INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

2012/02

Replaces DP 1046

BOUNDARY ESTIMATION IN THE PRESENCE OF MEASUREMENT ERROR WITH UNKNOWN VARIANCE

KNEIP, A., SIMAR, L. and I. VAN KEILEGOM

Boundary estimation in the presence of measurement error with unknown variance

Alois Kneip[1] Léopold Simar[2] Ingrid Van Keilegom[2]

January 10, 2012

Abstract

Boundary estimation appears naturally in economics in the context of productivity analysis. The performance of a firm is measured by the distance between its achieved output level (quantity of goods produced) and an optimal production frontier which is the locus of the maximal achievable output given the level of the inputs (labor, energy, capital, etc.). Frontier estimation becomes difficult if the outputs are measured with noise and most approaches rely on restrictive parametric assumptions. This paper contributes to the direction of nonparametric approaches.

We consider a general setup with unknown frontier and unknown variance of a normally distributed error term, and we propose a nonparametric method which allows to identify and estimate both quantities simultaneously. The asymptotic consistency and the rate of convergence of our estimators are established, and simulations are carried out to verify the performance of the estimators for small samples. We also apply our method on a dataset concerning the production output of American electricity utility companies.

Key words: deconvolution, stochastic frontier estimation, nonparametric estimation, penalized likelihood

[1] Department of Economics, University of Bonn, Bonn, Germany, E-mail: akneip@uni-bonn.de

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

1 Introduction

Boundary estimation problems arise naturally in economics, in the context of productivity analysis. When analyzing the productivity of firms, one may compare how the firms transform their inputs W (labor, energy, capital, etc.) into an output X (the quantity of goods produced). In this context, the set of technically possible outputs is determined by a production frontier $\tau(W)$ which is the geometric locus of optimal production plans. The economic efficiency of the firm operating at the level (W_0, X_0) is then measured in terms of the distance between its production level X_0 and the boundary level $\tau(W_0)$.

Efficiency and productivity analysis have been applied in many different fields of economic activity, including industry, hospitals, transportation, schools, banks, public services, etc. Frontier models were even introduced to measure the performance of portfolios in finance, in the line with the seminal work of Markovitz (1959) using Capital Assets Pricing Models (CAPM), where W measures the risk of a portfolio and X its average return. Gattoufi, Oral and Reisman (2004) cite more than 1,800 published articles on efficiency analysis, appearing in more than 400 journals in business and economics.

In *deterministic* frontier models it is assumed that $\tau(W)$ corresponds to the boundary of the support of X. For a random sample (W_i, X_i) one then has $P(X_i \leq \tau(W_i)) = 1$. Most nonparametric approaches are then based on the idea of enveloping the data. Farrell (1957) introduced Data Envelopment Analysis (DEA), based on either the conical hull or the convex hull of the data. Deprins et al. (1984) extended the idea to non convex sets and suggested the Free Disposal Hull (FDH) estimator, equal to the smallest free disposal set containing all the data. Statistical properties of these estimators are well known (see Banker, 1993; Korostelev et al., 1995a,b; Kneip et al., 1998; Gijbels et al., 1999; Park et al. 2000; Jeong, 2004; Jeong and Park, 2006; Kneip et al. 2008; Park et al. 2010; Daouia et al. 2010). However all these methods rely on the unrealistic assumption of deterministic frontier models that the outputs X_i are observed without noise. In the presence of noise, the envelopment methods will be biased and not consistent.

More realistic *stochastic* frontier models assume that observed outputs Y_i represent underlying, "true" outputs X_i contaminated with some additional noise. In most of the stochastic frontier approaches developed in the econometric literature, a fully parametric model is assumed. For instance, in the pioneering work of Aigner et al. (1977) and Meeusen and van den Broek (1977), we have an iid sample of (W_i, Y_i) of inputs and outputs generated by the basic model

$$Y_i = \tau(W_i) \exp(-U_i) \exp(V_i), \tag{1.1}$$

where $\tau(W_i)$ is a parametric production function (e.g. a Cobb-Douglas) quantifying the optimal attainable output for a given input level W_i. Moreover, $U_i > 0$ is a positive random variable having a jump at the origin that represents the inefficiency; in parametric models, U_i has a known density depending on one or two unknown parameters (often a half normal, truncated normal or exponential). So the latent unobserved output is $X_i = \tau(W_i)\exp(-U_i)$. The noise term is $Z_i = \exp(V_i)$, where $V_i \in \mathbb{R}$ has usually a normal density with mean zero and unknown variance. Finally, U_i is supposed to be conditionally independent of V_i, given W_i. These approaches have been very popular in the econometric literature and estimation is based on standard parametric techniques, like maximum likelihood or modified least squares methods (see Greene 2008, for a survey).

However, these approaches rely on very restrictive assumptions on both the frontier function and on the stochastic part of the model. A crucial issue is the specification of the distribution of the inefficiencies U_i. While some central limit arguments can be advocated for the Gaussian noise, there does usually not exist any information justifying particular distributional assumptions on U_i.

Recent attempts have been made to attack the problem from a non- or semi-parametric point of view. Using nonparametric techniques it is possible to avoid any parametric assumptions on the structure of $\tau(W_i)$. Important contributions in this direction are Fan et al. (1996) and Kumbhakar et al. (2007). They, however, still rely on parametric specifications for the density of U_i.

Even when assuming Gaussian noise, dropping parametric assumptions on the structure of the distribution of U_i greatly complicates the problem and enforces to develop completely new methods. Estimation of the boundary $\tau(W)$ of X then necessitates to solve a complicated, non-standard deconvolution problem.

In order to concentrate on the core of the problem, we will start by analyzing a slightly simplified version of the general model which assumes that the boundary $\tau(\cdot)$ is constant, i.e. $\tau(W) \equiv \tau$ for all W and some fixed, but unknown $\tau > 0$. With $X = \tau\exp(-U)$ and $Z = \exp(V)$ the general setup then reduces to the following situation: There are i.i.d. observations $Y_1, \dots, Y_n$ with a density g on $\mathbb{R}_+$, generated by the model

$$Y_i = X_i \cdot Z_i, \tag{1.2}$$

where X_i is a latent unobserved true signal having a density f on the support $[0, \tau]$, with $f(\tau) > 0$ for some unknown $\tau > 0$, and Z_i is the noise. We assume that Z_i is independent of X_i and is log-normally distributed. More precisely, $\log Z_i \sim N(0, \sigma^2)$, where $\sigma^2 > 0$ is an unknown variance. The problem then consists in estimating τ as well as σ, when only the Y_i's are observable.

Our estimation procedure for the simplified model (1.2) is based on the maximization of a penalized profile likelihood. Based on local constant or local linear approximation techniques this approach is then generalized to define estimators for the stochastic frontier model (1.1). Precise descriptions of estimators and a corresponding asymptotic theory are given in Sections 2 and 3.

Our basic approach is similar to the setup described in Hall and Simar (2002). They propose a nonparametric approach where the noise has an unspecified symmetric density with variance σ^2 converging to zero when the sample size increases. Different from their approach we avoid the restriction of having the noise converging to zero when the sample size increases. We want to note, however, that a lognormal distribution of Z is crucial to ensure identifiability in our context, while Hall and Simar (2002) rely on unspecified error distributions.

As already mentioned above, (1.2) with unknown τ and σ leads to a non-standard deconvolution problem. The novelty of our approach consists in the simultaneous estimation of both parameters and the derivation of resulting convergence rates. The problem of estimating an unknown boundary τ for a *known* error variance σ^2 has already been studied in a number of papers, see e.g. Goldenshluger and Tsybakov (2004), Delaigle and Gijbels (2006), Meister (2006), or Aarts, Groeneboom and Jongbloed (2007). Another related problem is the deconvolution problem with unknown error variance, but without assuming the existence of a finite boundary. Butucea and Matias (2005), Butucea, Matias and Pouet (2008), as well as Schwarz and Van Bellegem (2009) proposed estimators under this model, and they proved (among others) the identifiability and consistency of their estimators.

The paper is organized as follows. Sections 2 and 3 describe our estimation procedure and corresponding asymptotic properties, respectively. Numerical illustrations are presented in Section 4. We first begin with a simulation study to verify the performance of the estimators in (1.2) for small samples. We then compare the performance of our estimator of a production frontier with the procedure proposed in Hall and Simar (2002). We also apply our procedure to analyze the production outputs of American electricity utility companies. Proofs of some core results can be found in Section 5.

2 Estimation procedure

2.1 Estimation under the simplified model

Recall that under model (1.2), the latent variable X is defined on $[0, \tau]$ and its density f satisfies $f(\tau) > 0$. In addition, let g be the density of the observed variable Y. Also note that the model can equivalently be written as $Y^* = X^* + Z^*$, where $Y^* = \log Y$,

$X^* = \log X$ and where $Z^* \sim N(0, \sigma^2)$ is independent of X^*, and σ^2 is unknown.

Whenever confusion is possible, we will add a subindex 0 to indicate the true quantities (e.g. $f_0, g_0, \tau_0, \ldots$ stand for the true densities f and g and the true value of τ). Let $\phi(z)$ denote the standard normal density, and recall that the density ρ_σ of a log-normal random variable with parameters $\mu = 0$, $\sigma^2 > 0$ is given by $\rho_\sigma(z) = \frac{1}{\sigma z}\phi(\frac{\log z}{\sigma})$ for $z > 0$. For all $y > 0$ we can then write

$$\begin{aligned} g_0(y) &= \int_0^{\tau_0} f_0(x)\frac{1}{x}\rho_{\sigma_0}(\frac{y}{x})dx = \int_0^1 h_0(t)\frac{1}{t\tau_0}\rho_{\sigma_0}(\frac{y}{t\tau_0})dt \\ &= \frac{1}{\sigma_0 y}\int_0^1 h_0(t)\phi\left(\frac{1}{\sigma_0}\log\frac{y}{t\tau_0}\right)dt, \end{aligned} \tag{2.1}$$

where

$$h_0(t) = \tau_0 f_0(t\tau_0) \qquad \text{for } 0 \le t \le 1.$$

For an arbitrary density h defined on $[0,1]$ and for arbitrary values of $\tau > 0$ and $\sigma > 0$, define

$$g_{h,\tau,\sigma}(y) = \frac{1}{\sigma y}\int_0^1 h(t)\phi\left(\frac{1}{\sigma}\log\frac{y}{t\tau}\right)dt.$$

Obviously, $g_0 \equiv g_{h_0,\tau_0,\sigma_0}$. It will be shown in the next section that under some additional assumptions all parameters of the model are identifiable.

Our estimation procedure relies on the maximization of a penalized profile likelihood. Obviously the likelihood function is based on the density $g_{h,\tau,\sigma}$. Note that this density not only depends on the parameters τ and σ, but also on the underlying density h. We will use histogram-type estimators to approximate h, and therefore maximization will be done over all possible values of τ, σ as well as all h in the specified class of histogram estimators. A penalization is introduced in order to account for a possible smoothness of h_0.

More precisely, for a pre-specified natural number M let

$$\Gamma = \left\{\gamma = (\gamma_1, \ldots, \gamma_M) : \gamma_k > 0 \text{ for all } k \text{ and } \sum_{k=1}^M \gamma_k = M\right\},$$

and define

$$h_\gamma(t) = \gamma_1 I(t=0) + \sum_{k=1}^M \gamma_k I(q_{k-1} < t \le q_k)$$

for $0 \le t \le 1$, where $q_k = k/M$ $(k = 0, 1, \ldots, M)$. It is clear that h_γ is a density for all $\gamma \in \Gamma$. Then

$$g_{h_\gamma,\tau,\sigma}(y) = \frac{1}{\sigma y}\int_0^1 h_\gamma(t)\phi\left(\frac{1}{\sigma}\log\frac{y}{t\tau}\right)dt = \frac{1}{\sigma y}\sum_{k=1}^M \gamma_k \int_{q_{k-1}}^{q_k} \phi\left(\frac{1}{\sigma}\log\frac{y}{t\tau}\right)dt.$$

Estimators $\widehat{\tau}$, $\widehat{\sigma}$ (and $\widehat{h} := h_{\widehat{\gamma}}$) of τ_0, σ_0 (and h_0) are now obtained by maximizing the following penalized likelihood:

$$(\widehat{\tau}, \widehat{\sigma}, \widehat{\gamma}) = \text{argmax}_{\tau>0,\sigma>0,\gamma\in\Gamma}\Big\{ n^{-1}\sum_{i=1}^{n} \log g_{h_\gamma,\tau,\sigma}(Y_i) - \lambda\, \text{pen}(g_{h_\gamma,\tau,\sigma})\Big\},$$

where $\lambda \geq 0$ is a fixed value independent of n, and where

$$\text{pen}(g_{h_\gamma,\tau,\sigma}) = \max_{3\leq j\leq M} |\gamma_j - 2\gamma_{j-1} + \gamma_{j-2}|$$

The procedure also leads to an estimator $\widehat{g} := g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}$ of the density g_0 of Y.

Note that λ can be taken equal to zero, which means that we consider both penalized and non-penalized estimators. However, as will be shown in the next section, the penalized estimator attains a better rate of convergence if h_0 is smooth. It is then preferable over the non-penalized one. It is also important to highlight here that λ is a parameter that is chosen independent of the sample size, which is in contrast to most other penalized estimation methods in the literature, where λ is usually chosen as a function of the sample size n.

Note that $\text{pen}(\widehat{g}) = \max_{3\leq j\leq M} |\widehat{h}(\frac{j}{M}) - 2\widehat{h}(\frac{j-1}{M}) + \widehat{h}(\frac{j-2}{M})|$. But $\max_{3\leq j\leq M} |h_0(\frac{j}{M}) - 2h_0(\frac{j-1}{M}) + h_0(\frac{j-2}{M})| = M^{-2} \max_{3\leq j\leq M} |h_0''(\frac{j-1}{M})|(1 + o_P(1))$. Since $\text{pen}(\widehat{g}) = O_P(M^{-2})$ we thus ensure that the structure of our discretized estimator appropriately reflects the underlying smoothness of h_0 if $M \equiv M_n \to \infty$. We also refer to the proofs of the asymptotic results shown in the next section for better understanding the motivation for the precise formula of the penalty term.

Also note that although the above procedure is valid for practical purposes, we need to be a bit more precise when developing the asymptotic results in Section 3. In particular, we will require that σ and τ belong to a compact interval and that the function $h_\gamma(t)$ is uniformly bounded above on $[0, t]$ and uniformly bounded below from 0 for t close to 1. However, except for the latter condition on the lower bound for h_γ, these conditions do not play any role in practice, since the intervals and the upper bound can be chosen in an arbitrary way. We refer to assumption (A1) below for the precise formulation of the support of σ, τ and h_γ.

Remark 2.1 Our use of histogram-type estimators of h_0 is motivated by computational feasibility. Note that $g_{h_\gamma,\tau,\sigma}(y)$ is a *linear* function of γ. This greatly simplifies the maximization problem. For fixed τ, σ, the corresponding optimal values $\widehat{\gamma}(\tau, \sigma)$ of γ can easily be determined by applying standard optimization techniques. Subsequent maximization over τ, σ leads to a numerically stable algorithm with moderate computation times.

One will usually tend to assume that the true density h_0 is smooth on $[0,1]$. In this case penalized likelihood estimation with $\lambda > 0$ simply corrects a default of the "pure" histogram estimator which due to its discontinuous nature does not make any use of smoothness of h_0. This type of penalization is unnecessary when relying on smooth sieve estimators. For example, h may alternatively be approximated by cubic B-splines with M equidistant knots. It will be shown in Section 3 that if h_0 is twice continuously differentiable, then spline-based estimators (without penalty) achieve the same rate of convergence as penalized histogram-based estimators. The drawback of this alternative approach is that it leads to a maximization problem which is much more complex from a computational point of view. When trying to implement spline-based estimators we encountered prohibitive computation times as well as serious problems with numerical stability.

2.2 Incorporating covariates

Let us return to the general model (1.1) incorporating some d-dimensional covariate $W \in \mathbb{R}^d$, $d \geq 1$. We then have $Y_i = \tau(W_i)\exp(-U_i)\exp(V_i)$, and data consist of i.i.d. observations (W_i, Y_i), $i = 1, \dots, n$. For some value w_0 in the interior of the support of W the problem to be considered is to estimate the boundary $\tau_0(w_0)$ of the *conditional* distribution of $X = \tau_0(W)\exp(-U)$ given $W = w_0$.

Our setup consists in a straightforward generalization of the conditions used for analyzing the simplified model. We suppose that the conditional distribution of V given $W = w_0$ is $N(0, \sigma^2(w_0))$ with a true variance $\sigma_0^2(w_0)$ possibly depending on w_0. As mentioned in the introduction we furthermore assume that the components U and V of the model are conditionally independent (conditionally to W), and that U is a positive random variable. Then obviously $\exp(-U)$ takes values in $[0,1]$, and it is assumed that the conditional distribution of $\exp(-U)$ given $W = w_0$ possesses a density h_{0w_0} with $h_{0w_0}(1) > 0$. Consequently, the conditional density f_{0w_0} of $X = \tau_0(W)\exp(-U)$ given $W = w_0$ satisfies $h_{0w_0}(t) = \tau_0(w_0) f_{0w_0}(t\tau_0)$, $0 \leq t \leq 1$. Using the same notation as above, the conditional density of Y given $W = w_0$ is then equal to $g_{0w_0} = g_{h_{0w_0},\tau(w_0),\sigma(w_0)}$.

Resorting to the same ideas as in Hall and Simar (2002), the problem of estimating $\tau_0(w_0)$ can be viewed as a local boundary problem. The approach then consists in specifying a bandwidth b and determining estimates $\widehat{\tau}(w_0)$ of $\tau(w_0)$ and $\widehat{\sigma}^2(w_0)$ of $\sigma^2(w_0)$ by the penalized likelihood procedure described above, using only those observations Y_i with $\|W_i - w_0\|_2 \leq b$ (where $\|\cdot\|_2$ is the Euclidean distance). More precisely, for given $b > 0$,

$\lambda > 0$, and $n_b := \sharp\{W_i : \|W_i - w_0\|_2 \leq b\}$, the estimators are to be determined by

$$(\widehat{\tau}(w_0), \widehat{\sigma}(w_0), \widehat{\gamma}(w_0)) = \operatorname{argmax}_{\tau>0,\sigma>0,\gamma\in\Gamma}\Big\{n_b^{-1} \sum_{i:\|W_i-w_0\|_2\leq b} \log g_{h_\gamma,\tau,\sigma}(Y_i) - \lambda \operatorname{pen}(g_{h_\gamma,\tau,\sigma})\Big\}. \tag{2.2}$$

The estimators $\widehat{h}_{w_0} = h_{\widehat{\gamma}(w_0)}$ and $\widehat{g}_{w_0} = g_{\widehat{h}_{w_0},\widehat{\tau}(w_0),\widehat{\sigma}(w_0)}$ then provide estimates of the conditional densities h_{0w_0} and g_{0w_0}.

The basic motivation of this estimation procedure of course consists in the fact that under suitable smoothness assumptions (see Section 3) the conditional density $\widetilde{g}_b$ of Y given $\|W_i - w_0\|_2 \leq b$ satisfies $\widetilde{g}_b(y) = g_{0w_0}(y) + O(b^2)$. Indeed, it will be shown in the next section that when $b \to 0$, $nb^d \to \infty$ as $n \to \infty$, the resulting estimators achieve the same (logarithmic) rates of convergence as the estimators obtained under the simplified model (1.2).

We want to emphasize that in the presence of covariates the proposed procedure (2.2) constitutes a very simple approach which is based on a "locally constant" approximation of $\tau_0(\cdot)$ and $\sigma_0^2(\cdot)$. While it is common practice to assume that the distributions of U and V are approximately constant, researchers usually cannot exclude considerable local variation of the frontier function $\tau_0(\cdot)$. In such situations one may tend to prefer some method relying on some "locally linear" approximation of $\tau_0(\cdot)$.

In order to construct such a procedure first note that when taking logarithms in (1.1) we obtain $\log Y_i = \log \tau(W_i) - U_i + V_i$. By definition, for any value w in the support of W we have $E(V|\, W = w) = 0$. Additionally assume that in a small neighborhood of the point w_0 of interest $E(U|\, W = w)$ is constant, while $\tau_0(\cdot)$ is sufficiently smooth and can be well represented by a Taylor expansion, i.e. $\log \tau_0(w) \approx \log \tau_0(w_0) + \beta(w_0)^T(w - w_0)$ with $\beta(w_0) = \frac{\partial}{\partial w}(\log \tau_0)(w)|_{w=w_0}$. For a small $b > 0$ we then obtain

$$\log Y_i \approx \alpha(w_0) + \beta(w_0)^T(W_i - w_0) - U_i^0 + V_i, \qquad \text{if } \|W_i - w_0\|_2 \leq b,$$

where $\alpha(w_0) = \log \tau_0(w_0) - E(U|\, W = w_0)$ and $U_i^0 = U_i - E(U|\, W = w_0)$. The terms $-U_i^0 + V_i$ can then be interpreted as (approximately) zero mean error terms, and hence $\alpha(w_0)$ and $\beta(w_0)$ can be estimated by ordinary local least squares. The estimate $\widehat{\beta}(w_0)$ of $\beta(w_0)$ provides information about the local variation of $\log \tau_0(\cdot)$. This can be used to calculate a suitable correction of the likelihood function. Obviously, for all W_i with $\|W_i - w_0\|_2 \leq b$ we have:

$$\log \widetilde{Y}_i := \log Y_i - \beta(w_0)^T(W_i - w_0) \approx \alpha(w_0) - U_i^0 + V_i = \log \tau_0(w_0) - U_i + V_i,$$

and thus

$$\widetilde{Y}_i \approx \tau_0(w_0) \exp(-U_i) \exp(V_i),$$

which suggests to estimate the parameters of interest using a penalized likelihood based on $\widetilde{Y}_i$ instead of Y_i. This will provide estimators $\widehat{\tau}(w_0)$ and $\widehat{\sigma}(w_0)$. Then by using $\widehat{\alpha}(w_0)$, a natural estimate of $E(U|\ W = w_0)$ is given by $\log \widehat{\tau}(w_0) - \widehat{\alpha}(w_0)$.

When combining these arguments we arrive at the following alternative estimation method:

[1] Fix a bandwidth $b > 0$ and determine estimates $\widehat{\alpha}(w_0)$ and $\widehat{\beta}(w_0)$ by minimizing

$$\sum_{i:\|W_i - w_0\|_2 \leq b} (\log Y_i - \alpha - \beta^T (W_i - w_0)^2)$$

over all α and β.

[2] Calculate $\log \widehat{\widetilde{Y}}_i := \log Y_i - \widehat{\beta}(w_0)^T (W_i - w_0)$, $i = 1, \ldots, n_b$. Then determine estimates $\widehat{\tau}(w_0), \widehat{\sigma}(w_0), \widehat{\gamma}(w_0)$ by using (2.2) with Y_i being replaced by $\widehat{\widetilde{Y}}_i$, $i = 1, \ldots, n_b$.

[3] Finally, we also have $\widehat{E}(U|\ W = w_0) = \log \widehat{\tau}(w_0) - \widehat{\alpha}(w_0)$, as an estimator of the conditional expected inefficiency level.

In order to avoid an overloading of the present paper a theoretical analysis of this modification is omitted. However, the general structure of the theory presented in Section 3 strongly indicates that these alternative estimators will achieve the same rates of convergence as those obtained by directly applying (2.2). In many applications one may nevertheless expect some improvement in finite sample behavior when using [1] and [2] instead of (2.2). In our simulation study (see Section 4) we concentrate on estimates based on [1] and [2], and it turns out that they show a surprisingly high level of accuracy.

Remark 2.2 In practice, one will of course be interested in estimating $\tau(\cdot)$ not only at a single point w_0 but at a whole sequence $w_{01}, w_{02}, \ldots$ of different possible values of W. The above procedures can then be used at any point w_{0i}, but estimation error will induce some random fluctuations of the resulting $\widehat{\tau}(w_{0i})$. It will therefore sometimes make sense to additionally use some nonparametric smoothing procedure in order to generate a final estimate of $\tau(\cdot)$ from the raw estimates $\widehat{\tau}(w_{01}), \widehat{\tau}(w_{02}), \ldots$. For example, one may use kernel estimators. Other interesting candidates are isotonic regression methods, since in most applications it can be assumed that $\tau(w)$ is monotone in w. Details are not in the scope of the present paper.

3 Asymptotic results

In this section we establish the rate of convergence of our estimators. The proofs of these results can be found in the online supplement, except for our core result in Theorem 3.2, whose proof is given in the appendix.

For the asymptotic results for the simplified model (1.2), we need to make the following assumptions:

(A1) For some $0 < \sigma_{min} < \sigma_{max} < \infty$, $0 < \tau_{min} < \tau_{max} < \infty$, $0 < h_{min} < h_{max} < \infty$, and $0 < \delta < 1$ the estimators $(\widehat{g}, \widehat{\tau}, \widehat{\sigma})$ defined in steps 1 and 2 of our procedure are determined by minimizing over all

$$(h_\gamma, \tau, \sigma) \in \mathcal{H}_n \times [\tau_{min}, \tau_{max}] \times [\sigma_{min}, \sigma_{max}],$$

where $\mathcal{H}_n \subset \mathcal{H}_{h_{min}, h_{max}, \delta}$. Here, $\mathcal{H}_{h_{min}, h_{max}, \delta}$ denotes the set of all square integrable densities h with support $[0, 1]$ satisfying

- $\sup_{t\in[0,1]} h(t) \leq h_{max}$, as well as
- $\inf_{t\in[1-\delta,1]} h(t) \geq h_{min}$.

(A2) $h_0 \in \mathcal{H}_{h_{min}, h_{max}, \delta}$ and is twice continuously differentiable, $\tau_0 \in [\tau_{min}, \tau_{max}]$, and $\sigma_0 \in [\sigma_{min}, \sigma_{max}]$.

(A3) For some $0 < \beta < 1/5$, $M = M_n \sim n^\beta$ as n tends to ∞.

(A4) For some $A > \sqrt{2}$, $P(\log Y < -A(\log n)^{1/2}\sigma_0) = o(n^{-1})$.

Remark 3.1 Note that condition (A4) is a natural condition, and is satisfied when e.g. $h_0 \equiv 0$ on a small interval $[0, \epsilon]$ close to 0.

For two arbitrary densities g_1 and g_2, let

$$H^2(g_1, g_2) = \frac{1}{2} \int \left(\sqrt{g_1(y)} - \sqrt{g_2(y)}\right)^2 dy$$

be the Hellinger distance between g_1 and g_2.

Theorem 3.1 *Assume (A1)–(A4). Then, if $\lambda \geq 0$,*

$$H^2(\widehat{g}, g_0) = O_P(M_n^{-2}),$$

and if $\lambda > 0$,

$$pen(\widehat{g}) = O_P(M_n^{-2}).$$

Theorem 3.2 *Under the assumptions of Theorem 3.1, σ_0 and τ_0 are identifiable, and we have:*

a) If $\lambda = 0$ (i.e. without penalization),

$$\widehat{\sigma} - \sigma_0 = O_P\Big((\log n)^{-1}\Big), \tag{3.1}$$

$$\widehat{\tau} - \tau_0 = O_P\Big((\log n)^{-\frac{1}{2}}\Big). \tag{3.2}$$

b) If $\lambda > 0$ (i.e. with penalization),

$$\widehat{\sigma} - \sigma_0 = O_P\Big((\log n)^{-2}\Big), \tag{3.3}$$

$$\widehat{\tau} - \tau_0 = O_P\Big((\log n)^{-\frac{3}{2}}\Big), \tag{3.4}$$

$$\widehat{h}(1) - h_0(1) = O_P\Big((\log n)^{-1}\Big). \tag{3.5}$$

It can easily be seen from the proof given in the appendix that the validity of (3.1) - (3.5) does not depend on the method employed to construct the estimator $\widehat{g}$ of g_0. It is only required that $\widehat{g}$ possesses an appropriate structure, i.e. $\widehat{g} \equiv g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}$, and that $\widehat{g}$ adopts a polynomial rate of convergence such that $H^2(\widehat{g}, g_0) = O_P(n^{-2\beta})$ for some $\beta > 0$. The precise value of β does not play any role.

For deriving (3.1) and (3.2) it is additionally only required that $\widehat{\tau}, \tau_0 \in [\tau_{min}, \tau_{max}]$, $\widehat{\sigma}, \sigma_0 \in [\sigma_{min}, \sigma_{max}]$ as well as $\widehat{h}, h_0 \in \mathcal{H}_{h_{min},h_{max},\delta}$. Smoothness of h_0 is of no importance. On the other hand, smoothness of h_0 as well as $\text{pen}(\widehat{g}) = O_P(M_n^{-2})$ constitute the further conditions used for deriving (3.3) - (3.5).

For the case with covariates, the following slightly different assumptions need to be imposed:

(B1) This equals (A1).

(B2) $h_{0w_0} \in \mathcal{H}_{h_{min},h_{max},\delta}$ and is twice continuously differentiable, $\tau_0(w_0) \in [\tau_{min}, \tau_{max}]$, and $\sigma_0(w_0) \in [\sigma_{min}, \sigma_{max}]$.

(B3) For some $0 < \beta < 1/5$, $M = M_n \sim (nb_n^d)^\beta$ as n tends to ∞ and the bandwidth b_n satisfies $b_n \to 0$, $nb_n^d \to \infty$ and $b_n M_n = O(1)$.

(B4) For some $A > \sqrt{2}$, $\sup_{\|w-w_0\|_2 \le b_n} P(\log Y < -A(\log n)^{1/2}\sigma_0(w)|W = w) = o((nb_n^d)^{-1})$.

(B5) $g_{0w}(y)$, $\sigma(w)$ and $\tau(w)$ are twice continuously differentiable with respect to w at $w = w_0$ (for all y).

(B6) w_0 is a point in the interior of the support of W. Furthermore, W possesses a density q, and $q(w)$ is continuously differentiable at $w = w_0$.

Theorem 3.3 *Assume (B1)–(B6). Then, if* $\lambda \geq 0$,

$$H^2(\widehat{g}_{w_0}, g_{0w_0}) = O_P(M_n^{-2}),$$

and if $\lambda > 0$,

$$pen(\widehat{g}_{w_0}) = O_P(M_n^{-2}).$$

Theorem 3.4 *Under the assumptions of Theorem 3.3, the conclusions of Theorem 3.2 remain valid (but with* $\widehat{\sigma}$, σ_0, $\ldots$, *replaced by* $\widehat{\sigma}(w_0)$, $\sigma_0(w_0)$, $\ldots$*).*

Adapting notations, the proof of Theorem 3.4 is exactly the same as that of Theorem 3.2.

We end this section with a result about the rate of convergence of our estimators when the true density is more than twice continuously differentiable. From now on, we suppress the dependence on w_0 in the case with covariates, and use the notations $\widehat{\sigma}$, σ_0, $\widehat{\tau}$, τ_0, $\ldots$, both for the case with and without covariates.

Our estimator does not make use of a higher degree of smoothness of h_0. However, if h_0 is m-times continuously differentiable for some $m > 2$ faster (logarithmic) rates of convergence may be achieved by relying on estimators which determine smooth approximations $\widehat{h}$ of h_0. For example, our histogram estimator may be replaced by suitable spline approximations. But as already mentioned in Section 2, determining a spline estimator $\widehat{h}$ as well as $\widehat{\sigma}$ and $\widehat{\tau}$ by maximizing the resulting likelihood seems to be extremely difficult from a computational point of view. But the following theorem shows that any estimation method can be applied which ensures that the corresponding convoluted density $\widehat{g} = g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}$ possesses some polynomial rate of convergence.

Theorem 3.5 *For some* $m = 0, 1, 2, \ldots$ *let* $\mathcal{H}^m_{h_{min},h_{max},h_{m,max},\delta} \subseteq \mathcal{H}_{h_{min},h_{max},\delta}$ *denote a space of m-times continuously differentiable functions with* $\sup_{t\in[0,1]} |h^{(m)}(t)| \leq h_{m,max}$. *Assume that*

1) $h_0 \in \mathcal{H}^m_{h_{min},h_{max},h_{m,max},\delta}$,
2) *there exist estimators* $(\widehat{h}, \widehat{\tau}, \widehat{\sigma}) \in \mathcal{H}^m_{h_{min},h_{max},h_{m,max},\delta} \times [\tau_{min}, \tau_{max}] \times [\sigma_{min}, \sigma_{max}]$ *such that* $\widehat{g} = g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}$ *satisfies*

$$H^2(\widehat{g}, g_0) = O_P(n^{-\kappa}) \quad \textit{for some } \kappa > 0.$$

Then,

$$\widehat{\sigma} - \sigma_0 = O_P\left((\log n)^{-(1+\frac{m}{2})}\right), \tag{3.6}$$

$$\widehat{\tau} - \tau_0 = O_P\left((\log n)^{-\frac{m+1}{2}}\right), \tag{3.7}$$

$$\widehat{h}(1) - h_0(1) = O_P\left((\log n)^{-\frac{m}{2}}\right). \tag{3.8}$$

4 Numerical illustrations

4.1 Some Monte-Carlo experiments for the simplified model

The Monte-Carlo scenario we consider is inspired by the econometric literature on stochastic frontier models, as described in the introduction, Section 1. Mimicking (1.1), we can write

$$Y = \tau \exp(-U)\exp(V), \text{ where } U > 0 \text{ and } V \sim \mathrm{N}(0, \sigma^2).$$

So in our notation (1.2), $X = \tau \exp(-U)$ is the signal and $Z = \exp(V)$ is the noise. Often U is an exponential or a half-normal random variable, as in Aigner et al. (1977) and Meeusen and van den Broek (1977), or a truncated normal random variable, as in Stevenson (1980). In the exponential case the density of U is

$$U \sim \mathrm{Exp}(\beta) \Longleftrightarrow f_U(u) = \beta \exp(-\beta u)\, I(u > 0),$$

where $I(\cdot)$ is the indicator function. Moreover, $\mu_U = \sigma_U = 1/\beta$. For the truncated normal case, the density of U is a normal density with mean α and variance β^2 but truncated at zero:

$$U \sim \mathrm{N}^+(\alpha, \beta^2) \Longleftrightarrow f_U(u) = \frac{\Phi^{-1}(\alpha/\beta)}{\sqrt{2\pi}\beta} \exp\left\{-\frac{1}{2}\left(\frac{u-\alpha}{\beta}\right)^2\right\} I(u > 0).$$

Mean and variance of U are then given by $\mu_U = \alpha + c\beta$ and $\sigma_U^2 = \beta^2(1 - c(\alpha/\beta) - c^2)$ with $c = \phi(\alpha/\beta)/\Phi(\alpha/\beta)$. Here ϕ and Φ represent the density and the cumulative distribution function of a standard normal variable. The very popular half-normal is the particular case where $\alpha = 0$. We will concentrate on four examples. Two exponential cases and two truncated normal cases.

Example 1: Exponential Signal, $U \sim \mathrm{Exp}(\beta)$
The density of X can be written as

$$f(x) = \frac{\beta}{\tau^\beta} x^{\beta-1} I(0 \le x \le \tau).$$

In our simulation study we consider the cases $\beta = 2$ as well as $\beta = 1$. For $\beta = 2$ the density $f(x)$ is linearly increasing from 0 to $2/\tau^2$ on $[0, \tau]$, while for $\beta = 1$ the random variable X is uniform on $[0, \tau]$. In the Monte-Carlo experiments below we tuned the value of σ (size of the noise V) as a factor of σ_U. We choose $\sigma = \rho_{nts}\sigma_U$ with $\rho_{nts} = 0, 0.01, 0.05, 0.25, 0.50, 0.75$.

Example 2: Truncated Normal Signal, $U \sim \mathrm{N}^+(\alpha, \beta^2)$
Here the density of X is given by

$$f(x) = \frac{\Phi^{-1}\big(\alpha/\beta\big)}{\sqrt{2\pi}\beta} \exp\left\{-\frac{1}{2}\left(\frac{\log(\tau) - \log(x) - \alpha}{\beta}\right)^2\right\} x^{-1} I(0 \leq x \leq \tau).$$

Here we first consider the case where U is a half-normal, i.e. the density of U is decreasing from zero: we choose $U \sim \mathrm{N}^+(0, (0.80)^2)$ providing values $E(U) = 0.6383$ and $\sigma_U = 0.4822$ which are not too far from the case of the Exp(2)-distribution. In the second scenario we consider $U \sim \mathrm{N}^+(0.60, (0.60)^2)$. The resulting mean $E(U) = 0.7726$ and standard deviation $\sigma_U = 0.4761$ are of the same order of magnitude as above, but here the density is increasing from zero to $E(U)$ and then decreasing. For, V we follow the same scenario as for the Exponential case with $\sigma = \rho_{nts}\sigma_U$.

In the simulations we fixed arbitrarily the boundary at $\tau = 1$, so that the signal is $X = e^{-U}$. Figure 1 displays the densities of X for the 4 cases considered here. Note that only in the first case of an Exp(2)-distribution the density of X is strictly decreasing from the boundary point. In the two situations with truncated normal distributions the density of X is increasing a little when leaving the boundary point ($\tau = 1$) and then decreases. In the fourth case, the jump of the density at the boundary is rather small and the mode is far from the boundary point. This latter scenario is certainly the most complicated one.

We want to emphasize that large values of ρ_{nts} may result in huge noise to signal ratios in the space of the observations. Table 1 evaluates the ratio σ_Z/σ_X for the four experiments. We see also that in all scenarios the variances of the corresponding signals X are of the same order of magnitude. This facilitates the comparison across the various experiments.

In Tables 2 to 5, we display the results obtained with $MC = 500$ replications of each experiment. In the columns $\log_{10}\lambda$, we indicate the optimal values (given by the Monte-Carlo experiment) obtained over the grid search $\log_{10}\lambda = -4, -3, -2, -1, 0, 1, 2, 3, 4$, where "optimal" is in terms of the sum of the Root Mean Squared Error ($RMSE$) of $\widehat{\tau}$ and of $\widehat{\sigma}$.[1] These are not the optimal values for estimating τ and σ separately, the individual optimal values may in some cases be different from the values reported in the table by an order 10 or 10^2, but globally the results are rather stable in terms of the $RMSE$.

For the number of bins we used the rule $M = \max\big(3, c \times \mathrm{round}(n^{1/5})\big)$ where round(a) is the nearest integer to a. We fix $c = 2$. Note that we obtained very similar results in

[1] Some previous pilot experiments showed indeed that finer grids for the values of λ were not necessary, the results being rather stable to small changes in λ. We choose this "rough" grid for limiting the numerical burden in the Monte-Carlo experiments.

some pilot experiments with $c = 3$ (and even with $c = 1$ but here the number of bins was very small). For the selected sample sizes $n = 50, 100, 500$ this rule of thumb gives $M = 4, 5$ and 7, respectively.

When we look to the 4 tables, we first see that our estimators behave rather well for reasonable sample sizes and not too much noise. Looking through each table, we also see that the performances behave as expected: horizontally, when the sample size increases we improve the performance of the estimators for both τ and σ. Vertically, we can investigate the effect of increasing the size of the noise. We see that when increasing the noise from $\rho_{nts} = 0$ to 0.50 the performance deteriorates. This effect is stronger for estimating τ than for estimating σ, in particular for large samples.

We also note that an increasing density for X from 0 to τ (Table 2) gives better results than the others. The most difficult case (small jump and mode far from the boundary) is reported in Table 5. As is to be expected, the performance is less good but still quite reasonable. We also note that when $\rho_{nts} = 0.50$ or 0.75, in some cases, the performance seems to be quite similar. This might be due to the "rough" grid we used for selecting the optimal λ.

To summarize, we can conclude that the procedure for estimating τ and σ works pretty well even for moderate sample sizes. In applications in economics, where X will be output, U firm inefficiencies and V the noise, we may expect the size of the noise to be relatively small with respect to the size of the signal. In this case, even with small samples we may expect good behavior of our estimators. The suggested rule of thumb for selecting the number of bins also seems to be a good choice in our experiments. Finally, the selection of the penalty parameter λ seems not to be crucial. In practice, with a real sample, we suggest to use a bootstrap procedure to estimate the $RMSE$ of the estimators as a tool for selecting λ. In the next subsection we give an example of such algorithm.

4.2 Estimation of a production frontier: Monte-Carlo experiments

We investigate now the performance of the procedure described in Section 2.2 to estimate a production frontier, i.e., estimation of a boundary in the presence of covariates. We will compare our estimator of the frontier function with the one suggested in Hall and Simar (2002) (henceforth HS), by using the same scenario used by HS, which is a noisy version of the setup described in Gijbels et al. (1999). We have

$$Y = \tau(W) \exp(-U) \exp(V),$$

where $\tau(w) = w^{1/2}$, $U \sim \text{Exp}(3)$, $V \sim \text{N}(0, (0.0667)^2)$ and $W \sim \text{U}[0,1]$, with the random variables W, U and V independent in this scenario. Note that $\rho_{nts} = \sigma_V/\sigma_U = 0.20$. Figure 2 depicts a typical sample and the corresponding estimates on a grid of 11 values of w. We can see a very good fit even for a small sample size of $n = 100$. For each value of w, the value of λ was chosen by the following bootstrap algorithm. At each given w, we draw bootstrap random samples of size n_b, $(Y_1^{*,m}, \ldots, Y_{n_b}^{*,m})$, for $m = 1, \ldots, B$, by sampling with replacement from the n_b values Y_j in the original sample such that $\|W_j - w\|_2 \leq b$. Next, for a given value of λ, we compute the original estimator $\widehat{\tau}_\lambda(w)$ and its bootstrap analogue $\widehat{\tau}_\lambda^{*,m}(w)$, for $m = 1, \ldots, B$. Then several measures could be used to select the optimal λ over a grid of values. The results below were obtained by using the bootstrap estimate of the relative root mean squared error of $\tau(w)$:

$$\text{crit}(\lambda) = \sqrt{\frac{1}{B}\sum_{m=1}^{B}\left(\widehat{\tau}_\lambda^{*,m}(w) - \widehat{\tau}_\lambda(w)\right)^2} / \widehat{\tau}_\lambda(w).$$

This criterion was evaluated over a grid of 5 values in the log-scale of λ, $\log_{10}\lambda = -2, -1, 0, 1, 2$ with $B = 200$. Other measures have been used (like the simple $RMSE$, etc,...) giving qualitatively the same results.

Table 6 presents the results of Monte-Carlo experiments where the bias and the MSE of the estimators are calculated using 500 Monte-Carlo replications. For the estimation of the frontier levels, we can compare with the results in Table 10 in Hall and Simar (2002). We noticed through the experiments that the choice of λ does not seem to be so crucial, and we used our bootstrap algorithm to select its value. To summarize the results of Table 6, we see that for $\tau(w)$ the MSE of our new estimator behaves much better than in HS: for $n = 100$ our MSE is 30% of the MSE in HS (50% when $w = 0.75$) and for $n = 500$ the MSE is less than 50% of the MSE in HS for all cases. For the estimation of $\sigma(w)$ the MSE seems to be good, but no comparison with HS is possible.

4.3 American electricity utility companies

As in Hall and Simar (2002), we now consider an empirical example coming from Christensen and Greene (1976) concerning 123 American electricity utility companies. We use, as in HS, only the variables $Y = \log Q$ and $W = \log C$, where Q is the production output and C is the total cost involved in the production. Figure 5 displays the data along with point-wise estimates of the production frontier $\tau(w)$, over a selected grid of 21 values of w. The bandwidth taken is, as in HS, $b = 0.51$. The figure also shows a smooth estimator of the frontier obtained by running a kernel smoother (quartic kernel with same bandwidth

0.51) through the estimated boundary points. As in HS, to avoid edge effects we restrict the estimation to $w \in [1.5, 5]$.

Comparing with Figure 3 in HS, we observe that the estimation of the frontier function is much smoother than in HS, where important jumps appear in their Figure 3 for their estimate of the boundary points. Here we see that some data points are above the estimated frontier, which was not the case in HS. This is not a surprise since their setup is based on the assumption that the noise is converging to zero when the sample size increases. Here the model allows for large noise levels.

Note also that in our procedure we have much more information on the production process, because we have at each value of w an estimate of the standard deviation $\sigma(w)$ of the noise and an estimate of the conditional expected inefficiency level $E(U|W = w)$. The results are shown in Table 7. It appears that the noise is rather stable over the various levels of the cost, but that the mean of the inefficiency level seems to decrease when the production cost increases. Note that the table also provides the number of points $n_b(w)$ selected around w by the bandwidth b.

5 Appendix

Proof of Theorem 3.2. We first consider Assertion a) and take $\lambda \geq 0$. By assumption $H^2(\widehat{g}, g_0) = O_P(M_n^{-2}) = O_P(n^{-\kappa})$ for $\kappa = 2\beta$. Recall that $g_{h,\tau,\sigma}(y) = \frac{1}{\sigma y}\int_0^1 h(t)\phi\left(\frac{1}{\sigma}\log\frac{y}{t\tau}\right)dt$ for all possible values $(h, \tau, \sigma) \in \mathcal{H}_{h_{min},h_{max},\delta} \times [\tau_{min}, \tau_{max}] \times [\sigma_{min}, \sigma_{max}]$. By definition of ϕ we have

$$g_{h,\tau,\sigma}(y) = w(y, \tau, \sigma) \cdot v(y, h, \tau, \sigma), \tag{5.1}$$

where $w(\cdot)$ and $v(\cdot)$ are defined by

$$w(y, \tau, \sigma) := \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma y}\exp\left(-\frac{(\log y)^2}{2\sigma^2} + \frac{(\log y)(\log \tau)}{\sigma^2}\right), \tag{5.2}$$

$$v(y, h, \tau, \sigma) := \int_0^1 h(t)\exp\left(\frac{(\log y)(\log t)}{\sigma^2}\right)\exp\left(-\frac{(\log t\tau)^2}{2\sigma^2}\right)dt. \tag{5.3}$$

Let $0 < z_{min} < z_{max} < \frac{\kappa\sigma_{min}^2}{\sigma_0^2}$, and for $z \in [z_{min}, z_{max}]$ set

$$y_{z,n} := \exp\left(\left(2\sigma_0^2 z \log n\right)^{1/2}\right).$$

Obviously, $\exp\left(\frac{-(\log(t\tau))^2}{2\sigma^2}\right)$ is bounded above and below by $\exp\left(\frac{-(\log((1-\delta)\tau))^2}{2\sigma^2}\right)$ and $\exp\left(-\frac{(\log\tau)^2}{2\sigma^2}\right)$ for all $1 - \delta \leq t \leq 1$. This implies that there exist constants $0 < C_0 <$

$C_1 < \infty$ such that (see equations (A.10) - (A.12) in the online supplement for some details on the arguments)

$$\frac{C_0}{(\log n)^{1/2}} \leq v(y_{z,n}, h, \tau, \sigma) \leq \frac{C_1}{(\log n)^{1/2}}, \tag{5.4}$$

for all $z \in [z_{min}, z_{max}]$, all $(h, \tau, \sigma) \in \mathcal{H}_{h_{min},h_{max},\delta} \times [\tau_{min}, \tau_{max}] \times [\sigma_{min}, \sigma_{max}]$, and all sufficiently large n.

Since by construction of z_{min} and z_{max}, we have that $\exp\left(-\frac{(\log y_{z,n})^2}{2\sigma^2}\right) = n^{-\sigma_0^2 z/\sigma^2}$ with $\sup_{z\in[z_{min},z_{max}]} \sup_{\sigma\in[\sigma_{min},\sigma_{max}]} \sigma_0^2 z/\sigma^2 < \kappa$, the definition of w in (5.2) as well as relations (5.1) and (5.4) imply the existence of some $\kappa^* > 0$ such that

$$\sup_{z\in[z_{min},z_{max}]} \sup_{(h,\tau,\sigma)\in\mathcal{H}_{h_{min},h_{max},\delta}\times[\tau_{min},\tau_{max}]\times[\sigma_{min},\sigma_{max}]} \frac{n^{-\kappa}}{g_{h,\tau,\sigma}(y_{z,n})} = O(n^{-\kappa^*}) \quad \text{as } n \to \infty. \tag{5.5}$$

Now, let N_n denote the largest integer with $N_n \leq y_{z_{max},n} - y_{z_{min},n}$. There then exists a unique sequence $z_{min} =: z_0 < z_1 < \cdots < z_{N_n} \leq z_{max}$ such that $y_{z_j,n} - y_{z_{j-1},n} = 1$ for all $j = 1, \ldots, N_n$. Obviously, $N_n \to \infty$ as $n \to \infty$. By assumption $H^2(\widehat{g}, g_0) = O_P(n^{-\kappa})$, and hence $\sum_{j=1}^{N_n} \int_{y_{z_{j-1},n}}^{y_{z_j,n}} (\sqrt{\widehat{g}(y)} - \sqrt{g_0(y)})^2 dy = O_P(n^{-\kappa})$. The mean value theorem implies that for every $j = 1, \ldots, N_n$ there exists some $\tilde{z}_j \in [z_{j-1}, z_j]$ such that

$$\int_{y_{z_{j-1},n}}^{y_{z_j,n}} (\sqrt{\widehat{g}(y)} - \sqrt{g_0(y)})^2 dy = (\sqrt{\widehat{g}(y_{\tilde{z}_j,n})} - \sqrt{g_0(y_{\tilde{z}_j,n})})^2.$$

We can infer that $\sum_{\tilde{z}\in\mathcal{Z}_n} (\sqrt{\widehat{g}(y_{\tilde{z},n})} - \sqrt{g_0(y_{\tilde{z},n})})^2 = O_P(n^{-\kappa})$, where $\mathcal{Z}_n = \{\tilde{z}_1, \ldots, \tilde{z}_{N_n}\}$. It follows that $\sup_{\tilde{z}\in\mathcal{Z}_n} (\sqrt{\widehat{g}(y_{\tilde{z},n})} - \sqrt{g_0(y_{\tilde{z},n})})^2 = O_P(n^{-\kappa})$. At the same time $\widehat{g} \equiv g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}$ as well as $g_0 \equiv g_{h_0,\tau_0,\sigma_0}$, and (5.5) thus leads to

$$\sup_{\tilde{z}\in\mathcal{Z}_n} \left(\sqrt{\frac{g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}(y_{\tilde{z},n})}{g_{h_0,\tau_0,\sigma_0}(y_{\tilde{z},n})}} - 1\right)^2 = O_P(n^{-\kappa^*}), \; \sup_{\tilde{z}\in\mathcal{Z}_n} \left(\sqrt{\frac{g_{h_0,\tau_0,\sigma_0}(y_{\tilde{z},n})}{g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}(y_{\tilde{z},n})}} - 1\right)^2 = O_P(n^{-\kappa^*}). \tag{5.6}$$

Together with (5.1) and the definitions of w and v in (5.2) and (5.3) we therefore obtain

$$\begin{aligned}
&\sup_{\tilde{z}\in\mathcal{Z}_n} \left|\log \frac{g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}(y_{\tilde{z},n})}{g_{h_0,\tau_0,\sigma_0}(y_{\tilde{z},n})}\right| \\
&= \sup_{\tilde{z}\in\mathcal{Z}_n} \left| -(\frac{\sigma_0^2}{\widehat{\sigma}^2} - 1)\tilde{z}\log n + \frac{(2\sigma_0^2\tilde{z}\log n)^{1/2}}{\widehat{\sigma}^2}\log\widehat{\tau} - \frac{(2\sigma_0^2\tilde{z}\log n)^{1/2}}{\sigma_0^2}\log\tau_0 + \log\frac{v(y_{\tilde{z},n},\widehat{h},\widehat{\tau},\widehat{\sigma})}{v(y_{\tilde{z},n},h_0,\tau_0,\sigma_0)}\right| \\
&= O_P(n^{-\kappa^*/2}).
\end{aligned} \tag{5.7}$$

But (5.4) implies that $\sup_{\tilde{z}\in\mathcal{Z}_n} |\log \frac{v(y_{\tilde{z},n},\widehat{h},\widehat{\tau},\widehat{\sigma})}{v(y_{\tilde{z},n},h_0,\tau_0,\sigma_0)}| = O_P(1)$. Consequently,

$$\sup_{\tilde{z}\in\mathcal{Z}_n} \left| -(\frac{\sigma_0^2}{\widehat{\sigma}^2} - 1)\tilde{z}\log n + \frac{(2\sigma_0^2\tilde{z}\log n)^{1/2}}{\widehat{\sigma}^2}\log\widehat{\tau} - \frac{(2\sigma_0^2\tilde{z}\log n)^{1/2}}{\sigma_0^2}\log\tau_0\right| = O_P(1) \tag{5.8}$$

Since $\mathcal{Z}_n$ contains an increasing number of N_n elements, $\widehat{\tau} - \tau_0 = O_P\left((\log n)^{-1/2}\right)$ as well as $\widehat{\sigma} - \sigma_0 = O_P\left((\log n)^{-1}\right)$, and hence the identifiability of σ_0 and τ_0, are immediate consequences of (5.8).

This completes the proof of Assertion a). Note that the above arguments only require $H^2(\widehat{g}, g_0) = O_P(n^{-\kappa})$ and do not at all depend on λ. The proof of Assertion b) is based on an analysis of the structure of $\widehat{h}$ to be obtained under penalized estimation. This then allows a more precise evaluation of the difference $\log \frac{v(y_{\tilde{z},n}, \widehat{h}, \widehat{\tau}, \widehat{\sigma})}{v(y_{\tilde{z},n}, h_0, \tau_0, \sigma_0)}$.

Let $\widehat{\gamma} = \widehat{\gamma}_{\widehat{\tau},\widehat{\sigma}}$ and $p_j := \widehat{\gamma}_{M_n-j} - 2\widehat{\gamma}_{M_n-j+1} + \widehat{\gamma}_{M_n-j+2}$ and recall that by construction of our estimator $\widehat{h}$ we have $\widehat{h}(\frac{M_n-j}{M_n}) = \widehat{\gamma}_{M_n-j}$, $j = 0, 1, \ldots, M_n - 1$, and $\widehat{h}(0) = \widehat{\gamma}_1 = \widehat{h}(\frac{1}{M})$. Obviously, with $\gamma^{(1)} := M_n(\widehat{h}(\frac{M_n-1}{M_n}) - \widehat{h}(1))$ we then obtain $\widehat{h}(\frac{M_n-1}{M_n}) = \gamma^{(1)}\frac{1}{M_n} + \widehat{h}(1)$ as well as

$$\widehat{h}(\frac{M_n - j}{M_n}) = \widehat{h}(1) + \gamma^{(1)}\frac{j}{M_n} + \sum_{k=2}^{j}(j-k+1)p_k, \quad j = 2, \ldots, M_n - 1. \tag{5.9}$$

For all $j = 2, \ldots, M_n - 1$,

$$|\sum_{k=2}^{j}(j-k+1)p_k| \leq (M_n^2 \max_k |p_k|)\frac{\sum_{k=2}^{j}(j-k+1)}{M_n^2} \leq (M_n^2 \max_k |p_k|)\frac{1}{2}(\frac{j}{M_n})^2. \tag{5.10}$$

Furthermore, Theorem 3.1 implies that

$$M_n^2 \max_{2\leq k\leq M_n-1} |p_k| = M_n^2 \mathrm{pen}(\widehat{g}) = O_P(1). \tag{5.11}$$

Let J denote the largest integer such that $\frac{J}{M_n} \leq \delta$. Relations (5.9) and (5.10) then imply that

$$\begin{aligned} |\gamma^{(1)}| &= \frac{|\widehat{h}(\frac{M_n-J}{M_n}) - \widehat{h}(1) - \sum_{k=2}^{J}(J-k+1)p_k|}{J/M_n} \\ &\leq \frac{|h_{max} - h_{min}| + |\sum_{k=2}^{J}(J-k+1)p_k|}{\delta - 1/M_n} = O_P(1). \end{aligned} \tag{5.12}$$

Recall that $\widehat{h}$ is constant between the points $\frac{1}{M_n}, \frac{2}{M_n}, \ldots$. Therefore, there exists a constant $B_2 < \infty$, which can be chosen independently of M_n, such that for all $t \in [0,1]$,

$$\widehat{h}(t) = \widehat{h}(1) - \gamma^{(1)}(t-1) + R_M(t), \quad |R_M(t)| \leq (M_n^2 \max_k |p_k|) \cdot \left(\frac{1}{2}(t-1)^2 + \frac{B_2}{M_n}\right). \tag{5.13}$$

On the other hand, a Taylor expansion of the true function h_0 yields

$$h_0(t) = h_0(1) + h_0'(1)(t-1) + R_2(t), \quad \text{where } R_2(t) \leq \max_{s\in[0,1]} |h_0''(s)| \cdot (t-1)^2 \tag{5.14}$$

Using partial integration, some straightforward calculations show that for each $j = 1, 2, 3, \ldots$ there exist constants $0 < D_{0,j} < D_{1,j} < \infty$ such that

$$\frac{D_{0,j}}{(z \log n)^{(j+1)/2}} \leq \int_0^1 |t-1|^j \exp\left(\frac{(\log y_{z,n})(\log t)}{\sigma^2}\right) \exp\left(\frac{-(\log t\tau)^2}{2\sigma^2}\right) dt$$
$$= \int_0^1 |t-1|^j t^{(2\sigma_0^2 z \log n/\sigma^4)^{1/2}} \exp\left(\frac{-(\log t\tau)^2}{2\sigma^2}\right) dt \leq \frac{D_{1,j}}{(z \log n)^{(j+1)/2}} \tag{5.15}$$

for all $z \in [z_{min}, z_{max}]$, all $(h, \tau, \sigma) \in \mathcal{H}_{h_{min}, h_{max}, \delta} \times [\tau_{min}, \tau_{max}] \times [\sigma_{min}, \sigma_{max}]$, and all sufficiently large n.

Note that the derivatives of $t^{(2\sigma_0^2 z \log n/\sigma^4)^{1/2}} \exp\left(\frac{-(\log t\tau)^2}{2\sigma^2}\right)$ with respect to σ and τ are sums of terms which are of the general form $D_3 \cdot (z \log n)^{1/2} (\log t) t^{(2\sigma_0^2 z \log n/\sigma^4)^{1/2}} \exp\left(\frac{-(\log t\tau)^2}{2\sigma^2}\right)$ and $D_4 (\log t)^j t^{(2\sigma_0^2 z \log n/\sigma^4)^{1/2} - s} \exp\left(\frac{-(\log t\tau)^2}{2\sigma^2}\right)$, where D_3, D_4 are constants, and where $j = 0, 1, 2$ as well as $s = 0, 1$. But for a suitable choice of constants relation (5.15) remains valid when replacing $|t-1|^j$ by $|\log t|^j$ and $t^{(2\sigma_0^2 z \log n/\sigma^4)^{1/2}}$ by $t^{(2\sigma_0^2 z \log n/\sigma^4)^{1/2} - s}$. It then follows from a straightforward Taylor expansion that for all $j = 0, 1, \ldots$ there exist some constants $A_j, A_j^* < \infty$ such that

$$\left| \int_0^1 (t-1)^j t^{(\sigma_0^2 z \log n/\widehat{\sigma}^4)^{1/2}} \exp\left(\frac{-(\log t\widehat{\tau})^2}{2\widehat{\sigma}^2}\right) dt - \int_0^1 (t-1)^j t^{(\sigma_0^2 z \log n/\sigma_0^4)^{1/2}} \exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right) dt \right|$$
$$\leq \frac{A_j}{(z \log n)^{(j+1)/2}} |\widehat{\sigma} - \sigma_0| + \frac{A_j^*}{(z \log n)^{(j+1)/2}} |\widehat{\tau} - \tau_0| \tag{5.16}$$

It has already been shown above that $\widehat{\sigma} - \sigma_0 = O_P\left((\log n)^{-1}\right)$ as well as $\widehat{\tau} - \tau_0 = O_P\left((\log n)^{-1/2}\right)$. By definition of the function $v(\cdot)$, relations (5.4) as well as (5.9) - (5.16) then obviously imply that for $\alpha = \frac{1}{2}$

$$\sup_{\tilde{z} \in \mathcal{Z}_n} |\log \frac{v(y_{\tilde{z},n}, \widehat{h}, \widehat{\tau}, \widehat{\sigma})}{v(y_{\tilde{z},n}, h_0, \tau_0, \sigma_0)}|$$
$$= \sup_{\tilde{z} \in \mathcal{Z}_n} \left| \log \frac{\int_0^1 \left(\widehat{h}(1) - \gamma^{(1)}(t-1)\right) t^{(2\sigma_0^2 \tilde{z} \log n/\widehat{\sigma}^4)^{1/2}} \exp\left(\frac{-(\log t\widehat{\tau})^2}{2\widehat{\sigma}^2}\right) dt}{\int_0^1 (h_0(1) + h_0'(1)(t-1)) t^{(2\sigma_0^2 \tilde{z} \log n/\sigma_0^4)^{1/2}} \exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right) dt} \right| + O_P((\log n)^{-\alpha})$$
$$= \sup_{\tilde{z} \in \mathcal{Z}_n} \left| \frac{\int_0^1 \left(\widehat{h}(1) - h_0(1) - (\gamma^{(1)} + h_0'(1))(t-1)\right) t^{(2\sigma_0^2 \tilde{z} \log n/\sigma_0^4)^{1/2}} \exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right) dt}{\int_0^1 (h_0(1) + h_0'(1)(t-1)) t^{(2\sigma_0^2 \tilde{z} \log n/\sigma_0^4)^{1/2}} \exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right) dt} \right|$$
$$+ O_P((\log n)^{-\alpha}) \tag{5.17}$$

By (5.6) and (5.17) we now obtain the following generalization of (5.8):

$$\begin{aligned}
&\sup_{\tilde z\in\mathcal{Z}_n} |\log \frac{g_{\widehat h,\widehat\tau,\widehat\sigma}(y_{\tilde z,n})}{g_{h_0,\tau_0,\sigma_0}(y_{\tilde z,n})}| \\
&\le \sup_{\tilde z\in\mathcal{Z}_n}\Bigg| -(\frac{\sigma_0^2}{\widehat\sigma^2}-1)\tilde z\log n + \frac{(2\sigma_0^2\tilde z\log n)^{1/2}}{\widehat\sigma^2}\log\widehat\tau - \frac{(2\sigma_0^2\tilde z\log n)^{1/2}}{\sigma_0^2}\log\tau_0 \\
&\qquad + \frac{\int_0^1\left(\widehat h(1)-h_0(1)-(\gamma^{(1)}+h_0'(1))(t-1)\right)t^{(2\sigma_0^2\tilde z\log n/\sigma_0^4)^{1/2}}\exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right)dt}{\int_0^1\left(h_0(1)+h_0'(1)(t-1)\right)t^{(2\sigma_0^2\tilde z\log n/\sigma_0^4)^{1/2}}\exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right)dt}\Bigg| \\
&\quad +O_P((\log n)^{-\alpha}) \\
&= O_P((\log n)^{-\alpha}). \qquad (5.18)
\end{aligned}$$

Here, again $\alpha = \frac{1}{2}$. Since $\mathcal{Z}_n$ contains an increasing number of N_n elements, this already leads to $|\widehat\sigma-\sigma_0| = O_P\left((\log n)^{-1.5}\right)$, $|\widehat\tau-\tau_0| = O_P\left((\log n)^{-1}\right)$, and $|\widehat h(1)-h_0(1)| = O_P\left((\log n)^{-1/2}\right)$.

But $|\widehat\tau-\tau_0| = O_P\left((\log n)^{-1}\right)$ (instead of $|\widehat\tau-\tau_0| = O_P\left((\log n)^{-1/2}\right)$) implies that for $j=0$ the error of the Taylor expansions in (5.16) can even be bounded by $O_P\left((\log n)^{-1.5}\right)$ (instead of $O_P\left((\log n)^{-1}\right)$). Together with (5.9) - (5.15) we can then conclude that (5.17) and (5.18) even hold with $\alpha=1$.

Since $\mathcal{Z}_n$ contains an increasing number of N_n elements, $|\widehat\tau-\tau_0| = O_P\left((\log n)^{-1}\right)$ $|\widehat\sigma-\sigma_0| = O_P\left((\log n)^{-2}\right)$, $|\widehat\tau-\tau_0| = O_P\left((\log n)^{-1.5}\right)$, $|\widehat h(1)-h_0(1)| = O_P\left((\log n)^{-1}\right)$, as well as $|\gamma^{(1)}+h_0'(1)| = O_P\left((\log n)^{-1/2}\right)$ are immediate consequences of (5.15) and (5.18) with $\alpha=1$. □

Acknowledgments. The second and third author acknowledge support from IAP research network nr. P6/03 of the Belgian government (Belgian Science Policy). In addition, the third author acknowledges financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650.

References

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

[2] Aigner, D.J., Lovell, C.A.K. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier models. *Journal of Econometrics*, **6**, 21–37.

[3] Banker, R.D. (1993). Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science*, **39**, 1265–1273.

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

[6] Christensen, L., and Greene, W. H. (1976). Economies of Scale in U.S. Electric Power Generation. *Journal of Political Economy*, 84, 653-667.

[7] Daouia, A., Florens, J.P. and Simar, L. (2010). Frontier estimation and extreme values theory. *Bernoulli*, **16**, 1039–1063.

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

[9] Deprins, D., Simar, L. and Tulkens, H. (1984). Measuring labor inefficiency in post offices. In: *The Performance of Public Enterprises: Concepts and Measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243–267.

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

[11] Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society - Series A*, **120**, 253–281.

[12] Gattoufi, S., Oral, M. and Reisman, A. (2004). Data envelopment analysis literature: a bibliography update (1951-2001). In: *Socio-Economic Planning Sciences*, **38**, 159–229.

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

[14] Gijbels, I., Mammen, E., Park, B.U. and Simar, L. (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association*, **94**, 220–228.

[15] Greene, W.H. (2008). The econometric approach to efficiency analysis. In: *The Measurement of Productive Efficiency*, 2nd Edition, Harold Fried, C.A.Knox Lovell and Shelton Schmidt, editors, Oxford University Press.

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

[17] Jeong, S.-O. (2004). Asymptotic distribution of DEA efficiency scores. *Journal of the Korean Statistical Society*, **33**, 449–458.

[18] Jeong, S.-O. and Park, B. U. (2006). Large sample approximation of the distribution for convex-hull estimators of boundaries. *Scandinavian Journal of Statistics*, **33**, 139–151.

[19] Kneip, A., Park, B. and Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory*, **14**, 783–793.

[20] Kneip, A, Simar, L. and Wilson, P.W. (2008). Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models. *Econometric Theory*, **24**, 1663–1697.

[21] Korostelev, A., Simar, L. and Tsybakov, A. (1995a). Efficient estimation of monotone boundaries. *Annals of Statistics*, **23**, 476–489.

[22] Korostelev, A., Simar, L. and Tsybakov, A. (1995b). On estimation of monotone and convex boundaries. *Pub. Inst. Stat. Univ. Paris*, XXXIX, **1**, 3–18.

[23] Kumbhakar, S.C. , Park, B.U., Simar, L. and Tsionas, E.G. (2007). Nonparametric stochastic frontiers: a local likelihood approach. *Journal of Econometrics*, **137**, 1–27.

[24] Markovitz, H.M. (1959). *Portfolio Selection: Efficient Diversification of Investments*, John Wiley, New York.

[25] Meeusen, W. and van den Broek, J. (1977). Efficiency estimation from Cobb-Douglas production function with composed error. *International Economic Review*, **8**, 435–444.

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

[27] Park, B.U., Jeong, S.-O. and Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges. *Annals of Statistics*, **38**, 1320–1340.

[28] Park, B., Simar, L. and Weiner, C. (2000). The FDH estimator for productivity efficiency scores : asymptotic properties. *Econometric Theory*, **16**, 855–877.

[29] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics*. Springer-Verlag, New York.

[30] Schwarz, M. and Van Bellegem, S. (2009). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters*, **80**, 236-241.

[31] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

[32] Stevenson, R.E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, **13**, 57–66.

Figure 1: *Density of X in the 4 scenarios considered.*

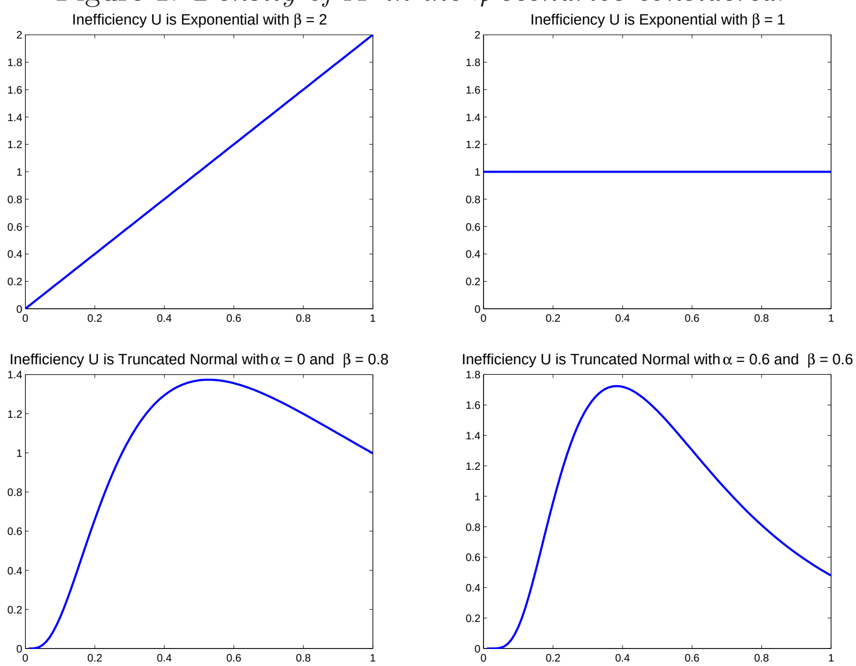

Table 1: *Links between various noise to signal ratios for the 4 scenarios: the table evaluates the ratios σ_Z/σ_X for different values of $\rho_{nts} > 0$.*

Case	σ_X	$\rho_{nts} = 0.01$	$\rho_{nts} = 0.05$	$\rho_{nts} = 0.10$	$\rho_{nts} = 0.25$	$\rho_{nts} = 0.50$	$\rho_{nts} = 0.75$
$U \sim \text{Exp}(2)$	0.2357	0.0212	0.1061	0.2125	0.5365	1.1115	1.7686
$U \sim \text{Exp}(1)$	0.2887	0.0346	0.1735	0.3490	0.9076	2.0921	3.9881
$U \sim \text{N}^+(0, 0.8^2)$	0.2319	0.0208	0.1040	0.2082	0.5252	1.0851	1.7188
$U \sim \text{N}^+(0.6, 0.6^2)$	0.2179	0.0218	0.1093	0.2188	0.5519	1.1391	1.8017

Table 2: *Example 1.a:* $U \sim \text{Exp}(2)$ *with* $\mu_U = \sigma_U = 0.5$

Noise to signal ratios: $\rho_{nts} = 0, \sigma = 0, \sigma_Z/\sigma_X = 0$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0129	0.45e-03	1	0.0066	0.45e-03	1	0.0013	0.08e-03	-3
BIAS	-0.0087	0.41e-03		-0.0042	0.42e-03		-0.0009	0.02e-03	
STD	0.0095	0.18e-03		0.0050	0.17e-03		0.0009	0.08e-03	

Noise to signal ratios: $\rho_{nts} = 0.01, \sigma = 0.005, \sigma_Z/\sigma_X = 0.0212$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0151	0.0118	-1	0.0100	0.0079	-2	0.0035	0.0036	-2
BIAS	-0.0082	0.0083		-0.0055	-0.0010		-0.0010	0.0006	
STD	0.0127	0.0084		0.0084	0.0079		0.0034	0.0036	

Noise to signal ratios: $\rho_{nts} = 0.05, \sigma = 0.025, \sigma_Z/\sigma_X = 0.1061$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0239	0.0279	-1	0.0178	0.0190	-2	0.0079	0.0073	-2
BIAS	-0.0092	0.0112		-0.0019	-0.0054		0.0000	0.0001	
STD	0.0221	0.0255		0.0177	0.0182		0.0079	0.0073	

Noise to signal ratios: $\rho_{nts} = 0.10, \sigma = 0.05, \sigma_Z/\sigma_X = 0.2125$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0312	0.0363	-1	0.0215	0.0250	-1	0.0125	0.0101	-2
BIAS	-0.0054	0.0083		-0.0038	0.0090		-0.0009	0.0008	
STD	0.0308	0.0353		0.0212	0.0234		0.0125	0.0101	

Noise to signal ratios: $\rho_{nts} = 0.25, \sigma = 0.125, \sigma_Z/\sigma_X = 0.5365$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0485	0.0492	-1	0.0352	0.0332	-1	0.0161	0.0144	-1
BIAS	0.0022	0.0039		0.0020	0.0049		0.0055	0.0050	
STD	0.0485	0.0491		0.0351	0.0329		0.0152	0.0136	

Noise to signal ratios: $\rho_{nts} = 0.50, \sigma = 0.250, \sigma_Z/\sigma_X = 1.1115$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0660	0.0554	1	0.0540	0.0432	-1	0.0312	0.0189	-1
BIAS	0.0297	0.0253		0.0134	-0.0019		0.0200	-0.0022	
STD	0.0590	0.0493		0.0523	0.0432		0.0239	0.0188	

Noise to signal ratios: $\rho_{nts} = 0.75, \sigma = 0.375, \sigma_Z/\sigma_X = 1.7686$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0957	0.0627	1	0.0750	0.0544	-1	0.0527	0.0251	-1
BIAS	0.0604	0.0103		0.0250	-0.0090		0.0385	-0.0094	
STD	0.0743	0.0619		0.0708	0.0537		0.0360	0.0233	

Table 3: *Example 1.b:* $U \sim \text{Exp}(1)$ *with* $\mu_U = \sigma_U = 1$

Noise to signal ratios: $\rho_{nts} = 0, \sigma = 0, \sigma_Z/\sigma_X = 0$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0259	0.46e-03	1	0.0138	0.39e-04	-3	0.0026	0.24e-04	-4
BIAS	-0.0180	0.43e-03		-0.0098	0.13e-04		-0.0018	0.08e-04	
STD	0.0187	0.19e-03		0.0098	0.37e-04		0.0019	0.23e-04	

Noise to signal ratios: $\rho_{nts} = 0.01, \sigma = 0.01, \sigma_Z/\sigma_X = 0.0346$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0339	0.0198	-2	0.0228	0.0169	-2	0.0079	0.0072	-2
BIAS	-0.0210	-0.0021		-0.0130	-0.0008		-0.0026	0.0007	
STD	0.0266	0.0197		0.0188	0.0169		0.0075	0.0072	

Noise to signal ratios: $\rho_{nts} = 0.05, \sigma = 0.05, \sigma_Z/\sigma_X = 0.1735$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0535	0.0469	-2	0.0370	0.0350	-2	0.0191	0.0155	-2
BIAS	-0.0196	-0.0143		-0.0067	-0.0121		-0.0013	-0.0020	
STD	0.0498	0.0447		0.0365	0.0328		0.0191	0.0154	

Noise to signal ratios: $\rho_{nts} = 0.10, \sigma = 0.10, \sigma_Z/\sigma_X = 0.3490$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0783	0.0741	-2	0.0581	0.0552	-1	0.0356	0.0242	-2
BIAS	-0.0131	-0.0262		-0.0249	0.0196		0.0002	-0.0040	
STD	0.0773	0.0694		0.0525	0.0517		0.0356	0.0239	

Noise to signal ratios: $\rho_{nts} = 0.25, \sigma = 0.25, \sigma_Z/\sigma_X = 0.9076$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.1280	0.1154	-1	0.0988	0.0840	-1	0.0444	0.0399	-1
BIAS	-0.0205	0.0062		-0.0251	0.0182		-0.0230	0.0239	
STD	0.1264	0.1154		0.0956	0.0821		0.0380	0.0320	

Noise to signal ratios: $\rho_{nts} = 0.50, \sigma = 0.50, \sigma_Z/\sigma_X = 2.0921$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.1141	0.1552	1	0.0919	0.1431	1	0.0627	0.0502	-1
BIAS	-0.0668	0.1110		-0.0628	0.1204		-0.0239	0.0245	
STD	0.0926	0.1085		0.0671	0.0774		0.0580	0.0438	

Noise to signal ratios: $\rho_{nts} = 0.75, \sigma = 0.75, \sigma_Z/\sigma_X = 3.9881$

	$n = 50$			$n = 100$			$n = 500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.1138	0.1697	2	0.0872	0.1495	1	0.0678	0.0549	-1
BIAS	-0.0414	0.1029		-0.0460	0.1153		-0.0197	0.0233	
STD	0.1061	0.1351		0.0742	0.0952		0.0649	0.0498	

Table 4: *Example 2.a:* $U \sim \mathrm{N}^{+}(0, 0.8^2)$ *so* $\mu_U = 0.6383$ *and* $\sigma_U = 0.4822$

Noise to signal ratios: $\rho_{nts} = 0, \sigma = 0, \sigma_Z/\sigma_X = 0$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0259	0.93e-04	-2	0.0127	0.29e-04	-4	0.0025	0.35e-03	0
BIAS	-0.0186	0.34e-04		-0.0090	0.14e-04		-0.0017	0.25e-03	
STD	0.0181	0.87e-04		0.0090	0.26e-04		0.0019	0.25e-03	

Noise to signal ratios: $\rho_{nts} = 0.01, \sigma = 0.0048, \sigma_Z/\sigma_X = 0.0208$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0314	0.0183	0	0.0210	0.0137	-2	0.0091	0.0083	-2
BIAS	-0.0225	0.0181		-0.0130	0.0034		-0.0052	0.0041	
STD	0.0219	0.0020		0.0165	0.0133		0.0074	0.0073	

Noise to signal ratios: $\rho_{nts} = 0.05, \sigma = 0.0241, \sigma_Z/\sigma_X = 0.1040$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0660	0.0519	-2	0.0441	0.0385	-2	0.0186	0.0155	-2
BIAS	-0.0397	0.0133		-0.0227	0.0083		-0.0096	0.0065	
STD	0.0528	0.0503		0.0378	0.0376		0.0159	0.0142	

Noise to signal ratios: $\rho_{nts} = 0.10, \sigma = 0.0482, \sigma_Z/\sigma_X = 0.2082$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0775	0.0640	-3	0.0552	0.0456	-2	0.0309	0.0224	-2
BIAS	-0.0401	0.0089		-0.0255	0.0068		-0.0167	0.0103	
STD	0.0663	0.0635		0.0491	0.0452		0.0260	0.0200	

Noise to signal ratios: $\rho_{nts} = 0.25, \sigma = 0.1206, \sigma_Z/\sigma_X = 0.5252$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.1112	0.0894	-2	0.0999	0.0672	-2	0.0640	0.0349	-2
BIAS	-0.0570	0.0102		-0.0436	0.0126		-0.0377	0.0186	
STD	0.0956	0.0889		0.0900	0.0661		0.0517	0.0295	

Noise to signal ratios: $\rho_{nts} = 0.50, \sigma = 0.2411, \sigma_Z/\sigma_X = 1.0851$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.1109	0.1009	1	0.0941	0.0920	1	0.0915	0.0451	-2
BIAS	-0.0898	0.0830		-0.0808	0.0824		-0.0634	0.0239	
STD	0.0652	0.0575		0.0483	0.0410		0.0660	0.0383	

Noise to signal ratios: $\rho_{nts} = 0.75, \sigma = 0.3617, \sigma_Z/\sigma_X = 1.7188$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0995	0.0856	1	0.0777	0.0716	1	0.0432	0.0531	1
BIAS	-0.0627	0.0554		-0.0529	0.0555		-0.0310	0.0469	
STD	0.0773	0.0653		0.0570	0.0452		0.0301	0.0249	

Table 5: *Example 2.b:* $U \sim \mathrm{N}^{+}(0.6, 0.6^2)$ *so* $\mu_U = 0.7726$ *and* $\sigma_U = 0.4761$

Noise to signal ratios: $\rho_{nts} = 0, \sigma = 0, \sigma_Z/\sigma_X = 0$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0448	0.98e-04	-2	0.0255	0.13e-03	-2	0.0052	0.38e-03	0
BIAS	-0.0312	0.37e-04		-0.0169	0.50e-04		-0.0036	0.29e-03	
STD	0.0322	0.90e-04		0.0191	0.12e-03		0.0038	0.24e-03	

Noise to signal ratios: $\rho_{nts} = 0.01, \sigma = 0.0048, \sigma_Z/\sigma_X = 0.0218$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.0574	0.0180	-1	0.0386	0.0187	-1	0.0222	0.0183	-2
BIAS	-0.0396	0.0151		-0.0294	0.0176		-0.0144	0.0128	
STD	0.0416	0.0098		0.0251	0.0064		0.0170	0.0130	

Noise to signal ratios: $\rho_{nts} = 0.05, \sigma = 0.0238, \sigma_Z/\sigma_X = 0.1093$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.1391	0.1030	-4	0.1038	0.0776	-4	0.0579	0.0473	-2
BIAS	-0.0998	0.0641		-0.0672	0.0428		-0.0410	0.0314	
STD	0.0971	0.0807		0.0792	0.0649		0.0409	0.0354	

Noise to signal ratios: $\rho_{nts} = 0.10, \sigma = 0.0476, \sigma_Z/\sigma_X = 0.2188$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.1592	0.1146	-4	0.1193	0.0843	-3	0.0759	0.0533	-2
BIAS	-0.1160	0.0644		-0.0772	0.0426		-0.0540	0.0343	
STD	0.1091	0.0949		0.0911	0.0728		0.0534	0.0408	

Noise to signal ratios: $\rho_{nts} = 0.25, \sigma = 0.1190, \sigma_Z/\sigma_X = 0.5519$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.1760	0.1138	-4	0.1483	0.0894	-3	0.1111	0.0554	-2
BIAS	-0.1183	0.0351		-0.0959	0.0301		-0.0824	0.0373	
STD	0.1304	0.1083		0.1132	0.0843		0.0747	0.0410	

Noise to signal ratios: $\rho_{nts} = 0.50, \sigma = 0.2381, \sigma_Z/\sigma_X = 1.1391$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.2120	0.1223	-4	0.1854	0.0885	-4	0.1436	0.0498	-3
BIAS	-0.1380	0.0123		-0.1212	0.0267		-0.0842	0.0195	
STD	0.1611	0.1218		0.1404	0.0844		0.1164	0.0459	

Noise to signal ratios: $\rho_{nts} = 0.75, \sigma = 0.3571, \sigma_Z/\sigma_X = 1.8017$

	$n=50$			$n=100$			$n=500$		
	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$	$\widehat{\tau}$	$\widehat{\sigma}$	$\log_{10}\lambda$
RMSE	0.2284	0.1247	-3	0.1812	0.0876	2	0.1465	0.0523	-3
BIAS	-0.1545	-0.0046		-0.1226	0.0711		-0.0841	0.0091	
STD	0.1684	0.1248		0.1336	0.0512		0.1200	0.0515	

Figure 2: *Frontier estimates on a grid of 11 values of w for $n = 100$ (left panel) and $n = 500$ (right panel).*

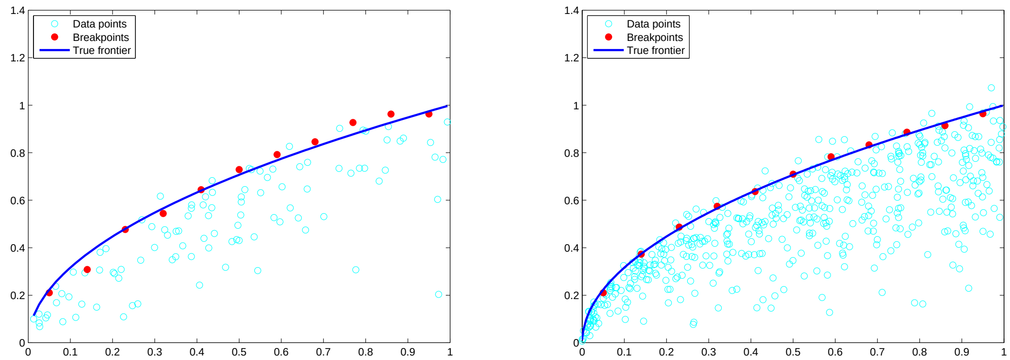

Table 6: *Estimated bias and MSE over 500 Monte-Carlo simulated samples, at 3 different values of w, for $n = 100$ and $n = 500$; HS = Hall-Simar (2002). At each value of w, 200 bootstrap replications are used for selecting λ.*

n				$w = 0.25$	$w = 0.50$	$w = 0.75$
100	$\tau(w)$	new	$BIAS$	.0034	.0046	-.0030
			MSE	.0004	.0008	.0014
		HS	$BIAS$	.0048	.0017	-.0003
			MSE	.0014	.0025	.0028
	$\sigma(w)$	new	$BIAS$	-.0071	-.0031	.0009
			MSE	.0003	.0003	.0003
500	$\tau(w)$	new	$BIAS$	.0083	.0105	.0064
			MSE	.0002	.0004	.0004
		HS	$BIAS$	.0009	.0041	.0041
			MSE	.0005	.0007	.0011
	$\sigma(w)$	new	$BIAS$	-.0053	-.0033	-.0005
			MSE	.0001	.0001	.0001

Figure 3: *American electricity utility data: frontier estimates on a grid of 21 values of* $w = \log(cost)$.

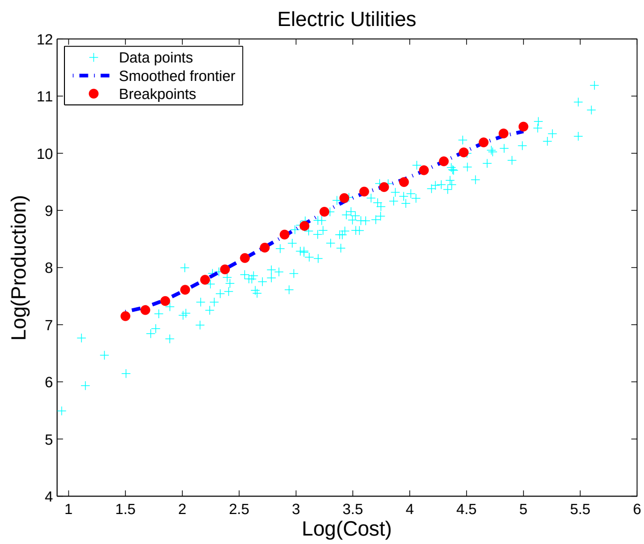

Table 7: *American electricity utility data: estimates of* $\tau(w), \sigma(w)$ *and* $E(U|W = w)$; $n_b(w)$ *indicates the active number of data points used for the local linear approximation.*

values of w	$n_b(w)$	$\widehat{\tau}(w)$	$\widehat{\sigma}(w)$	$\widehat{E}(U\|W = w)$
1.50	10	7.15	0.0342	0.0769
1.68	12	7.26	0.0421	0.0682
1.85	17	7.41	0.0404	0.0607
2.02	19	7.61	0.0398	0.0447
2.20	26	7.79	0.0357	0.0463
2.38	27	7.97	0.0340	0.0472
2.55	28	8.17	0.0257	0.0504
2.72	35	8.35	0.0250	0.0446
2.90	37	8.58	0.0193	0.0504
3.08	42	8.73	0.0258	0.0453
3.25	43	8.98	0.0246	0.0464
3.42	42	9.21	0.0214	0.0482
3.60	38	9.33	0.0226	0.0404
3.78	34	9.41	0.0272	0.0325
3.95	32	9.49	0.0232	0.0255
4.12	28	9.70	0.0250	0.0284
4.30	24	9.86	0.0227	0.0278
4.47	22	10.02	0.0237	0.0300
4.65	21	10.19	0.0201	0.0294
4.83	20	10.35	0.0199	0.0301
5.00	15	10.47	0.0123	0.0278

Online supplement of:

Boundary estimation in the presence of measurement error with unknown variance

Alois Kneip — University of Bonn

Léopold Simar — Université catholique de Louvain

Ingrid Van Keilegom — Université catholique de Louvain

Proof of Theorem 3.1. Let

$$\mathcal{F}_n = \Big\{ y \to g_{h_\gamma,\tau,\sigma}(y) = \frac{1}{\sigma y}\sum_{k=1}^{M_n} \gamma_k \int_{q_{k-1}}^{q_k} \phi\Big(\frac{1}{\sigma}\log\frac{y}{t\tau}\Big)dt : \tau_{min} \le \tau \le \tau_{max},$$

$$\sigma_{min} \le \sigma \le \sigma_{max}, \sum_{k=1}^{M_n} \gamma_k = M_n, 0 \le \gamma_k \le h_{max} \text{ for all } k = 1, \ldots, M_n,$$

$$\text{and } \inf_{1-\delta \le t \le 1} h_\gamma(t) \ge h_{min} \Big\}. \tag{A.1}$$

For any $g(\cdot) = \frac{1}{\sigma \cdot}\int_0^1 h(t)\phi\Big(\frac{1}{\sigma}\log\frac{\cdot}{t\tau}\Big)dt$, define the projection onto $\mathcal{F}_n$ by $\pi_n g(y) = g_{h_\gamma,\tau,\sigma}(y)$, where the vector γ is determined such that $H(g, \pi_n g)$ is minimal. Some easy calculations show that

$$\gamma_k = M_n \int_{q_{k-1}}^{q_k} h(t)dt \tag{A.2}$$

$(k = 1, \ldots, M_n)$, and hence

$$\pi_n g(y) = \frac{M_n}{\sigma y}\sum_{k=1}^{M_n} \int_{q_{k-1}}^{q_k} h(t)dt \int_{q_{k-1}}^{q_k} \phi\Big(\frac{1}{\sigma}\log\frac{y}{t\tau}\Big)dt.$$

Let $q_k^* = (q_{k-1} + q_k)/2$ and let $R_0(y,t) = \frac{\partial}{\partial t}\phi\Big(\frac{1}{\sigma_0}\log\frac{y}{t\tau_0}\Big)$. Consider

$$\begin{aligned}
&\Big|E(\log \pi_n g_0(Y)) - E(\log g_0(Y))\Big| = \Big|\int\Big[\log\frac{\pi_n g_0(y)}{g_0(y)}\Big]g_0(y)\,dy\Big| \\
&\leq \int\Big|\frac{\pi_n g_0(y)}{g_0(y)} - 1\Big|g_0(y)\,dy = \int\Big|\pi_n g_0(y) - g_0(y)\Big|\,dy \\
&\leq \frac{M_n}{\sigma_0}\sum_{k=1}^{M_n}\int\frac{1}{y}\Big|\int_{q_{k-1}}^{q_k}\int_{q_{k-1}}^{q_k}(h_0(s) - h_0(t))\,ds\,\phi\Big(\frac{1}{\sigma_0}\log\frac{y}{t\tau_0}\Big)\,dt\Big|\,dy \\
&\leq \frac{M_n}{\sigma_0}\sum_{k=1}^{M_n}\int\frac{1}{y}\Big|\int_{q_{k-1}}^{q_k}\int_{q_{k-1}}^{q_k}(h_0(s) - h_0(t))\,ds\,\phi\Big(\frac{1}{\sigma_0}\log\frac{y}{q_k^*\tau_0}\Big)\,dt\Big|\,dy \\
&\quad+\frac{M_n}{\sigma_0}\sum_{k=1}^{M_n}\int\frac{1}{y}\Big|\int_{q_{k-1}}^{q_k}\int_{q_{k-1}}^{q_k}(h_0(s) - h_0(t))\,ds\,R_0(y,\eta_k(t))(t - q_k^*)\,dt\Big|\,dy \\
&= \frac{M_n}{\sigma_0}\sum_{k=1}^{M_n}\int\frac{1}{y}\Big|\int_{q_{k-1}}^{q_k}\int_{q_{k-1}}^{q_k}h_0'(q_k^*)(s-t)\,dt\,ds\,\phi\Big(\frac{1}{\sigma_0}\log\frac{y}{q_k^*\tau_0}\Big)\Big|\,dy + O(M_n^{-2}) \\
&= O(M_n^{-2}),
\end{aligned} \tag{A.3}$$

using two Taylor expansions of first order, where $\eta_k(t)$ is between t and q_k^* for any $q_{k-1} < t < q_k$. Next, note that for all $g \in \mathcal{F}_n$, we have that

$$\begin{aligned}
E\Big[\log g(Y) - \log g_0(Y)\Big]^2 &= \int\Big[\log\frac{g(y)}{g_0(y)}\Big]^2 g_0(y)\,dy = 4\int\Big[\log\frac{\sqrt{g(y)}}{\sqrt{g_0(y)}}\Big]^2 g_0(y)\,dy \\
&\leq 4\int\Big[\frac{\sqrt{g(y)}}{\sqrt{g_0(y)}} - 1\Big]^2 g_0(y)\,dy = 4\int[\sqrt{g(y)} - \sqrt{g_0(y)}]^2\,dy \\
&= 4H^2(g, g_0),
\end{aligned}$$

which is uniformly bounded for all $g \in \mathcal{F}_n$. Hence, for all $g \in \mathcal{F}_n$, $\mathrm{Var}(\log g(Y)) \leq D$ for some $D < \infty$, and

$$n^{-1}\sum_{i=1}^{n}\log g(Y_i) - E(\log g(Y)) = O_P(n^{-1/2}).$$

Consider now a subset $\mathcal{F}_n^*$ of $\mathcal{F}_n$ of size $n^{\kappa_2 M_n}$ for some $\kappa_2 > 0$, and let $0 < \kappa_3 < 1/2$. Let $y_{\max} = \exp\{A(\log n)^{1/2}\sigma_0 + \log\tau_0\}$ and $y_{\min} = \exp\{-A(\log n)^{1/2}\sigma_0\}$, where $A > \sqrt{2}$, and define

$$\tilde{g}(y) = \begin{cases} g(y) & y_{\min} \leq y \leq y_{\max} \\ g(y_{\max}) & y > y_{\max} \\ g(y_{\min}) & y < y_{\min}. \end{cases}$$

Then, it is clear that $\max_{g\in\mathcal{F}_n^*}|E(\log g(Y)) - E(\log \tilde{g}(Y))| \leq \frac{C}{2}n^{-\kappa_3}(\log n)^{1/2}$ for some $0 < C < \infty$. Hence,

$$
\begin{aligned}
&P\Big(\max_{g\in\mathcal{F}_n^*}\Big|n^{-1}\sum_{i=1}^n \log g(Y_i) - E(\log g(Y))\Big| \geq Cn^{-\kappa_3}(\log n)^{1/2}\Big)\\
&\leq P\Big(\max_{g\in\mathcal{F}_n^*}\Big|n^{-1}\sum_{i=1}^n \log \tilde{g}(Y_i) - E(\log \tilde{g}(Y))\Big| \geq \frac{C}{2}n^{-\kappa_3}(\log n)^{1/2}\Big)\\
&\quad + P\Big(\max_{1\leq i\leq n}\frac{\log Y_i - \log\tau_0}{\sigma_0} > A(\log n)^{1/2}\Big) + P\Big(\max_{1\leq i\leq n}\frac{-\log Y_i}{\sigma_0} > A(\log n)^{1/2}\Big)\\
&= P_1 + P_2 + P_3 \quad \text{(say)}.
\end{aligned}
$$

Using Bernstein's inequality (see e.g. Serfling (1980), p. 95), we obtain that

$$
\begin{aligned}
P_1 &\leq 2n^{\kappa_2 M_n}\exp\Big(-\frac{1}{4}\frac{C^2 n^{-2\kappa_3}\log n}{2n^{-1}D + Kn^{-1}(\log n)n^{-\kappa_3}(\log n)^{1/2}}\Big)\\
&\leq 2n^{\kappa_2 M_n}\exp\Big(-K'n^{1-2\kappa_3}\log n\Big) \leq 2n^{\kappa_2 M_n}n^{-K'n^{1-2\kappa_3}} = o(1),
\end{aligned}
$$

for some $0 < K, K' < \infty$, provided $M_n = O(n^{1-2\kappa_3})$ and provided K' (and hence C) is chosen sufficiently large. This together with assumption (A3) implies that κ_3 should be chosen at most equal to $(1-\beta)/2$ which is strictly between 2/5 and 1/2 depending on the value of β. Now, note that

$$
\begin{aligned}
P_2 &\leq nP(\log Y > \log\tau_0 + A(\log n)^{1/2}\sigma_0) \leq nP(\log Z/\sigma_0 > A(\log n)^{1/2})\\
&\leq \frac{n}{\sqrt{2\pi}}\frac{1}{A\sqrt{\log n}}\exp\Big\{-\frac{A^2}{2}\log n\Big\} = \frac{1}{A\sqrt{2\pi}}\frac{1}{\sqrt{\log n}}n^{1-\frac{A^2}{2}} = o(1),
\end{aligned}
$$

since $A > \sqrt{2}$. Moreover, from assumption (A4) we know that $P_3 \leq nP(Y < y_{\min}) = o(1)$. Hence,

$$
\max_{g\in\mathcal{F}_n^*}\Big|n^{-1}\sum_{i=1}^n \log g(Y_i) - E(\log g(Y))\Big| = O_P(n^{-\kappa_3}(\log n)^{1/2}). \tag{A.4}
$$

Next, we specify the set $\mathcal{F}_n^*$. Divide the interval $[0, h_{max}]$ into $O(n^{\kappa_2})$ intervals $[\alpha_j, \alpha_{j+1}]$ with $\alpha_j = jn^{-\kappa_2}$ $(j = 0, 1, \ldots, O(n^{\kappa_2}))$. Also, divide $[\tau_{min}, \tau_{max}]$ into $O(n^{\kappa_2})$ intervals $[\tau_i, \tau_{i+1}]$ with $\tau_i = \tau_{min} + in^{-\kappa_2}$ $(i = 0, 1, \ldots, O(n^{\kappa_2}))$ and similarly, divide $[\sigma_{min}, \sigma_{max}]$ into $O(n^{\kappa_2})$ intervals $[\sigma_l, \sigma_{l+1}]$ with $\sigma_l = \sigma_{min} + ln^{-\kappa_2}$, $(l = 0, 1, \ldots, O(n^{\kappa_2}))$. Let

$$
\begin{aligned}
\mathcal{F}_n^* = \Big\{ & y \to g_{h_\gamma,\tau,\sigma}(y) \in \mathcal{F}_n : \text{ there exist } i, l, j_1, \ldots, j_{M_n} \text{ such that}\\
& \tau = \tau_i, \sigma = \sigma_l, \gamma_k = \alpha_{j_k} \text{ for all } k = 1, \ldots, M_n\Big\}.
\end{aligned} \tag{A.5}
$$

Then, it is clear that the number of elements of $\mathcal{F}_n^*$ is $n^{\kappa_2(M_n+2)}$. We will show that

$$
\sup_{g\in\mathcal{F}_n}\inf_{g^*\in\mathcal{F}_n^*}\Big|n^{-1}\sum_{i=1}^n\Big[\log g(Y_i) - \log g^*(Y_i)\Big]\Big| = o_P(n^{-\kappa_4}) \tag{A.6}
$$

for some $\kappa_4 > 0$. In a similar way it can also be shown that

$$\sup_{g\in\mathcal{F}_n} \inf_{g^*\in\mathcal{F}_n^*} \Big| E\Big[\log g(Y) - \log g^*(Y)\Big]\Big| = o(n^{-\kappa_4}). \tag{A.7}$$

To prove (A.6), note that for any $g_{h_\gamma,\tau,\sigma}(\cdot) \in \mathcal{F}_n$, there exist a τ_i, a σ_l, and $\alpha_{j_1}, \ldots, \alpha_{j_{M_n}}$ such that $0 < \tau - \tau_i < n^{-\kappa_2}$, $0 < \sigma - \sigma_l < n^{-\kappa_2}$, $0 < \gamma_k - \alpha_{j_k} < n^{-\kappa_2}$ for all $k = 1, \ldots, M_n$. Denote this element of $\mathcal{F}_n^*$ by g_{h^*,τ^*,σ^*}. Then, for any $y > 0$,

$$\begin{aligned}
&\log g_{h,\tau,\sigma}(y) - \log g_{h^*,\tau^*,\sigma^*}(y) \\
&= \Big[\log w(y,\tau,\sigma) - \log w(y,\tau^*,\sigma^*)\Big] + \Big[\log v(y,h,\tau,\sigma) - \log v(y,h^*,\tau^*,\sigma^*)\Big] \\
&= T_1(y,g,g^*) + T_2(y,g,g^*),
\end{aligned}$$

where

$$w(y,\tau,\sigma) := \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma y} \exp\left(-\frac{(\log y)^2}{2\sigma^2} + \frac{(\log y)(\log \tau)}{\sigma^2} \right), \tag{A.8}$$

$$v(y,h,\tau,\sigma) := \int_0^1 h(t) \exp\left(\frac{(\log y)(\log t)}{\sigma^2} \right) \exp\left(-\frac{(\log t\tau)^2}{2\sigma^2} \right) dt. \tag{A.9}$$

It can be easily shown that $\sup_{g\in\mathcal{F}_n} \inf_{g^*\in\mathcal{F}_n^*} |n^{-1}\sum_{i=1}^n T_1(Y_i,g,g^*)| = o_P(n^{-\kappa_4})$ if $\kappa_4 < \kappa_2$. In fact, write

$$\begin{aligned}
&P\Big(\sup_{g\in\mathcal{F}_n} \inf_{g^*\in\mathcal{F}_n^*} \Big| n^{-1} \sum_{i=1}^n T_1(Y_i,g,g^*)\Big| > n^{-\kappa_4}\Big) \\
&\le P\Big(\min_i \log Y_i < -A(\log n)^{1/2}\sigma_0\Big) + P\Big(\max_i \log Y_i > \log\tau_0 + A(\log n)^{1/2}\sigma_0\Big) \\
&\quad + P\Big(\sup_{g\in\mathcal{F}_n} \inf_{g^*\in\mathcal{F}_n^*} \Big| n^{-1} \sum_{i=1}^n T_1(Y_i,g,g^*) I\Big\{ - A(\log n)^{1/2}\sigma_0 \le \log Y_i \\
&\qquad\qquad \le \log\tau_0 + A(\log n)^{1/2}\sigma_0\Big\}\Big| > \frac{1}{2} n^{-\kappa_4}\Big).
\end{aligned}$$

As before (see the derivation for P_2 and P_3) we have that the first two terms above are $o(1)$. For the third term note that the expression between absolute values is $O_P((\log n)n^{-\kappa_2}) = o_P(n^{-\kappa_4})$ uniformly over g. Next , write

$$\begin{aligned}
&P\Big(\sup_{g\in\mathcal{F}_n} \inf_{g^*\in\mathcal{F}_n^*} \Big| n^{-1} \sum_{i=1}^n T_2(Y_i,g,g^*)\Big| > n^{-\kappa_4}\Big) \\
&\le P\Big(\min_i \log Y_i < -A(\log n)^{1/2}\sigma_0\Big) \\
&\quad + P\Big(\sup_{g\in\mathcal{F}_n} \inf_{g^*\in\mathcal{F}_n^*} \Big| n^{-1} \sum_{i=1}^n T_2(Y_i,g,g^*) I\Big\{ - A(\log n)^{1/2}\sigma_0 \le \log Y_i \le 0\Big\}\Big| > \frac{1}{2} n^{-\kappa_4}\Big) \\
&\quad + P\Big(\sup_{g\in\mathcal{F}_n} \inf_{g^*\in\mathcal{F}_n^*} \Big| n^{-1} \sum_{i=1}^n T_2(Y_i,g,g^*) I\{Y_i > 1\}\Big| > \frac{1}{2} n^{-\kappa_4}\Big) \\
&= S_1 + S_2 + S_3.
\end{aligned}$$

Clearly, by assumption (A4), $S_1 = o(1)$. Next, consider S_2. We focus attention on the term involving $\log v(Y_i, h, \tau, \sigma) - \log v(Y_i, h^*, \tau, \sigma)$, as the terms dealing with $\tau - \tau^*$ and $\sigma - \sigma^*$ can be dealt with in a similar way. It is easy to see that on the interval $[\exp\{-A(\log n)^{1/2}\sigma_0\}, 1]$, the function $y \to \int_0^1 \exp\{\frac{(\log y)(\log t)}{\sigma^2}\}\exp\{-\frac{(\log t\tau)^2}{2\sigma^2}\}dt$ attains its maximum for $y = \exp\{-A(\log n)^{1/2}\sigma_0\}$ and hence $v(y, h, \tau, \sigma) - v(y, h^*, \tau, \sigma)$ is bounded by $Cn^{-\kappa_2}n^{2A^2\sigma_0^2/\sigma_{min}^2}$ for some constant $0 < C < \infty$. By choosing κ_2 large enough, this will be $o(n^{-\kappa_4})$ uniformly on the interval $[\exp\{-A(\log n)^{1/2}\sigma_0\}, 1]$. Since the function $v(y, h, \tau, \sigma)$ is bounded below from zero uniformly over all y in $[\exp\{-A(\log n)^{1/2}\sigma_0\}, 1]$ and all $g_{h,\tau,\sigma} \in \mathcal{F}_n$, it follows that also $\log v(y, h, \tau, \sigma) - \log v(y, h^*, \tau, \sigma)$ is $o(n^{-\kappa_4})$. Hence, $S_2 = 0$ for n large enough.

In order to deal with S_3, define $v_\delta(y, h, \tau, \sigma)$ by integrating over the interval $[1-\delta, 1]$ instead of $[0, 1]$ in (A.9). Let $0 < z_{min} < z_{max} < \kappa\sigma_{min}^2\sigma_0^{-2}$ where $\kappa = 2\beta$ and β is defined in condition (A3), and for $z \in [z_{min}, z_{max}]$ set

$$y_{z,n} := \exp\left(\left(2\sigma_0^2 z \log n\right)^{1/2}\right).$$

Since $h \in \mathcal{H}_{h_{min},h_{max},\delta}$ is bounded by h_{max}, it is now immediately seen that

$$\sup_{z\in[z_{min},z_{max}]}\left|\frac{v(y_{z,n}, h, \tau, \sigma)}{v_\delta(y_{z,n}, h, \tau, \sigma)} - 1\right| \to 0 \quad \text{as } n \to \infty \tag{A.10}$$

uniformly for all $(h, \tau, \sigma) \in \mathcal{H}_{h_{min},h_{max},\delta} \times [\tau_{min}, \tau_{max}] \times [\sigma_{min}, \sigma_{max}]$. Furthermore,

$$\begin{aligned}\int_{1-\delta}^1 h(t)\exp\left(\frac{(\log y_{z,n})(\log t)}{\sigma^2}\right)dt &\geq h_{min}\int_{1-\delta}^1 t^{(2\sigma_0^2 z\log n/\sigma^4)^{1/2}}dt \\ &= \frac{h_{min}(1-(1-\delta)^{(2\sigma_0^2 z\log n/\sigma^4)^{1/2}+1})}{(2\sigma_0^2 z\log n/\sigma^4)^{1/2}+1} \\ &\geq \frac{h_{min}}{(2\sigma_0^2 z\log n/\sigma^4)^{1/2}+1}, \end{aligned}\tag{A.11}$$

and

$$\begin{aligned}\int_0^1 h(t)\exp\left(\frac{(\log y_{z,n})(\log t)}{\sigma^2}\right)dt &\leq h_{max}\int_0^1 t^{(2\sigma_0^2 z\log n/\sigma^4)^{1/2}}dt \\ &\leq \frac{h_{max}}{(2\sigma_0^2 z\log n/\sigma^4)^{1/2}+1}. \end{aligned}\tag{A.12}$$

Now, write

$$\begin{aligned}&n^{-1}\sum_{i=1}^n\Big(\log v(Y_i, h, \tau, \sigma) - \log v(Y_i, h^*, \tau^*, \sigma^*)\Big)I(Y_i > 1) \\ &= n^{-1}\sum_{i=1}^n\frac{I(Y_i > 1)}{v(Y_i, \tilde{h}, \tilde{\tau}, \tilde{\sigma})}\Big([v(Y_i, h, \tau, \sigma) - v(Y_i, h^*, \tau, \sigma)] \\ &\qquad +[v(Y_i, h^*, \tau, \sigma) - v(Y_i, h^*, \tau^*, \sigma)] + [v(Y_i, h^*, \tau^*, \sigma) - v(Y_i, h^*, \tau^*, \sigma^*)]\Big),\end{aligned}$$

for some intermediate $\tilde{h}, \tilde{\tau}$ and $\tilde{\sigma}$. Then, each Y_i in this sum can be written as $Y_i = \exp((2\sigma_0^2 \tilde{Z}_i \log n)^{1/2})$ for some $\tilde{Z}_i > 0$. It is now easily seen using (A.11) and (A.12) that $n^{-1}\sum_{i=1}^n T_2(Y_i, g, g^*) I(Y_i > 1) = O_P(n^{-\kappa_2}) = o_P(n^{-\kappa_4})$ uniformly over g.

It now follows from (A.4), (A.6) and (A.7) that

$$\sup_{g \in \mathcal{F}_n} \left| n^{-1} \sum_{i=1}^n \log g(Y_i) - E(\log g(Y)) \right| = O_P(n^{-\min(\kappa_3, \kappa_4)} (\log n)^{1/2}). \qquad \text{(A.13)}$$

Now, let $\kappa = \min(\kappa_3, \kappa_4)$. Since κ_3 needs to be at most $(1-\beta)/2$ and κ_4 can be any value smaller than κ_2, which on its turn can be chosen as large as needed, it follows that the highest possible value for κ is $\kappa = (1-\beta)/2$.

Next, denoting the γ-vector corresponding to $\pi_n g_0$ by γ_0, and defining the function

$$\tilde{h}_0(s) = M_n \int_{s-1/M_n}^{s} h_0(t)\, dt,$$

it follows from (A.2) that $\gamma_{0k} = \tilde{h}_0(q_k)$ for all $k = 3, \ldots, M_n$, and hence

$$\begin{aligned} \gamma_{0k} - 2\gamma_{0,k-1} + \gamma_{0,k-2} &= M_n^{-1}\tilde{h}_0'(q_k^*) + \frac{1}{2} M_n^{-2}\tilde{h}_0''(\xi_k) - M_n^{-1}\tilde{h}_0'(q_{k-1}^*) - \frac{1}{2} M_n^{-2}\tilde{h}_0''(\xi_{k-1}) \\ &= M_n^{-2}\tilde{h}_0''(\eta_k) + O(M_n^{-2}) = O(M_n^{-2}), \end{aligned}$$

uniformly in k, where $q_k^* = (q_{k-1} + q_k)/2$ as before, and where ξ_k, ξ_{k-1} and η_k are intermediate points. It now follows that $\text{pen}(\pi_n g_0) = O(M_n^{-2})$. Moreover,

$$n^{-1} \sum_{i=1}^n \log \widehat{g}(Y_i) - \lambda\, \text{pen}(\widehat{g}) \geq n^{-1} \sum_{i=1}^n \log g(Y_i) - \lambda\, \text{pen}(g)$$

for any $g \in \mathcal{F}_n$. Now, consider

$$\begin{aligned} & 0 \leq \lambda\, \text{pen}(\widehat{g}) \qquad \text{(A.14)} \\ & \leq E[\log g_0(Y) - \log \widehat{g}(Y)] + \lambda\, \text{pen}(\widehat{g}) \\ & = \Big[n^{-1} \sum_{i=1}^n \log g_0(Y_i) - n^{-1} \sum_{i=1}^n \log \widehat{g}(Y_i) + \lambda\, \text{pen}(\widehat{g}) \Big] \\ & \quad - \Big[n^{-1} \sum_{i=1}^n \log g_0(Y_i) - E(\log g_0(Y)) \Big] + \Big[n^{-1} \sum_{i=1}^n \log \widehat{g}(Y_i) - E(\log \widehat{g}(Y)) \Big] \\ & \leq n^{-1} \sum_{i=1}^n \log g_0(Y_i) - n^{-1} \sum_{i=1}^n \log \pi_n g_0(Y_i) + \lambda\, \text{pen}(\pi_n g_0) + O_P(n^{-\kappa} (\log n)^{1/2}) \\ & \leq E[\log g_0(Y) - \log \pi_n g_0(Y)] + \lambda\, \text{pen}(\pi_n g_0) + O_P(n^{-\kappa} (\log n)^{1/2}) \\ & = O(M_n^{-2}) + O_P(n^{-\kappa} (\log n)^{1/2}) = O_P(M_n^{-2}), \end{aligned}$$

by assumption (A3) and since $\kappa = (1-\beta)/2$. Here, the third and the fourth inequality follow from (A.13) and the first equality in the last line is a consequence of (A.3). This shows that

$$\lambda\,\text{pen}(\widehat{g}) = O_P(M_n^{-2}),$$

and also that

$$E[\log g_0(Y) - \log \widehat{g}(Y)] = O_P(M_n^{-2}).$$

It now follows from Reiss (1989) (p. 99) that

$$H^2(\widehat{g}, g_0) \leq \int \log\Big(\frac{g_0(y)}{\widehat{g}(y)}\Big) g_0(y)\,dy = E[\log g_0(Y) - \log \widehat{g}(Y)] = O_P(M_n^{-2}),$$

which finishes the proof. □

Proof of Theorem 3.3. The proof is similar to the one of Theorem 3.1, and we therefore only focus on the most important differences. Using the abbreviated notations $n_b = \sum_{i=1}^n I(\|W_i - w_0\|_2 \leq b_n)$ and $g_{0i}(\cdot) = g_{0W_i}(\cdot)$, and by ordering the data in such a way that the first n_b observations are the ones for which $\|W_i - w_0\|_2 \leq b_n$, we can write with $\kappa = (1-\beta)/2$

$$\begin{aligned}
0 &\leq \lambda\,\text{pen}(\widehat{g}_{w_0}) \\
&\leq n_b^{-1}\sum_{i=1}^{n_b} E\Big[\log g_{0i}(Y_i) - \log \widehat{g}_{w_0}(Y_i)|W = W_i\Big] + \lambda\,\text{pen}(\widehat{g}_{w_0}) \\
&= \Big[n_b^{-1}\sum_{i=1}^{n_b}\log g_{0i}(Y_i) - n_b^{-1}\sum_{i=1}^{n_b}\log \widehat{g}_{w_0}(Y_i) + \lambda\,\text{pen}(\widehat{g}_{w_0})\Big] \\
&\quad -\Big[n_b^{-1}\sum_{i=1}^{n_b}\Big\{\log g_{0i}(Y_i) - E(\log g_{0i}(Y_i)|W = W_i)\Big\}\Big] \\
&\quad +\Big[n_b^{-1}\sum_{i=1}^{n_b}\Big\{\log \widehat{g}_{w_0}(Y_i) - E(\log \widehat{g}_{w_0}(Y_i)|W = W_i)\Big\}\Big] \\
&\leq \Big[n_b^{-1}\sum_{i=1}^{n_b}\log g_{0w_0}(Y_i) - n_b^{-1}\sum_{i=1}^{n_b}\log \pi_n g_{0w_0}(Y_i) + \lambda\,\text{pen}(\pi_n g_{0w_0})\Big] \\
&\quad +\Big[n_b^{-1}\sum_{i=1}^{n_b}\Big\{\log g_{0i}(Y_i) - \log g_{0w_0}(Y_i)\Big\}\Big] + O_P((nb_n^d)^{-\kappa}(\log n)^{1/2}) \\
&\leq E[\log g_{0w_0}(Y) - \log \pi_n g_{0w_0}(Y)|W = w_0] + \lambda\,\text{pen}(\pi_n g_{0w_0}) \\
&\quad + O_P(b_n^2) + O_P((nb_n^d)^{-\kappa}(\log n)^{1/2}) \\
&= O(M_n^{-2}) + O_P((nb_n^d)^{-\kappa}(\log n)^{1/2}) = O_P(M_n^{-2}).
\end{aligned}$$

The remainder of the proof is similar to the proof of Theorem 3.1. □

Proof of Theorem 3.5. Recall that $\sup_{t\in[0,1]}|h^{(m)}(t)| \leq h_{m,max}$ as well as $\sup_{t\in[0,1]}|h(t)| \leq h_{max}$ for all $h \in \mathcal{H}^m_{h_{min},h_{max},h_{m,max}\delta}$. By using Taylor expansions it is now immediately seen that this implies the existence of a constant H_{max} such that $\sup_{t\in[0,1]}|h^{(j)}(t)| \leq H_{max}$ for all $j = 0, 1, \ldots, m$ and all $h \in \mathcal{H}^m_{h_{min},h_{max},h_{m,max},\delta}$.

The theorem is proved by induction over $m = 0, 1, 2, \ldots$. The arguments used in the proof of Assertion a) of Theorem 3.2 readily generalize to the present situation and already show that (3.6) - (3.8) hold for $m = 0$.

Now consider the case that $m > 0$ and assume that the assertions of the theorem hold for all $m^* = 0, \ldots, m-1$. The proof follows from a generalization of the arguments used in the proof of Assertion b) of Theorem 3.2. We already know that $\widehat{\sigma} - \sigma_0 = O_P\left((\log n)^{-(1+(m-1)/2)}\right)$ as well as $\widehat{\tau} - \tau_0 = O_P\left((\log n)^{-m/2}\right)$. Therefore, (5.16) implies

$$\Bigg|\int_0^1 (t-1)^j t^{(\sigma_0^2 z \log n/\widehat{\sigma}^4)^{1/2}} \exp\left(\frac{-(\log t\widehat{\tau})^2}{2\widehat{\sigma}^2}\right) dt \qquad (A.15)$$
$$- \int_0^1 (t-1)^j t^{(\sigma_0^2 z \log n/\sigma_0^4)^{1/2}} \exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right) dt\Bigg| = O_P((z\log n)^{-(m+1)/2})$$

for all $j = 0, \ldots, m$. Taylor expansions of h_0 and $\widehat{h}$ then provide

$$h_0(t) = h_0(1) + \sum_{j=1}^{m-1} h_0^{(j)}(1)(t-1)^j + R_3(t), \quad \text{where } |R_3(t)| \leq H_{max}\cdot(t-1)^m$$
$$\widehat{h}(t) = \widehat{h}(1) + \sum_{j=1}^{m-1} \widehat{h}^{(j)}(1)(t-1)^j + R_4(t), \quad \text{where } |R_4(t)| \leq H_{max}\cdot(t-1)^m$$

Recall that $\sup_{t\in[0,1]}|h_0^{(j)}(t)| \leq H_{max}$ and $\sup_{t\in[0,1]}|\widehat{h}^{(j)}(t)| \leq H_{max}$ for all $j = 0, \ldots, m$. By (5.15) and (A.15) a straightforward generalization of the arguments leading to (5.17) and (5.18) then yields

$$\sup_{\tilde{z}\in\mathcal{Z}_n} |\log \frac{g_{\widehat{h},\widehat{\tau},\widehat{\sigma}}(y_{\tilde{z},n})}{g_{h_0,\tau_0,\sigma_0}(y_{\tilde{z},n})}|$$
$$= \sup_{\tilde{z}\in\mathcal{Z}_n}\Bigg| -(\frac{\sigma_0^2}{\widehat{\sigma}^2} - 1)\tilde{z}\log n + \frac{(2\sigma_0^2\tilde{z}\log n)^{1/2}}{\widehat{\sigma}^2}\log\widehat{\tau} - \frac{(2\sigma_0^2\tilde{z}\log n)^{1/2}}{\sigma_0^2}\log\tau_0$$
$$+ \frac{\sum_{j=0}^{m-1}(\widehat{h}^{(j)}(1) - h_0^{(j)}(1))\int_0^1 (t-1)^j t^{(2\sigma_0^2\tilde{z}\log n/\sigma_0^4)^{1/2}} \exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right) dt}{\sum_{j=0}^{m-1} h_0^{(j)}(1)\int_0^1 (t-1)^j t^{(2\sigma_0^2\tilde{z}\log n/\sigma_0^4)^{1/2}} \exp\left(\frac{-(\log t\tau_0)^2}{2\sigma_0^2}\right) dt}\Bigg|$$
$$= O_P((\log n)^{-m/2}). \qquad (A.16)$$

Since $\mathcal{Z}_n$ contains an increasing number of N_n elements, $\widehat{\tau} - \tau_0 = O_P\left((\log n)^{-(1+m)/2}\right)$, $\widehat{\sigma} - \sigma_0 = O_P\left((\log n)^{-(1+m/2)}\right)$, and $|\widehat{h}^{(j)}(1) - h_0^{(j)}(1)| = O_P\left((\log n)^{-m/2}\right)$ are immediate consequences of (5.15) and (A.16). □