2002) can be described as follows. We want to estimate a production frontier (instead of a cost frontier), giving the maximum achievable level of the output Y when using the amount of input W. We use the maximum instead of the minimum, in order to be coherent with what was done in KSVK. This has however no influence on the methodology. The stochastic frontier model can thus be written as

$$Y = \tau(W)\exp(-Z)\exp(\varepsilon), \tag{5.1}$$

where we only know that $Z \geq 0$ and ε is symmetric around zero. As in KSVK, we simulate n data points from the following model: $\tau(w) = w^{1/2}$, $Z \sim \text{Exp}(3)$ and $\varepsilon \sim \text{N}(0, 0.0667^2)$, with $W \sim \text{U}(0,1)$, all the variables W, Z and ε are independent in this scenario. Note that $\rho_{\text{nts}} = 0.20$. As in KSVK, we analyze the estimation of $\tau(w)$ for 3 values of $w = (0.25, 0.50, 0.75)$ and we select a neighborhood for each w, by using the same bandwidth h_w (given by the usual rule of thumb). Note that taking the logarithms we have

$$\log Y = \log \tau(W) - Z + \varepsilon, \tag{5.2}$$

which has exactly the additive form of our model. As in KSVK, we do not use directly the observations $\log Y_i$ for which $|W_i - w| \leq h_w$ but we rather use a "local linear" approximation of $\tau(\cdot)$ around the point of interest w. To be more precise, the idea is as follows : if $\tau(\cdot)$ is sufficiently smooth we have from a Taylor expansion that $\log \tau(W)$ is approximately $\log \tau(w) + \beta^t(w)(W - w)$ and for small h_w we then obtain

$$\log Y_i \approx \alpha(w) + \beta^t(w)(W_i - w) - \widetilde{Z}_i(w) + \varepsilon_i, \text{ if } |W_i - w| \leq h_w, \tag{5.3}$$

where $\alpha(w) = \log \tau(w) - E(Z|W = w)$ and $\widetilde{Z}_i(w) = Z_i - E(Z|W = w)$. Since in the last equation the error term has locally zero mean, $\alpha(w)$ and $\beta(w)$ can be estimated by ordinary least squares. Now we have for all W_i with $|W_i - w| \leq h_w$:

$$\begin{aligned} \log \widetilde{Y}_i &:= \log Y_i - \beta^t(w)(W_i - w) \\ &\approx \log \tau(w) - Z_i + \varepsilon_i, \end{aligned}$$

which suggests to estimate the frontier function at $W = w$ by the upper boundary of the local linear least squares estimator of $\log \widetilde{Y}_i$ for $|W_i - w| \leq h_w$ rather than the upper boundary of the local observations $\log Y_i$ themselves (see the practical details in Kneip et al., 2015).

We used 200 Monte-Carlo replications to evaluate both the bias and the MSE, and compared the results with those obtained in KSVK for the case $n = 100$ and $n = 500$. The results are shown in Table 7. The table deserves some comments. First we see that the results are going in the expected direction: the MSE for both $\tau(w)$ and $\sigma(w)$ decreases when n increases. Comparing with the results in KSVK, we see that for $n = 100$ the results are, as expected, less accurate in our approach, since we only use the symmetry assumption for the distribution of the noise. Note also that we have very few observations for each value of w (on average only 24 data points for $\log \widetilde{Y}_i$). As a result, and comparing with Table 8 in KSVK, we see that the MSE for the estimation of $\tau(w)$ has an order augmented by a factor 2. For the estimation of $\sigma(w)$ the situation is worse and we have a factor 10. The

situation is much better when the sample size increases to $n = 500$. Here for the estimation of $\tau(w)$ we reach for the MSE the same order as the values in KSVK. For $\sigma(w)$ we still have disappointing results with the MSE greater by a factor 7. In this case ($n = 500$), note that for each value of w, we only use on average 88 data points, which is not much for a procedure that does not exploit the normality of the noise, as opposed to KSVK. This is the price to pay for being more general. [4]

Only for illustration, we show also in Table 8 the results obtained by conditioning not on $W = w$ but rather on $W \leq w$ as described above (we are in a production setup here so we look for the upper support of Y given that $W \leq w$). In this case there is no local linear smoothing step. To summarize, we see that in this scenario the results are similar for the estimation of $\tau(w)$, but they are still worse for the estimation of $\sigma(w)$, where too many Monte-Carlo estimates were set to zero. For the chosen scenario, the first approach seems more appropriate.

To end this section, we illustrate our procedure with a real data set coming from 2326 post offices (delivery offices) of a French operator, observed in 2013. For each post office we have the level of the activity, the output W (the volume of delivered mail), and the labor cost measured by the quantity of labor (number of hours) X. This data set has been used in a different context (deterministic frontier) in Cazals et al. (2016).

We are here interested in the cost frontier, conditional on the value $W = w$, from the sample of $n = 2326$ observations (W_i, Y_i), where $i = 1, \ldots, n$ and Y_i are the noisy versions of X_i. Due to the shape of the cloud of points (quite asymmetric with heavy right tails for both dimensions) we decided to work on a logarithmic scale for both input and output. The localization in the $\log W$ is done by selecting a bandwidth $h_w = 0.1944$ which was obtained by least-squares cross validation for estimation of the conditional distribution of $\log Y$ given $\log W$ (see Li et al, 2013). As explained above for estimating the cost frontier around $\log w$, we use a local linear approximation, as suggested in Kneip et al. (2015). In particular, we do not use directly the observations $\log Y_i$ for which $|\log W_i - \log w| \leq h_w$, but their local linear approximations $\log \widetilde{Y}_i = \log Y_i - \beta^t(w)(\log W_i - \log w)$ where $\beta(w)$ is the local slope obtained by local ordinary least squares approximation of $\log Y_i$ over $(\log W_i - \log w)$ (see details above).

The estimates are computed for a fixed grid of 21 values of $\log w$. The values for m

[4] We might expect better performance of our approach compared to KSVK when the noise is not normal. Unfortunately we know from the results in the univariate case, that for the Student-t and the Laplace cases we need a lot of observations to improve the KSVK approach, see the comments above on Tables 5 and 6. Here for each value of w, we only have roughly 24 observations when $n = 100$ and 88 when $n = 500$.

are selected over a grid of values $m = 0, \ldots, 5$ and for the penalty term over a grid of 21 equidistant values of $\alpha \in [0, 3]$. The optimal values of the pairs (m, α) are obtained, at each selected value of $\log w$, by minimizing the bootstrap estimate of the RMSE of the frontier estimates (500 bootstrap replications at each point).

The results are displayed in Figure 1 and the 21 values of $\log w$ are shown by the blue circles. The final estimate of the cost frontier is obtained by smoothing these pointwise estimates (we used a standard B-splines smoothing technique, cubic splines, 20 knots, penalty on second derivatives with penalty factor 10, see Eilers and Marx, 1996). We can see how the cost frontier "envelops" the cloud of points by letting as expected a few points outside the frontier. In particular a few points are really extreme (e.g. around the coordinate (11.7,5)) and our model is quite robust to these outliers. In the logarithmic units that we have chosen, the magnitude of $\widehat{\sigma}_\varepsilon(w)$ is stable over the 21 pointwise estimates and is of the order of 10^{-4}.

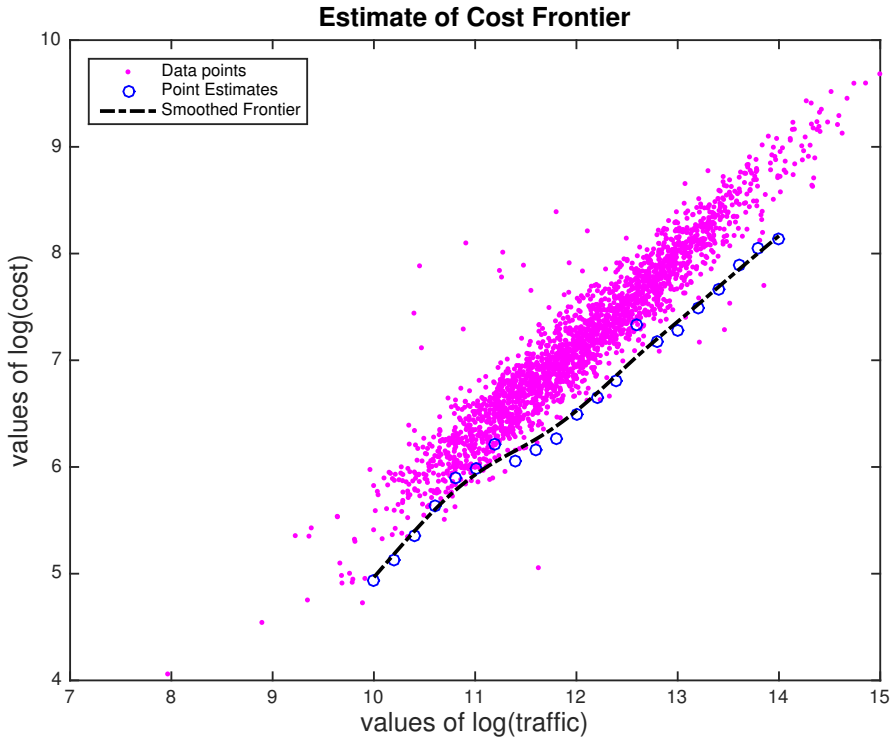

Figure 1: *Example with data from post offices, the curve is a smoothed version of the pointwise estimates (blue circles). The* $n = 2326$ *data points are also displayed.*

6 Conclusion and further research

This paper presents a new nonparametric analysis of the model $Y = \tau + Z + \varepsilon$ (with $Z \geq 0$ and $\varepsilon \in \mathbb{R}$), where ε is only constrained to be symmetric, the distribution of Z is supposed to belong to a flexible parametric class, and Z and ε are independent. We also consider the extension to stochastic frontier models. We provide an estimation approach based on the cumulant functions of odd order and we develop asymptotic properties of our estimator. A tractable family for the distribution of the inefficiency Z is proposed, and the power of our method is shown via simulations.

In this paper we estimate the true frontier, i.e. the minimal value of the cost. It is well known that this estimator is not robust, in particular to outliers. We may argue that the introduction of a stochastic error improves the robustness of the estimation of the frontier, but it is possible to adapt the technology of this paper to a robust estimator. From the information contained in the imaginary part of the characteristic function we may estimate the survival function $S(x|w_1, w_2) = P(X \geq x|W_1 \geq w_1, W_2 = w_2)$ of X given $W_1 \geq w_1$ and $W_2 = w_2$, which leads to an estimator of the quantile $\tau_\alpha(w_1, w_2) = S^{-1}(1-\alpha|w_1, w_2)$, or the m-frontier $\tau_m(w_1, w_2) = \int_0^\infty S^m(u|w_1, w_2)\, du$, see e.g. Daouia and Simar (2007) and Cazals et al. (2002). A deeper analysis of the relation between stochastic frontiers and robustness will be the topic of future research.

Several extensions of deterministic frontiers may be considered in the case of stochastic frontiers with symmetric errors, like endogeneity, separability, among others.

Finally, note that our model is an example of a more general situation. We want to estimate a function or a set of functions which are non identified (different functions are compatible with the data generating process). However, there exists a sieve family of parametric approximations which satisfies the two following properties. First for any finite dimensional parametric approximation the parameters are identified and, second, there exists a sequence of parametric approximations which converges (for a suitable topology) to any function. Intuitively, if the dimension of the parametric model increases the approximation is more accurate but the model is less identified and the choice of the parametric dimension follows from a trade-off between these two arguments. In our example the lack of identification requires the introduction of the penalty term which introduces an estimation bias. Two tuning parameters should be determined: the size of the approximation and the penalty parameter. In our paper we assume that the true function belongs to a finite parametric family and we do not develop the general theory.

References

[1] Aarts, L.P., Groeneboom, P. and G. Jongbloed (2007). Estimating the upper support point in deconvolution. *Scandinavian Journal of Statistics*, 34, 552–568.

[2] Adler, R.J. (1981). *The Geometry of Random Fields*. Wiley, Chichester.

[3] Bertrand, A., Van Keilegom, I. and C. Legrand (2017). Flexible parametric approach to classical measurement error variance estimation without auxiliary data. *Biometrics* (in press).

[4] Butucea, C. and C. Matias (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11, 309–340.

[5] Butucea, C., Matias, C. and C. Pouet (2008). Adaptivity in convolution models with partially known noise distribution. *Electronic Journal of Statistics*, 2, 897–915.

[6] Butucea, C., Nguyepe Zumpe, R. and P. Vandekerkhove (2017). Semiparametric topographical mixture models with symmetric errors. *Bernoulli*, 23, 825–862.

[7] Butucea, C. and P. Vandekerkhove (2014). Semiparametric mixtures of symmetric distributions. *Scandinavian Journal of Statistics*, 41, 227–239.

[8] Cazals, C., Fève F., Florens, J.P. and L. Simar (2016). Nonparametric instrumental variables estimation for efficiency frontier. *Journal of Econometrics*, 190, 349–359.

[9] Cazals, C. Florens, J.P. and L. Simar (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106, 1–25.

[10] Daouia, A., Florens, J.-P. and L. Simar (2018). Robust frontier estimation from noisy data: a Tikhonov regularization approach. *Econometrics and Statistics* (in press), https://doi.org/10.1016/j.ecosta.2018.07.003.

[11] Daouia, A. and L. Simar (2007). Nonparametric efficiency analysis: a multivariate conditional quantile approach. *Journal of Econometrics*, 140, 375–400.

[12] Delaigle, A. and I. Gijbels (2006). Data-driven boundary estimation in deconvolution problems. *Computational Statistics and Data Analysis*, 50, 1965–1994.

[13] Delaigle, A. and P. Hall (2016). Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society - Series B*, 78, 231–252.

[14] Eilers, P.H.C. and B.D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121.

[15] Engl, H.W., Hanke, M. and A. Neubauer (1996). *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht.

[16] Fan, J., Li, Q. and A. Weersink (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business & Economic Statistics*, 14, 460–468.

[17] Geerdens, C., Claeskens, G. and P. Janssen (2013). Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty. *Biostatistics*, 14, 433–446.

[18] Goldenshluger, A. and A. Tsybakov (2004). Estimating the endpoint of a distribution in presence of additive observation errors. *Statistics & Probability Letters*, 68, 39–49.

[19] Hall, P. and L. Simar (2002). Estimating a changepoint, boundary or frontier in the presence of observation error. *Journal of the American Statistical Association*, 97, 523–534.

[20] Kneip, A., Simar, L. and I. Van Keilegom (2015). Frontier estimation in the presence of measurement error with unknown variance. *Journal of Econometrics*, 184, 379–393.

[21] Kuosmanen, T. and M. Kortelainen (2012). Stochastic non-smooth envelopment of data: Semiparametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38, 11–28.

[22] Li, Q., Lin, J. and J.S. Racine (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics*, 31, 57–65.

[23] Lukacs, E. (1960). *Characteristic Functions*. London: Griffin.

[24] Lukacs, E. (1983). *Developments in Characteristic Function Theory*. New York: Macmillan.

[25] Meister, A. (2006a). Support estimation via moment estimation in presence of noise. *Statistics*, 40, 259–275.

[26] Meister, A. (2006b). Density estimation with normal measurement error with unknown variance. *Statistica Sinica*, 16, 195–211.

[27] Meister, A. (2007). Deconvolving compactly supported densities. *Math. Methods Statist.*, 16, 63–76.

[28] Nikiforov, A. and V.B. Uvarov (1988). *Special Functions of Mathematical Physics*. Basel: Birkhäuser Verlag. A unified introduction with applications, translated from the Russian and with a preface by Ralph P. Boas, with a foreword by A. A. Samarskiĭ.

[29] Rice, J.J. (1964). *The Approximation of Functions*. Reading : Addison - Wesley.

[30] Schwarz, M. and S. Van Bellegem (2010). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters*, 80, 236–241.

[31] Van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

[32] Van der Vaart, A.W. and J.A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.

A Appendix A: Proofs of the main results

We start by stating the regularity conditions under which our main asymptotic results are valid.

(C1) $T_\lambda = \eta_\lambda - \eta_\lambda(0)$ is an operator from Λ to the space $\mathcal{E}$, T_λ is injective and Lipschitz continuous, and $\sup_{\lambda\in\Lambda}\int |T_\lambda|^2 w < \infty$.

(C2) T_λ is Fréchet differentiable and the Fréchet derivative T'_λ is a linear, injective and bounded operator from $\mathbb{R}^m$ to $\mathcal{E}$ for $\lambda \in \Lambda$. The adjoint T'^*_λ of T'_λ satisfies $\|T'^*_\lambda - T'^*_{\lambda_0}\| \leq \gamma\|\lambda - \lambda_0\|$, and $T'^*_\lambda g$ ($g \in \mathcal{E}$) is continuously differentiable for λ in the interior of Λ.

(C3) There exists a λ_α in the interior of Λ that is the unique solution of $\min_{\lambda\in\Lambda}\left\{\|T_{\lambda_0} - T_\lambda\|^2 + \alpha\|\lambda_\alpha\|^2\right\}$, for all $\alpha > 0$.

(C4) The matrix Δ (defined in Theorem 3.2) is positive definite.

(C5) The function

$$u : \Lambda \to \mathbb{R}^2 : \lambda \to \begin{pmatrix} a(\lambda) \\ b(\lambda) \end{pmatrix}$$

is continuously differentiable.

Proof of Theorem 3.1. It is easily seen that in the space $\mathcal{E}$ the process $n^{1/2}(\hat{\psi}_Y(\cdot) - \psi_Y(\cdot))$ converges weakly to a zero-mean Gaussian process with covariance function $(\beta(s,t))_{s,t}$ given by $\beta(s,t) = \psi_Y(s+t) - \psi_Y(s)\psi_Y(t)$ (see e.g. Van der Vaart and Wellner, 1996, Chapter 1.8 for weak convergence in Hilbert spaces). Next, using the Delta-method it follows that the process

$$n^{1/2}\big(\Im \log \hat{\psi}_Y(\cdot) - \Im \log \psi_Y(\cdot)\big)$$

converges weakly to a zero-mean Gaussian process, whose covariance function is given by

$$\frac{1}{4}\Big\{\frac{\psi_Y(t-s)}{\psi_Y(t)\overline{\psi_Y(s)}} + \frac{\overline{\psi_Y(t-s)}}{\overline{\psi_Y(t)}\psi_Y(s)} - \frac{\psi_Y(t+s)}{\psi_Y(t)\psi_Y(s)} - \frac{\overline{\psi_Y(t+s)}}{\overline{\psi_Y(t)}\,\overline{\psi_Y(s)}}\Big\},$$

since $\Im \log \hat{\psi}_Y(t) = (2i)^{-1}(\log \hat{\psi}_Y(t) - \log \overline{\hat{\psi}}_Y(t))$. Since the covariance function of this process is twice continuously differentiable, it follows from Theorem 2.2.2 in Adler (1981) that

$$n^{1/2}\big(\hat{T} - T_{\lambda_0}\big) = n^{1/2}\frac{\partial}{\partial t}\Big(\Im \log \hat{\psi}_Y(\cdot) - \Im \log \hat{\psi}_Y(0) - \Im \log \psi_Y(\cdot) + \Im \log \psi_Y(0)\Big)$$

converges also weakly to a zero-mean Gaussian process whose covariance operator is given in the statement of the theorem. □

Proof of Theorem 3.2. For the consistency note that we need to show that (see e.g. Van der Vaart (1998), Theorem 5.7, p. 45)

$$\sup_{\lambda \in \Lambda} \Big| \int \Big\{ \big[(\hat{T} - T_\lambda)^2 w + \alpha \|\lambda\|^2\big] - \big[(T_{\lambda_0} - T_\lambda)^2 w + \alpha \|\lambda\|^2\big] \Big\} \Big| \xrightarrow{P} 0.$$

The integral in the above expression is bounded in absolute value by

$$\begin{aligned} &\Big| \int \hat{T}^2 w - \int T_{\lambda_0}^2 w \Big| + 2 \sup_\lambda \Big| \int (\hat{T} - T_{\lambda_0}) T_\lambda w \Big| \\ &\leq \Big| \int \hat{T}^2 w - \int T_{\lambda_0}^2 w \Big| + 2 \Big(\int (\hat{T} - T_{\lambda_0})^2 w \Big)^{1/2} \sup_\lambda \Big(\int |T_\lambda|^2 w \Big)^{1/2}, \end{aligned}$$

and this tends to zero in probability.

Next we prove the asymptotic normality of $\hat{\lambda}_\alpha$. It follows from (3.2) and (3.3) that

$$\alpha(\hat{\lambda}_\alpha - \lambda_\alpha) + T'^*_{\lambda_\alpha}(T_{\hat{\lambda}_\alpha} - T_{\lambda_\alpha}) - (T'^*_{\hat{\lambda}_\alpha} - T'^*_{\lambda_\alpha})(\hat{T} - T_{\hat{\lambda}_\alpha}) = T'^*_{\lambda_\alpha}(\hat{T} - T_{\lambda_0}).$$

Note that

$$\begin{aligned} &(T'^*_{\hat{\lambda}_\alpha} - T'^*_{\lambda_\alpha})(\hat{T} - T_{\hat{\lambda}_\alpha}) \\ &= (T'^*_{\hat{\lambda}_\alpha} - T'^*_{\lambda_\alpha})\big\{(\hat{T} - T_{\lambda_0}) + (T_{\lambda_0} - T_{\lambda_\alpha}) + (T_{\lambda_\alpha} - T_{\hat{\lambda}_\alpha})\big\} \\ &= -\nabla(T'^*_\lambda S)|_{S = T_{\lambda_\alpha} - T_{\lambda_0}, \lambda = \lambda_\alpha}(\hat{\lambda}_\alpha - \lambda_\alpha) + o_P(\|\hat{\lambda}_\alpha - \lambda_\alpha\|), \end{aligned}$$

since $\|T'^*_{\hat{\lambda}_\alpha} - T'^*_{\lambda_\alpha}\| \leq \gamma \|\hat{\lambda}_\alpha - \lambda_\alpha\|$ for some $\gamma < \infty$, and similarly for T_{λ_α}. Hence,

$$\hat{\lambda}_\alpha - \lambda_\alpha = \Big(\alpha I + T'^*_{\lambda_\alpha} T'_{\lambda_\alpha} + \nabla(T'^*_\lambda S)|_{S = T_{\lambda_\alpha} - T_{\lambda_0}, \lambda = \lambda_\alpha}\Big)^{-1} T'^*_{\lambda_\alpha}(\hat{T} - T_{\lambda_0}) + o_P(\|\hat{\lambda}_\alpha - \lambda_\alpha\|),$$

from which the asymptotic normality follows. □

Proof of Theorem 3.3. By following similar arguments as in the proof of Theorem 3.1, we can show that the process

$$n^{1/2} \begin{pmatrix} \overline{Y} - E(Y) \\ n^{-1} \sum_{j=1}^n Y_j^2 - E(Y^2) \\ \hat{\eta}_Y(\cdot) - \eta_Y(\cdot) \end{pmatrix} = n^{1/2} \begin{pmatrix} -i\big(\frac{\partial}{\partial t} \hat{\psi}_Y(t)|_{t=0} - E(Y)\big) \\ -\big(\frac{\partial^2}{\partial t^2} \hat{\psi}_Y(t)|_{t=0} - E(Y^2)\big) \\ \frac{\partial}{\partial t} \Im \log \hat{\psi}_Y(\cdot) - \eta_Y(\cdot) \end{pmatrix}$$

converges weakly to a zero-mean Gaussian process with variance operator Ξ characterized for fixed s and t by

$$\begin{pmatrix} -\frac{\partial^2}{\partial s\partial t}\beta(s,t)|_{s,t=0} & i\frac{\partial^3}{\partial s\partial t^2}\beta(s,t)|_{s,t=0} & -\frac{1}{2}\frac{\partial^2}{\partial s\partial t}\left[\frac{\beta(s,t)}{\psi_Y(t)}-\frac{\beta(s,-t)}{\psi_Y(-t)}\right]\Big|_{s=0} \\ i\frac{\partial^3}{\partial s^2\partial t}\beta(s,t)|_{s,t=0} & \frac{\partial^4}{\partial s^2\partial t^2}\beta(s,t)|_{s,t=0} & -\frac{1}{2i}\frac{\partial^3}{\partial s^2\partial t}\left[\frac{\beta(s,t)}{\psi_Y(t)}-\frac{\beta(s,-t)}{\psi_Y(-t)}\right]\Big|_{s=0} \\ -\frac{1}{2}\frac{\partial^2}{\partial s\partial t}\left[\frac{\beta(s,t)}{\psi_Y(s)}-\frac{\beta(-s,t)}{\psi_Y(-s)}\right]\Big|_{t=0} & -\frac{1}{2i}\frac{\partial^3}{\partial s\partial t^2}\left[\frac{\beta(s,t)}{\psi_Y(s)}-\frac{\beta(-s,t)}{\psi_Y(-s)}\right]\Big|_{t=0} & \sigma(s,t) \end{pmatrix},$$

where $\beta(s,t)=\psi_Y(s+t)-\psi_Y(s)\psi_Y(t)$, and where $\sigma(s,t)$ is defined in Theorem 3.1. Let us rewrite this matrix as

$$\begin{pmatrix} a_{11} & a_{12} & c_1(t) \\ a_{21} & a_{22} & c_2(t) \\ c_1(s) & c_2(s) & \sigma(s,t) \end{pmatrix}.$$

Use the notation

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

let Σ be the operator characterized by $\sigma(s,t)$, and let C be the operator from $\mathcal{E}$ to $\mathbb{R}^2$ given by

$$Cg=\begin{pmatrix} \langle c_1, g\rangle_{\mathcal{E}} \\ \langle c_2, g\rangle_{\mathcal{E}} \end{pmatrix}.$$

Then, Ξ may be partitioned into

$$\begin{pmatrix} A & C \\ C^* & \Sigma \end{pmatrix},$$

where C^* is the adjoint operator of C. Define the 2×2 matrix

$$M=\begin{pmatrix} 1 & -2E(Y) \\ 0 & 1 \end{pmatrix},$$

and let Υ be the linear operator from $\mathcal{E}$ to $\mathbb{R}^m$ that is such that the asymptotic variance of $\hat{\lambda}_\alpha$ for fixed α is given by $\Upsilon\Sigma\Upsilon^*$ (this is possible thanks to Theorem 3.2). Also, define the function $u:\mathbb{R}^m\to\mathbb{R}^2$ such that

$$u(\lambda)=\begin{pmatrix} a(\lambda) \\ b(\lambda) \end{pmatrix},$$

and let $\frac{\partial u}{\partial\lambda}$ be the $2\times m$ matrix of partial derivatives. Then,

$$n^{1/2}\begin{pmatrix} \overline{Y}-E(Y) \\ \hat{\sigma}_Y^2-[E(Y^2)-E(Y)^2] \\ u(\hat{\lambda}_\alpha)-u(\lambda) \end{pmatrix} \xrightarrow{d} N\left(0,\begin{pmatrix} MAM^t & MC\Upsilon^*\frac{\partial u}{\partial\lambda}^t \\ \frac{\partial u}{\partial\lambda}\Upsilon C^*M^t & \frac{\partial u}{\partial\lambda}\Upsilon\Sigma\Upsilon^*\frac{\partial u}{\partial\lambda}^t \end{pmatrix}\right).$$

The latter variance matrix is a 4×4 matrix which we denote by

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

It now follows that

$$n^{1/2} \begin{pmatrix} \hat{\tau}_\alpha - \tau_\alpha \\ \hat{\sigma}^2_\alpha - \sigma^2_\alpha \end{pmatrix} \xrightarrow{d} N\big(0, B_{11} + B_{22} - B_{12} - B_{21}\big).$$

□

Proof of Theorem 3.4. We follow the method of proof given in Engl et al (1996). First note that by definition of $\hat{\lambda}_\alpha$,

$$\|\hat{T} - T_{\hat{\lambda}_\alpha}\|^2 \leq \|\hat{T} - T_{\hat{\lambda}_\alpha}\|^2 + \alpha\|\hat{\lambda}_\alpha\|^2 \leq \|\hat{T} - T_{\lambda_0}\|^2 + \alpha\|\lambda_0\|^2, \tag{A.1}$$

and this converges to zero in probability if $\alpha \to 0$, since $\|\hat{T} - T_{\lambda_0}\|^2 = O_P(n^{-1})$ thanks to Theorem 3.1. Hence, $T_{\hat{\lambda}_\alpha} \xrightarrow{P} T_{\lambda_0}$, since

$$\|T_{\hat{\lambda}_\alpha} - T_{\lambda_0}\| \leq \|T_{\hat{\lambda}_\alpha} - \hat{T}\| + \|\hat{T} - T_{\lambda_0}\| \xrightarrow{P} 0.$$

Moreover, $\|\hat{\lambda}_\alpha\| = O_P(1)$, since by (A.1), $\|\hat{\lambda}_\alpha\| = O_P((n\alpha)^{-1}) + \|\lambda_0\|^2 = O_P((n\alpha)^{-1})$. It now follows that there exists a subsequence $\hat{\lambda}_{\alpha,ss}$ of $\hat{\lambda}_\alpha$ that converges in probability to a certain limit λ. Hence, $T_{\hat{\lambda}_{\alpha,ss}} \xrightarrow{P} T_\lambda$. We also know that $T_{\hat{\lambda}_{\alpha,ss}} \xrightarrow{P} T_{\lambda_0}$. Hence, by the injectivity of T, $\lambda = \lambda_0$. Since $\hat{\tau}_\alpha$ and $\hat{\sigma}^2_\alpha$ are functions of $\hat{\lambda}_\alpha$, $\overline{Y}$ and $\hat{\sigma}^2_Y$, the same holds true for $\hat{\tau}_\alpha$ and $\hat{\sigma}^2_\alpha$. □

B Appendix B: Some properties of Laguerre polynomials

The Laguerre polynomials are easy to build. For $k = 0, 1, \ldots$, they can be written as

$$v_k(z) = \sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^{k-j}}{j!} z^j, \tag{B.1}$$

and it is easy to prove that

$$\int_0^\infty v_{k_1}(z) v_{k_2}(z) e^{-z} dz = \begin{cases} 0 & \text{if } k_1 \neq k_2 \\ 1 & \text{if } k_1 = k_2. \end{cases} \tag{B.2}$$

In Figure 2 we illustrate how flexible the family is with only $m = 3$. We display 5 densities randomly selected from this family. Here, the parameters θ_1, θ_2 and θ_3 are randomly

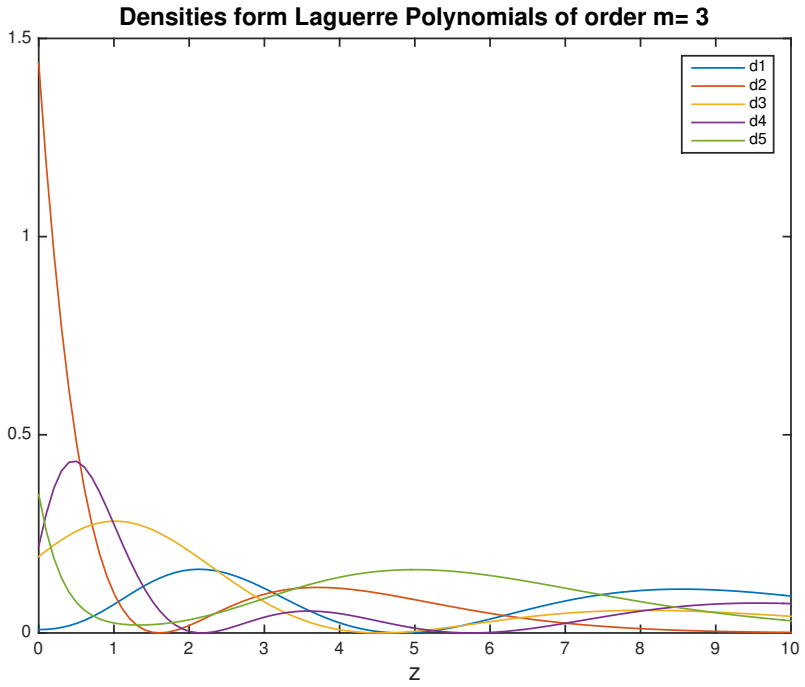

Figure 2: *The 5 densities* $d1$ *to* $d5$ *are randomly selected from the parametric family with* $m = 3$. *The parameters* θ_1, θ_2 *and* θ_3 *are drawn from independent* $N(0, 1)$ *variables.*

generated from independent $N(0, 1)$ variables. We see how even with only $m = 3$ we cover a wide variety of shapes of the resulting densities. The coefficients $a_1, \ldots, a_7$ are given in Table 9 for each of the 5 densities, and are derived from formulas (4.3)-(4.5).

In Figures 3 and 4 we show how usual parametric densities used for positive random variables are easily approximated by a density in our family (4.1). We consider the half-normal and the truncated-normal. We see that even with low order for m we have already a good 'visual' approximation.

To conclude we give the proof of Lemma 4.1.

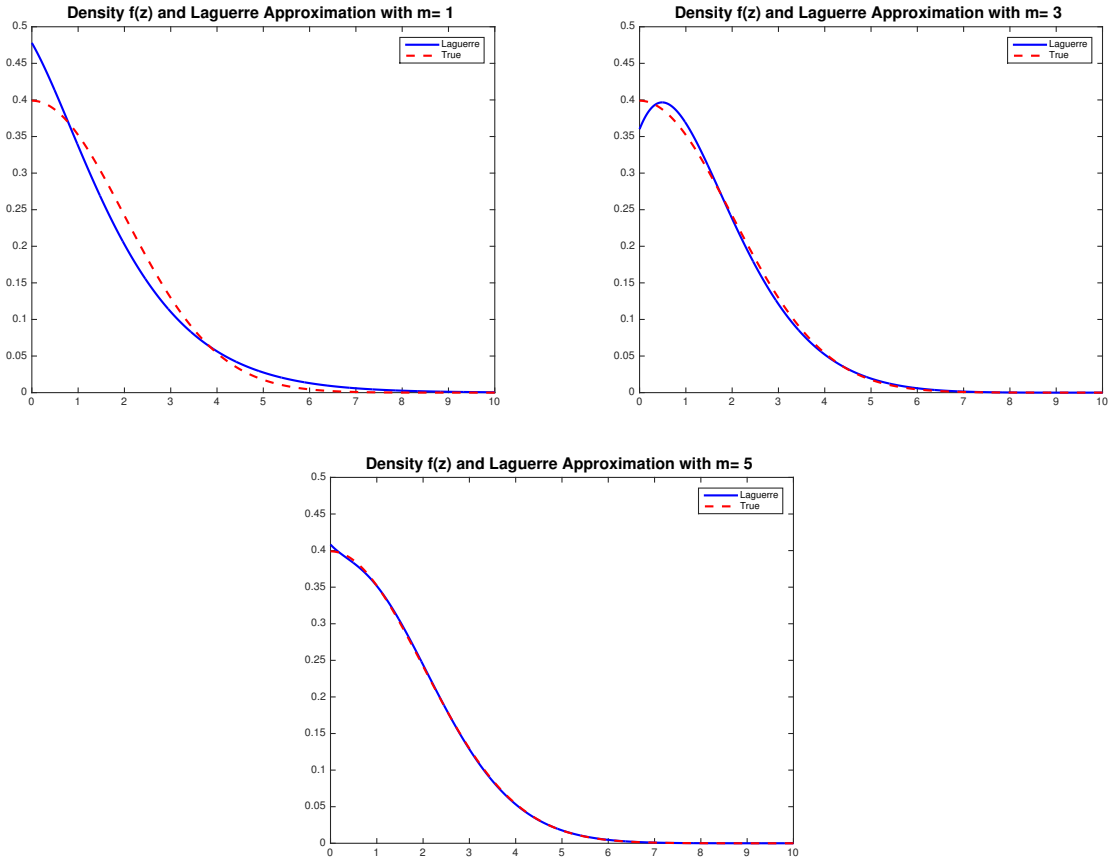

Figure 3: *Plots of the true half-normal density, $f_Z(z)$ (dashed-red), together with the best parametric approximation $f_m(z|\theta)$ (solid blue) for $m = 1, 3, 5$ (respectively).*

Proof of Lemma 4.1. Write

$$\begin{aligned} f_m(z|\theta) &= \frac{e^{-z}}{\|\theta\|^2}\Big\{\sum_{k=0}^{m}\theta_k v_k(z)\Big\}^2 \\ &= \frac{e^{-z}}{\|\theta\|^2}\Big\{\sum_{k=0}^{m}\theta_k\sum_{j=0}^{k}\binom{k}{j}\frac{(-1)^{k-j}}{j!}z^j\Big\}^2 \\ &= \frac{e^{-z}}{\|\theta\|^2}\Big\{\sum_{j=0}^{m}\sum_{k=j}^{m}\theta_k\binom{k}{j}\frac{(-1)^{k-j}}{j!}z^j\Big\}^2. \end{aligned}$$

The expression between braces can be written as $\sum_{j=0}^{m} c_j z^j$ with, for $j = 0, \ldots, m$,

$$c_j = \sum_{k=j}^{m}\theta_k\binom{k}{j}\frac{(-1)^{k-j}}{j!}.$$

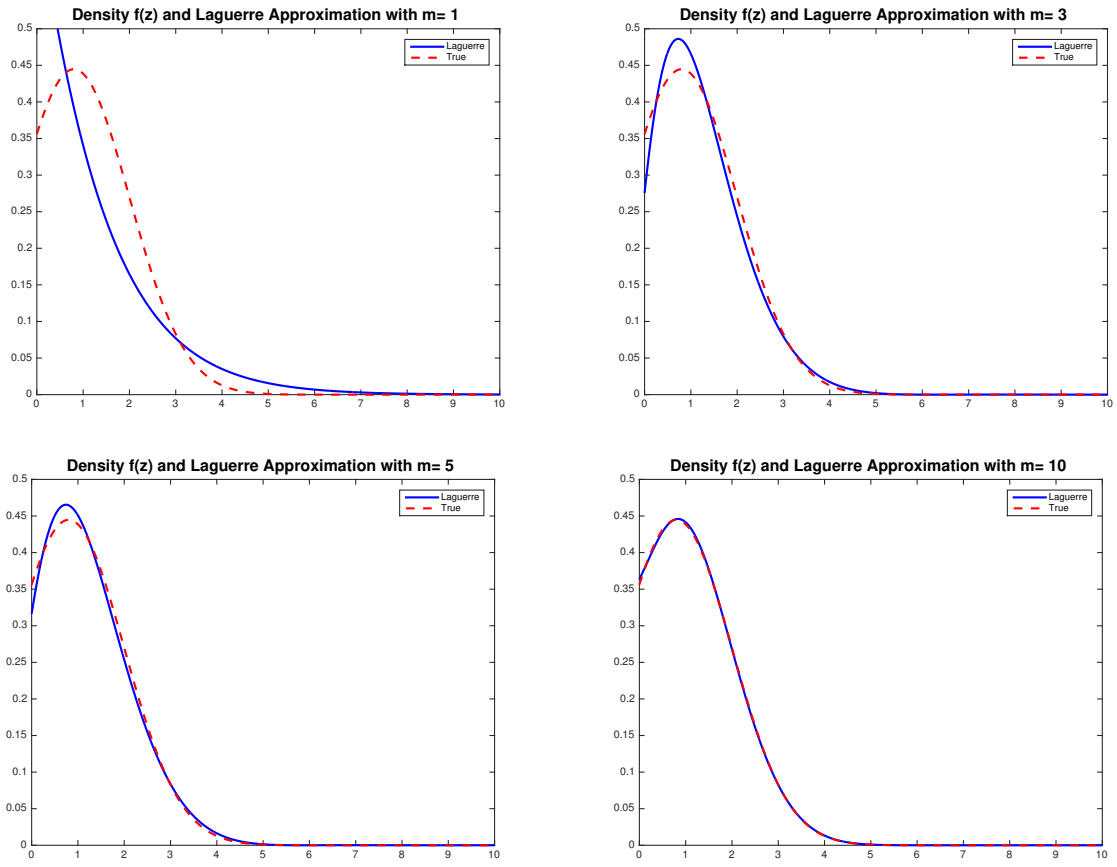

Figure 4: *Plots of the true truncated-normal density, $f_Z(z)$ (dashed-red), together with the best parametric approximation $f_m(z|\theta)$ (solid blue) for $m = 1, 3, 5, 10$ (respectively).*

Moreover,

$$\begin{aligned}\Big(\sum_{j=0}^{m} c_j z^j\Big)^2 &= \sum_{k=0}^{m}\sum_{j=0}^{m} c_k c_j z^{k+j} = \sum_{k=0}^{m}\sum_{j=k}^{m+k} c_k c_{j-k} z^j \\ &= \sum_{j=0}^{m}\Big(\sum_{k=0}^{j} c_k c_{j-k}\Big) z^j + \sum_{j=m+1}^{2m}\Big(\sum_{k=j-m}^{m} c_k c_{j-k}\Big) z^j = \sum_{j=0}^{2m} b_j z^j,\end{aligned}$$

where

$$b_j = \begin{cases} \sum_{k=0}^{j} c_k c_{j-k} & \text{if } j = 0, 1, \ldots, m \\ \sum_{k=j-m}^{m} c_k c_{j-k} & \text{if } j = m+1, m+2, \ldots, 2m. \end{cases}$$

Hence,

$$f_m(z|\theta) = \frac{e^{-z}}{\|\theta\|^2}\sum_{j=0}^{2m} b_j z^j = \sum_{j=0}^{2m} a_j \frac{z^j e^{-z}}{\Gamma(j+1)},$$

where $a_j = b_j\Gamma(j+1)/\|\theta\|^2$. This provides (4.3). Since $f_m(z|\theta)$ is a density on $\mathbb{R}^+$, it integrates to one, yielding $\sum_{j=0}^{2m} a_j = 1$. □

C Appendix C: Some details about the practical implementation

The method has been implemented in MATLAB, but should be easy to implement in R too. The basic calculations are summarized in Section 3.1. First, the values appearing in the optimization problem (3.1) are calculated for all $t \in \mathbb{R}$ as

$$\widehat{T}(t) = \Im\left[\frac{\sum_{j=1}^n iY_j \exp(itY_j)}{\sum_{j=1}^n \exp(itY_j)}\right] - \overline{Y},$$
$$T_\lambda(t) = \Im\Big[\psi_Z'(t|\theta)/\psi_Z(t|\theta)\Big] - \mathbb{E}(Z|\theta),$$

where explicit expressions for $\psi_Z'(t|\theta), \psi_Z(t|\theta)$ and $\mathbb{E}(Z|\theta)$ are given in (4.7), (4.8) and (4.6), respectively. The $\Im$ operator is obtained through the build-in function "`imag`" from MATLAB.

The integral in (3.1) is computed by trapezoidal methods over a grid of 2000 equidistant points over the interval $[-t^*, t^*]$, where t^* is selected following the rule motivated in Delaigle and Hall (2016), i.e., t^* is the smallest $t > 0$ such that $|\widehat{\psi}_Y(t)| \leq n^{-1/4}$. This is only a rule of thumb based on the fact that $n^{-1/2}$ is the order of magnitude of the error when estimating $\psi_Y(t)$.[5] For the weight function $w(t)$ we used the Epanechnikov kernel rescaled to the interval $[-t^*, t^*]$.[6] For the optimization procedure we used the unconstrained optimization program "`fminunc`" from MATLAB.

For the estimation of the variance, we slightly modify the formula given in (3.4) to obtain positive values for $\widehat{\sigma}_\alpha^2 = \max\{0, \sigma_Y^2 - b(\hat{\lambda}_\alpha)\}$. As suggested in the supplement of Delaigle and Hall (2016) we obtained slightly better results by fitting the polynomial $1 - (1/2)t^2\sigma^2$ to $\widehat{\psi}_\varepsilon(t)$ in a neighborhood of $t = 0$. Here

$$\widehat{\psi}_\varepsilon(t) = \frac{\widehat{\psi}_Y(t)}{\widehat{\psi}_Z(t)\exp(it\widehat{\tau}_\alpha(t))},$$

[5]Note that we tried also with a grid of only 1000 values and a threshold value for $|\widehat{\psi}_Y(t)|$ such as $n^{-1/2}$, without significant changes in the results.

[6]We tried also other kernels in some pilot experiments, like uniform, triangular, quartic, with similar results. Symmetric kernels giving more weight near zero gave slightly better results. So the results in our tables correspond to the Epanechnikov case.

and then we can define, as suggested in Delaigle and Hall (2016), the estimator

$$\widehat{\sigma}_{\alpha}^{2} = \arg\min_{\sigma \geq 0} \int_{-t_0}^{t_0} \left[\widehat{\psi}_{\varepsilon}(t) - 1 + (1/2)t^2\sigma^2 \right]^2 dt,$$

where t_0 is small and fixed as $\sup_{|t| \leq t_0} |\widehat{\psi}_{\varepsilon}(t) - 1| \leq 0.05$. In fact the two estimation approaches gave very similar results, but we report in our tables the values obtained by the Delaigle and Hall approach.

Table 1: Scenario 1: $Z \sim \text{Exp}(2)$ with $\mu_Z = \sigma_Z = 0.5$, and Gaussian noise $\text{N}(0, \sigma^2)$, see Table 2 in Kneip et al. (2015). The results are based on 200 Monte-Carlo trials.

	$n = 50$				$n = 100$				$n = 500$			
	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α
Noise to signal ratio $\rho_{\text{nts}} = 0.00, \sigma = 0.0000$												
$RMSE$	0.0652	0.0288	4	0.150	0.0523	0.0172	4	0.150	0.0249	0.0101	2	0.050
$BIAS$	0.0443	0.0090			0.0405	0.0039			0.0220	0.0024		
STD	0.0480	0.0274			0.0333	0.0168			0.0117	0.0099		
Noise to signal ratio $\rho_{\text{nts}} = 0.10, \sigma = 0.0500$												
$RMSE$	0.0598	0.0517	4	0.100	0.0387	0.0510	4	0.050	0.0160	0.0406	2	0.000
$BIAS$	0.0282	-0.0260			0.0076	-0.0114			-0.0000	-0.0118		
STD	0.0528	0.0448			0.0380	0.0499			0.0160	0.0390		
Noise to signal ratio $\rho_{\text{nts}} = 0.25, \sigma = 0.1250$												
$RMSE$	0.0695	0.0749	4	0.040	0.0472	0.0593	4	0.030	0.0262	0.0289	4	0.020
$BIAS$	0.0094	-0.0205			0.0020	-0.0085			0.0049	-0.0038		
STD	0.0690	0.0723			0.0473	0.0588			0.0258	0.0288		
Noise to signal ratio $\rho_{\text{nts}} = 0.50, \sigma = 0.2500$												
$RMSE$	0.0833	0.0776	2	0.000	0.0653	0.0607	2	0.000	0.0449	0.0271	4	0.010
$BIAS$	0.0117	0.0003			0.0071	0.0199			0.0040	-0.0036		
STD	0.0827	0.0778			0.0651	0.0575			0.0448	0.0269		
Noise to signal ratio $\rho_{\text{nts}} = 0.75, \sigma = 0.3750$												
$RMSE$	0.1116	0.0851	2	0.000	0.0858	0.0570	2	0.000	0.0459	0.0369	2	0.010
$BIAS$	0.0268	-0.0270			0.0156	-0.0039			0.0158	0.0021		
STD	0.1086	0.0810			0.0846	0.0570			0.0432	0.0369		

Table 2: Scenario 2: $Z \sim \text{Exp}(1)$ with $\mu_Z = \sigma_Z = 1$, and Gaussian noise $\text{N}(0, \sigma^2)$, see Table 3 in Kneip et al. (2015). The results are based on 200 Monte-Carlo trials.

	$n = 50$				$n = 100$				$n = 500$			
	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α
Noise to signal ratio $\rho_{\text{nts}} = 0.00, \sigma = 0.0000$												
$RMSE$	0.0896	0.1687	10	0.800	0.0633	0.1200	10	0.950	0.0247	0.0605	10	0.950
$BIAS$	-0.0200	0.1077			-0.0172	0.0789			-0.0085	0.0381		
STD	0.0875	0.1302			0.0611	0.0906			0.0232	0.0472		
Noise to signal ratio $\rho_{\text{nts}} = 0.10, \sigma = 0.1000$												
$RMSE$	0.0903	0.1360	10	0.590	0.0654	0.1025	10	0.980	0.0279	0.0513	10	0.590
$BIAS$	-0.0211	0.0341			-0.0156	0.0090			-0.0092	-0.0094		
STD	0.0881	0.1320			0.0637	0.1024			0.0264	0.0505		
Noise to signal ratio $\rho_{\text{nts}} = 0.25, \sigma = 0.2500$												
$RMSE$	0.1005	0.1230	10	0.250	0.0768	0.0887	8	0.400	0.0362	0.0394	10	0.600
$BIAS$	-0.0207	0.0132			-0.0205	0.0153			-0.0064	0.0002		
STD	0.0986	0.1226			0.0742	0.0876			0.0357	0.0395		
Noise to signal ratio $\rho_{\text{nts}} = 0.50, \sigma = 0.5000$												
$RMSE$	0.1318	0.1239	10	0.200	0.0983	0.0807	10	0.350	0.0525	0.0434	10	0.600
$BIAS$	-0.0038	-0.0177			-0.0029	0.0016			-0.0033	-0.0019		
STD	0.1321	0.1229			0.0985	0.0808			0.0525	0.0435		
Noise to signal ratio $\rho_{\text{nts}} = 0.75, \sigma = 0.7500$												
$RMSE$	0.1604	0.1592	8	0.600	0.1207	0.1013	8	0.600	0.0630	0.0504	10	0.600
$BIAS$	0.0088	-0.0432			0.0039	-0.0123			-0.0015	-0.0066		
STD	0.1605	0.1537			0.1209	0.1008			0.0631	0.0501		

Table 3: Scenario 3: $Z \sim \mathrm{N}^{+}(0, 0.8^2)$ with $\mu_Z = 0.6383$ and $\sigma_Z = 0.4822$, and Gaussian noise $\mathrm{N}(0, \sigma^2)$, see Table 4 in Kneip et al. (2015). The results are based on 200 Monte-Carlo trials.

	$n = 50$				$n = 100$				$n = 500$			
	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α
Noise to signal ratio $\rho_{\text{nts}} = 0.00, \sigma = 0.0000$												
$RMSE$	0.0689	0.0007	4	0.500	0.0461	0.0001	2	0.550	0.0250	0.0001	2	0.450
$BIAS$	0.0221	0.0001			0.0013	0.0001			0.0028	0.0001		
STD	0.0655	0.0007			0.0462	0.0000			0.0249	0.0000		
Noise to signal ratio $\rho_{\text{nts}} = 0.10, \sigma = 0.0482$												
$RMSE$	0.0662	0.0474	2	0.550	0.0463	0.0480	2	0.550	0.0254	0.0481	2	0.450
$BIAS$	-0.0031	-0.0467			0.0023	-0.0479			0.0015	-0.0480		
STD	0.0663	0.0086			0.0463	0.0027			0.0254	0.0012		
Noise to signal ratio $\rho_{\text{nts}} = 0.25, \sigma = 0.1206$												
$RMSE$	0.0837	0.0820	8	0.050	0.0836	0.0501	4	0.125	0.0656	0.0409	4	0.075
$BIAS$	0.0338	-0.0259			-0.0556	-0.0181			-0.0491	-0.0161		
STD	0.0767	0.0780			0.0625	0.0468			0.0436	0.0377		
Noise to signal ratio $\rho_{\text{nts}} = 0.50, \sigma = 0.2411$												
$RMSE$	0.1192	0.0832	2	0.100	0.0933	0.0634	4	0.100	0.0730	0.0496	10	0.025
$BIAS$	-0.0876	-0.0412			-0.0605	-0.0362			0.0559	-0.0346		
STD	0.0811	0.0724			0.0712	0.0522			0.0471	0.0356		
Noise to signal ratio $\rho_{\text{nts}} = 0.75, \sigma = 0.3617$												
$RMSE$	0.1290	0.0910	2	0.075	0.1015	0.0778	2	0.100	0.0790	0.0573	6	0.050
$BIAS$	-0.0884	-0.0441			-0.0772	-0.0464			0.0473	-0.0421		
STD	0.0942	0.0798			0.0660	0.0626			0.0634	0.0390		

Table 4: Scenario 4: $Z \sim \mathrm{N}^{+}(0.6, 0.6^2)$ with $\mu_Z = 0.7726$ and $\sigma_Z = 0.4761$, and Gaussian noise $\mathrm{N}(0, \sigma^2)$, see Table 5 in Kneip et al. (2015). The results are based on 200 Monte-Carlo trials.

	$n = 50$				$n = 100$				$n = 500$			
	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α
Noise to signal ratio $\rho_{\text{nts}} = 0.00, \sigma = 0.0000$												
$RMSE$	0.0673	0.0001	6	1.000	0.0494	0.0001	6	1.200	0.0199	0.0002	8	1.400
$BIAS$	-0.0089	0.0001			-0.0054	0.0001			0.0010	0.0002		
STD	0.0669	0.0000			0.0492	0.0000			0.0199	0.0000		
Noise to signal ratio $\rho_{\text{nts}} = 0.10, \sigma = 0.0476$												
$RMSE$	0.0681	0.0475	6	1.000	0.0495	0.0475	6	1.200	0.0204	0.0475	6	1.400
$BIAS$	-0.0086	-0.0475			-0.0039	-0.0475			-0.0015	-0.0475		
STD	0.0678	0.0000			0.0494	0.0000			0.0204	0.0000		
Noise to signal ratio $\rho_{\text{nts}} = 0.25, \sigma = 0.1190$												
$RMSE$	0.1038	0.0673	10	0.200	0.0877	0.0515	10	0.250	0.0738	0.0222	8	0.250
$BIAS$	-0.0712	-0.0228			-0.0695	-0.0193			-0.0700	-0.0068		
STD	0.0757	0.0635			0.0537	0.0478			0.0233	0.0212		
Noise to signal ratio $\rho_{\text{nts}} = 0.50, \sigma = 0.2381$												
$RMSE$	0.1097	0.0949	10	0.150	0.0994	0.0590	8	0.125	0.0891	0.0280	8	0.125
$BIAS$	-0.0684	-0.0589			-0.0810	-0.0259			-0.0846	-0.0129		
STD	0.0859	0.0746			0.0578	0.0531			0.0281	0.0249		
Noise to signal ratio $\rho_{\text{nts}} = 0.75, \sigma = 0.3571$												
$RMSE$	0.1250	0.1027	8	0.075	0.0935	0.0815	6	0.100	0.0884	0.0385	6	0.100
$BIAS$	-0.0766	-0.0510			-0.0638	-0.0557			-0.0802	-0.0280		
STD	0.0990	0.0894			0.0685	0.0597			0.0374	0.0266		

Table 5: Scenario 5: $Z \sim \mathrm{N}^{+}(0, 0.8^2)$ with $\mu_Z = 0.6383$ and $\sigma_Z = 0.4822$, and Student noise $C_1 \times \mathrm{Stud}(4)$, where $C_1 = \sigma/\sqrt{2}$, see Table 6 in Kneip et al. (2015). The results are based on 200 Monte-Carlo trials.

	$n = 50$				$n = 100$				$n = 500$			
	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α
Noise to signal ratio $\rho_{\mathrm{nts}} = 0.00, \sigma = 0.0000$												
$RMSE$	0.0689	0.0007	4	0.500	0.0461	0.0001	2	0.550	0.0250	0.0001	2	0.450
$BIAS$	0.0221	0.0001			0.0013	0.0001			0.0028	0.0001		
STD	0.0655	0.0007			0.0462	0.0000			0.0249	0.0000		
Noise to signal ratio $\rho_{\mathrm{nts}} = 0.10, \sigma = 0.0482$												
$RMSE$	0.0663	0.0475	2	0.550	0.0468	0.0480	2	0.550	0.0253	0.0481	2	0.450
$BIAS$	-0.0032	-0.0467			0.0023	-0.0479			0.0016	-0.0480		
STD	0.0664	0.0090			0.0469	0.0028			0.0253	0.0012		
Noise to signal ratio $\rho_{\mathrm{nts}} = 0.25, \sigma = 0.1206$												
$RMSE$	0.0814	0.0848	8	0.075	0.0790	0.0569	4	0.150	0.0649	0.0414	4	0.075
$BIAS$	0.0394	-0.0371			-0.0478	-0.0335			-0.0488	-0.0173		
STD	0.0714	0.0764			0.0631	0.0460			0.0428	0.0377		
Noise to signal ratio $\rho_{\mathrm{nts}} = 0.50, \sigma = 0.2411$												
$RMSE$	0.0967	0.1155	8	0.050	0.0998	0.0643	2	0.100	0.0731	0.0582	10	0.025
$BIAS$	0.0412	-0.0564			-0.0837	-0.0374			0.0540	-0.0420		
STD	0.0876	0.1010			0.0545	0.0525			0.0494	0.0404		
Noise to signal ratio $\rho_{\mathrm{nts}} = 0.75, \sigma = 0.3617$												
$RMSE$	0.1282	0.1137	2	0.075	0.1098	0.0807	2	0.075	0.0768	0.0677	6	0.050
$BIAS$	-0.0912	-0.0542			-0.0890	-0.0412			0.0376	-0.0511		
STD	0.0903	0.1002			0.0645	0.0696			0.0671	0.0445		

Table 6: Scenario 6: $Z \sim \mathrm{N}^{+}(0, 0.8^2)$ with $\mu_Z = 0.6383$ and $\sigma_Z = 0.4822$, and Laplace noise $C_2 \times$ Laplace, where $C_2 = \sigma/\sqrt{2}$, see Table 7 in Kneip et al. (2015). The results are based on 200 Monte-Carlo trials.

	$n = 50$				$n = 100$				$n = 500$			
	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α	$\widehat{\tau}$	$\widehat{\sigma}$	m	α
Noise to signal ratio $\rho_{\text{nts}} = 0.00, \sigma = 0.0000$												
$RMSE$	0.0689	0.0007	4	0.500	0.0461	0.0001	2	0.550	0.0250	0.0001	2	0.450
$BIAS$	0.0221	0.0001			0.0013	0.0001			0.0028	0.0001		
STD	0.0655	0.0007			0.0462	0.0000			0.0249	0.0000		
Noise to signal ratio $\rho_{\text{nts}} = 0.10, \sigma = 0.0482$												
$RMSE$	0.0673	0.0474	2	0.550	0.0460	0.0480	2	0.550	0.0254	0.0481	2	0.450
$BIAS$	-0.0031	-0.0468			0.0012	-0.0480			0.0015	-0.0480		
STD	0.0674	0.0073			0.0461	0.0014			0.0254	0.0011		
Noise to signal ratio $\rho_{\text{nts}} = 0.25, \sigma = 0.1206$												
$RMSE$	0.0888	0.0845	8	0.075	0.0805	0.0570	4	0.150	0.0643	0.0374	4	0.075
$BIAS$	0.0416	-0.0394			-0.0521	-0.0315			-0.0498	-0.0143		
STD	0.0787	0.0749			0.0615	0.0477			0.0407	0.0346		
Noise to signal ratio $\rho_{\text{nts}} = 0.50, \sigma = 0.2411$												
$RMSE$	0.1285	0.0777	2	0.075	0.0805	0.0831	8	0.050	0.0703	0.0533	10	0.025
$BIAS$	-0.0999	-0.0226			0.0521	-0.0438			0.0547	-0.0356		
STD	0.0810	0.0745			0.0616	0.0708			0.0444	0.0398		
Noise to signal ratio $\rho_{\text{nts}} = 0.75, \sigma = 0.3617$												
$RMSE$	0.1319	0.1033	2	0.075	0.0923	0.0944	8	0.025	0.0761	0.0588	10	0.025
$BIAS$	-0.0947	-0.0450			0.0527	-0.0414			0.0586	-0.0380		
STD	0.0920	0.0932			0.0760	0.0851			0.0487	0.0450		

Table 7: Bivariate example. Same scenario as for Table 8 in Kneip et al. (2015) and conditioning on $W = w$. The results are based on 200 Monte-Carlo trials.

n			$w = 0.25$	$w = 0.50$	$w = 0.75$
100	$\tau(w)$	$BIAS$	0.0027	0.0071	0.0092
		STD	0.0272	0.0426	0.0489
		MSE	0.0007	0.0019	0.0025
	$\sigma(w)$	$BIAS$	-0.0265	-0.0085	-0.0075
		STD	0.0573	0.0644	0.0708
		MSE	0.0040	0.0042	0.0051
500	$\tau(w)$	$BIAS$	0.0003	0.0036	0.0054
		STD	0.0141	0.0186	0.0246
		MSE	0.0002	0.0004	0.0006
	$\sigma(w)$	$BIAS$	-0.0187	0.0019	0.0028
		STD	0.0431	0.0492	0.0500
		MSE	0.0022	0.0024	0.0025

Table 8: Bivariate example. Same scenario as for Table 8 in Kneip et al. (2015) and conditioning on $W \leq w$. The results are based on 200 Monte-Carlo trials.

n			$w = 0.25$	$w = 0.50$	$w = 0.75$
100	$\tau(w)$	$BIAS$	-0.0090	-0.0032	-0.0034
		STD	0.0437	0.0474	0.0478
		MSE	0.0020	0.0023	0.0023
	$\sigma(w)$	$BIAS$	-0.0580	-0.0649	-0.0661
		STD	0.0441	0.0138	0.0061
		MSE	0.0053	0.0044	0.0044
500	$\tau(w)$	$BIAS$	0.0002	-0.0027	-0.0001
		STD	0.0220	0.0215	0.0195
		MSE	0.0005	0.0005	0.0004
	$\sigma(w)$	$BIAS$	-0.0661	-0.0666	-0.0665
		STD	0.0067	0.0000	0.0000
		MSE	0.0044	0.0044	0.0044

Table 9: Values of the coefficients $a_1, \ldots, a_7$ for the densities in Figure 2.

density	a_1	a_2	a_3	a_4	a_5	a_6	a_7
d_1	0.0086	0.0058	0.1519	0.0564	3.8010	-8.3671	5.3434
d_2	1.4374	-1.3021	-0.1233	1.1649	0.1756	-0.4860	0.1336
d_3	0.1928	0.3393	0.4643	0.2048	-0.6054	-1.0000	1.4042
d_4	0.2183	1.0974	1.0840	-11.9649	25.8999	-25.3179	9.9832
d_5	0.3496	-0.7125	1.3596	-1.9028	1.5836	0.3062	0.0163