Ritov, Ya'acov; Härdle, Wolfgang Karl

Working Paper

From animal baits to investors' preference: estimating and demixing of the weight function in semiparametric models for biased samples

SFB 649 Discussion Paper, No. 2007,024

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Ritov, Ya'acov; Härdle, Wolfgang Karl (2007) : From animal baits to investors' preference: estimating and demixing of the weight function in semiparametric models for biased samples, SFB 649 Discussion Paper, No. 2007,024, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/25196

SFB 649 Discussion Paper 2007-024

From Animal Baits to Investors' Preference: Estimating and Demixing of the Weight Function in Semiparametric Models for Biased Samples

Ya'acov Ritov*
Wolfgang Härdle**

* The Hebrew University of Jerusalem, Israel
** Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

From animal baits to investors' preference: Estimating and demixing of the weight function in semiparametric models for biased samples

Ya'acov Ritov*
Department of Statistics
The Hebrew University of Jerusalem

Wolfgang K. Härdle †
CASE - Center for Applied Statistics and Economics
Institute for Statistics and Econometrics
Humboldt-Universität zu Berlin

May 1, 2007

Abstract

We consider two semiparametric models for the weight function in a biased sample model. The object of our interest parametrizes the weight function, and it is either Euclidean or non Euclidean. One of the models discussed in this paper is motivated by the estimation the mixing distribution of individual utility functions in the DAX market.

Key Words and Phrases: Mixture distribution. Inverse problem. Risk aversion. Exponential mixture. Empirical pricing kernel. DAX. Market utility function.

AMS subject classification: 62C05, 62G07, 62G30, 62G08, 62P05, 62P20, G2P12.

JEL subject classifications: C10, C14, D01, D81.

*Partially supported by an ISF grant.

†This research was supported by Deutsche Forschungsgemein schaft through the SFB 649 Economic Risk.

1 Introduction

A sample $X_1, \ldots, X_n$ is considered biased if it is sampled from a density p which is represented as

$$p(x) = \frac{q(x)w(x)}{\int q(u)w(u)\,du}. \tag{1}$$

Here q is some 'natural' pdf (probability density function) for the problem, representing the 'true' underlying distribution, while w is a given weight function that biases the sample. In a standard example, X represents the severity of the disease, and q is the density of X among patients at admission to the hospital. However, it may be more convenient to take a random sample from the population of patients who are in the hospital at a given time. If the time of hospitalization is proportion to the severity of the case, then the sample is taken from the density p, which is equal to q 'length biased' with $w(x) \equiv x$. Vardi (1985) was the first to analyze systematically these type of models. Asymptotic theory was develop in Gill, Vardi and Wellner (1988). Gilbert, Lele, and Vardi (1999) extended the model to the situation where the weight function depends on some parameter, $w(x) = w(x; f)$. The large sample properties were discussed in Gilbert (2000).

Our paper is about estimating f, the parameter of the weight function, $w(x) = w(x; f)$. Two semiparametric models are discussed in this paper. The first model is an experimental design problem in which the concept of biased sample is introduced in a situation where a direct sample is in fact impossible. The model for q is semiparametric, while the parameter of interest, f, is Euclidean. In the second model we consider, q is taken as known, while the weight function is parametrized by a non-Euclidean parameter. This brings us to an inverse problem of estimating and demixing the weight.

The purpose of the first model is to design an experiment in order to evaluate the relative attractiveness of different baits. See for example Ji, Veitch, and Craig (1999). It can be described as follows. Suppose that traps are distributed with different baits. We do not know much about the animal population. We do not know even how many animals visited each trap. We know only the number and characteristics of the animals who were tempted by each of the baits and were captured. For simplicity we assume that each trap and bait can attract and catch unlimited number of animals. We want to estimate the attractiveness of each type of bait as a function of

the animal characteristics, and do that from the sample in which a datum is a pair of an animal and a bait.

The second model is motivated by research on risk aversion and proclivity, more precisely on the empirical pricing kernel (EPK). See Detlefsen, Härdle and Moro (2007) (hereafter DHM). The EPK describes the apparent utility behavior as function of individual investors utility function.

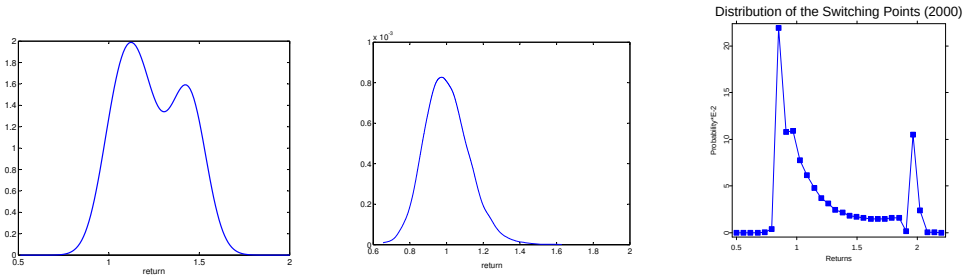
In this model q is the risk neutral density of assets pricing. It is derived from theoretical considerations. The density p on the other hand is the empirical (historical) prices. See Figure 1 (a) and (b) for an example. In asset pricing the EPK links risk neutral investor's behavior to individual utilities, which gives in our notation a semiparametric modeling of the weight function w. The integral function of the pricing kernel q/p is the utility function used by a representing individual. Knowing p and q yields the exact form of the utility function, cf. Ait-Sahalia and Lo (2000). The risk neutral (state price) density (SPD) q can be calculated from the market data on European options, in the DAX data-set. There are more than 5000 observations each day for maturity from 1 week to 2 years. The SPD can therefore be estimated very precisely. Much empirical research work has demonstrated the so called EPK paradox: The resulting utility function is partially concave and partially convex, more precisely of the Friedman and Savage type. See Friedman and Savage (1948).

It is assumed in DHM that the observed density of the DAX value has density of the form $p(x) = cq(x)w(x; f)$ where $q \in \{q_\nu, \nu \in N \subseteq \mathbb{R}^d\}$ is theoretical derived risk neutral density, assumed to follow a given parametric function and c is a normalization factor. That is, of the type of equation (1). The weight function is theoretically derived to be given by $1/U'(x)$, where U is the market utility function, and prime denotes derivative. The market utility function itself is assumed to be a function of the mixture of the individual investors:

$$x = U^{-1}(u) = \int g(u; \xi) f(\xi)\, d\xi.$$

Here ξ denotes an investor type, f is the density of the investors' distribution, and the function $g(\cdot; \cdot)$ is the inverse utility function and it is considered known. A subject of type ξ has the inverse utility function $g(\cdot; \xi)$, or equivalently, the utility function $u(\cdot; \xi)$ satisfying $g\{u(x; \xi); \xi\} \equiv x$. The problem we consider in this paper is to find the density f. We obtain the representation:

$$p(x) = cq(x) \int \frac{\partial}{\partial u} g(u; \xi) f(\xi)\, d\xi, \quad \text{with } x = \int g(u; \xi) f(\xi)\, d\xi. \qquad (2)$$



(a) (b) (c)

Figure 1: The DAX data, 24/03/2000 half a year look ahead: (a) p, the historical density; (b) q, the risk neutral density; (c) The estimate of f, the mixing density. Figures are taken from DHM

See Figure 1 for an example taken from DHM of estimates of p, q, and f.

We will investigate the estimation rate for a few utility functions. The result is typical for inverse problems, in that slightly different assumption will yield completely different results. In fact, we will present three similar models, similar to those investigated in DHM, that exhibit three different type of behavior:

(i) There is no consistent estimator of f;

(ii) f can be estimated at a regular nonparametric rate of $n^{-\alpha}$;

(iii) f can be estimated but at a very slow rate.

We will also show a sort of uncertainty principle. The better we can estimate the function $U^{-1}(u)$, the worse we can demix it and estimate f . This is reasonable. We cannot estimate f well, when large differences in f have only minor impact on $\int g(\cdot;\xi)f(\xi)\,d\xi$.

There are two unified themes to the paper. The first is the usefulness of the concept of biased sample of unknown weight function, even in models where this point of view is not automatic. The second is the technical analysis of inverse problem which starts with a naive estimator that not much is known about it, and then improving it. The structure of the rest of the paper is as follows. In Section 2 we consider a parametric model for the weight function, and show the simplicity of the result. Then, in Section 3 , we suggest an algorithm for calculating the generalized maximum-likelihood estimator (GMLE) for the semiparametric weight function of the

model suggested by DHM. Finally rates of convergence of the demixing estimator for the DHM's model are discussed in Section 4 , as well as of estimates of the mixture itself.

2 The parametric bait problem

Animals of type z are trapped by a bait of type u with probability which is proportional to $w(z, u; f)$. We do know about the animals only after they are caught by the traps. Our aim is to design an experiment and to estimate f. The solution in a nutshell is to distribute many different baits and compare the difference between the animals captured by the different baits, somewhat similar to what is done in case-control studies.

Let $X = (U, Z)$ where U denotes the bait property (e.g., percentage of fat) and Z is a vector of measurement for each animal (e.g., gender, weight, and age). Baits with known (designed!) distribution g are spread, and are encountered by the animals. We assume therefore that the data is sampled from the joint density given by

$$p(u, z) = \frac{w(u, z; f)g(u)h(z)}{\int\int w(\tilde{u}, \tilde{z}; f)g(\tilde{u})h(\tilde{z})\, \underset{\cdot}{\mu}\, (\tilde{u})\, d\nu(\tilde{z})} \tag{3}$$

with respect to the product measure $\mu \times \nu$, and where $w(u, z; f)$ is some parametric weight function, for example, if u is a scalar and $z \in \mathbb{R}^d$, we can consider

$$w(u, z; f) = \exp\{uf^{\mathsf{T}}z\}. \tag{4}$$

Our aim is to estimate f. The density h is an unknown distribution over $\mathbb{R}^d$, d large.

Suppose we consider the weight function (4), then the likelihood equation for f is

$$0 = \sum U_i Z_i - n\frac{\int\int uze^{uf^{\mathsf{T}}z}g(u)h(z)\, d\mu(u)\, d\nu(z)}{\int\int e^{uf^{\mathsf{T}}z}g(u)h(z)\, d\mu(u)\, d\nu(z)}$$

Assuming that Z has at least one Lebesgue continuous component, the GMLE (generalized maximum likelihood estimator) of its distribution is discrete with a point mass at each observation, so that we obtain

$$0 = \sum_{i=1}^{n} U_i Z_i - n\frac{\sum_j \int uz_j e^{uf^{\mathsf{T}}z_j}g(u)\hat{h}_j\, d\mu(u)}{\sum_j \int e^{uf^{\mathsf{T}}z_j}g(u)\hat{h}_j\, d\mu(u)}.$$

where $\hat{h}_j$ is the estimated mass at the point Z_j. However, the likelihood equation for h is

$$0 = \frac{1}{\hat{h}_j} - n\frac{\int e^{uf^{\mathsf{T}}Z_j} g(u)\, d\mu(u)}{\sum_{i=1}^n \int e^{uf^{\mathsf{T}}Z_i} g(u)\, d\mu(u)\hat{h}_i}$$

Plugging this into the profiled log-likelihood equation for f we obtain

$$\begin{aligned} 0 = \frac{1}{n}\dot{\ell}(f) &= \frac{1}{n}\sum_{i=1}^n U_i Z_i - \sum_{j=1}^n \frac{\int u z_j e^{uf^{\mathsf{T}}z_j} g(u)\, d\mu(u)}{\int e^{uf^{\mathsf{T}}z_j} g(u)\, d\mu(u)} \\ &= \frac{1}{n}\sum_{i=1}^n Z_i\big\{U_i - \mathrm{E}_f(U|f^{\mathsf{T}}Z_i)\big\} \\ &= \frac{1}{n}\sum_{i=1}^n T(U_i, Z_i; f), \quad \text{say.} \end{aligned}$$

For example, in the simple case where g is uniform on the interval (a, b):

$$\mathrm{E}_f(U|z) = \frac{\int_a^b u e^{uf^{\mathsf{T}}z}\, dz}{\int_a^b e^{uf^{\mathsf{T}}z}\, du} = \frac{be^{bf^{\mathsf{T}}u} - ae^{af^{\mathsf{T}}z}}{e^{bf^{\mathsf{T}}z} - e^{af^{\mathsf{T}}z}} - \frac{1}{f^{\mathsf{T}}z}$$

if $f^{\mathsf{T}}z \neq 0$, and $(a+b)/2$ otherwise.

Generally, the derivative of $\mathrm{E}_f(U|f^{\mathsf{T}}Z_i)$ is $-\mathrm{Var}_f(U|f^{\mathsf{T}}z)ZZ^{\mathsf{T}}$. Hence ℓ is concave in f. This yields that the maximizer of ℓ is simple to find and is asymptotically normal with asymptotic covariance function given by $\mathrm{E}\big\{\mathrm{Var}_f(U|f^{\mathsf{T}}z)ZZ^{\mathsf{T}}\big\}$

3 EPK: Model and an EM estimator

We consider now the EPK problem. We start now from (2) and we assume that q is known. In practice, it is assumed only to belong to some parametric family $\{q_\nu\}$. However, we will deal in the following in rates which are slower than the parametric $\sqrt{n}$ rate, and the estimate of ν is based on much larger sample than the estimates of the rest of the parameters. Therefore, the assumption that ν is known simplifies the discussion without a real impact on the results.

Rewrite (2) as

$$\begin{aligned} &p\Big\{\int g(u;\xi)f(\xi)\, d\mu(\xi)\Big\}\int \frac{\partial}{\partial u} g(u;\xi)f(\xi)\, d\mu(\xi) \\ &\qquad = cq\Big\{\int g(u;\xi)f(\xi)\, d\mu(\xi)\Big\}\Big\{\int \frac{\partial}{\partial u} g(u;\xi)f(\xi)\, d\mu(\xi)\Big\}^2, \end{aligned} \tag{5}$$

where μ is some dominating measure (i.e., the Lebesgue or the counting meausres). Noting that the LHS of (5) is integrated to 1, we can solve for c and obtain

$$p\Big\{\int g(u;\xi)f(\xi)\,d\mu(\xi)\Big\} = \frac{q\big\{\int g(u;\xi)f(\xi)\,d\mu(\xi)\big\}\int \frac{\partial}{\partial u}g(u;\xi)f(\xi)\,d\mu(\xi)}{\int q\big\{\int g(v;\xi)f(\xi)\,d\mu(\xi)\big\}\big\{\int \frac{\partial}{\partial u}g(v;\xi)f(\xi)\,d\mu(\xi)\big\}^2\,dv}$$

The market utility $U(x) = U(x;f)$ is given by

$$x \equiv \int g\big\{U(x;f);\xi\big\}f(\xi)\,d\mu(\xi) \equiv \psi_f\big\{U(x;f)\big\}.$$

We obtain

$$\begin{aligned} p(x) &= \frac{q(x)\int \frac{\partial}{\partial u}g(U(x;f);\xi)f(\xi)\,d\mu(\xi)}{\int q(y)\int \frac{\partial}{\partial u}g(U(y;f);\xi)f(\xi)\,d\mu(\xi)\,dy} \\ &= \frac{q(x)\psi_f'\big\{\psi_f^{-1}(x)\big\}}{\int q(y)\psi_f'\big\{\psi_f^{-1}(y)\big\}\,dy}. \end{aligned} \tag{6}$$

The statistical model assumed by DHM is that we obtain a simple random sample from p, where p is parametrized in (6) by the non-Euclidean parameter f. A natural approach is to estimate f by the MLE or a variant of it, which we develop now. Note that $\nabla_f\psi_f(u) = g(u;\cdot)$, and by taking the gradient of $x \equiv \int g\big\{\psi_f^{-1}(x);\xi\big\}f(\xi)\,d\mu(\xi)$ we obtain

$$0 = g\big\{\psi_f^{-1}(x);\cdot\big\} + \psi_f'\big\{\psi_f^{-1}(x)\big\}\nabla_f\psi_f^{-1}(x).$$

The derivative of the log-likelihood is given therefore by

$$\begin{aligned} \dot\ell_f(\xi) &= \sum_{i=1}^n \frac{1}{\psi_f'\big\{\psi_f^{-1}(X_i)\big\}}\Big[\frac{\partial}{\partial u}g\big\{\psi_f^{-1}(X_i);\xi\big\} - \frac{\psi_f''}{\psi_f'}\big\{\psi_f^{-1}(X_i)\big\}g\big\{\psi_f^{-1}(X_i);\xi\big\}\Big] \\ &\qquad - nA_f(\xi), \quad \xi \in \operatorname{supp} f \\ &= \sum_{i=1}^n \frac{1}{\psi_f'\big\{U_i\big\}}\big\{\frac{\partial}{\partial u}g\big\{U_i;\xi\big\} - \frac{\psi_f''}{\psi_f'}(U_i)g(U_i;\xi)\big\} - nA_f(\xi), \\ &\qquad U_i = \psi_f^{-1}(X_i), \quad \xi \in \operatorname{supp} f, \end{aligned}$$

where $A_f(\xi)$ is the mean of the first term under f. Since the density of U_i is given by

$$r_f(u) = p\big\{\psi_f(u)\big\}\psi_f'(u) = \frac{q\big\{\psi_f(u)\big\}\big\{\psi_f'(u)\big\}^2}{\int q\big\{\psi_f(v)\big\}\big\{\psi_f'(v)\big\}^2\,dv}.$$

We obtain that

$$A_f(\xi) = \frac{\int q\{\psi_f(u)\}\{\psi_f'(u)\frac{\partial}{\partial u}g(u;\xi) - \psi_f''(u)g(u;\xi)\}\,du}{\int q\{\psi_f(v)\}\{\psi_f'(v)\}^2\,dv}.$$

The model of random sample from the density p can be well approximated as $\sigma \to 0$ by a $X_i = \psi_f(U_i) + \varepsilon_i$, $i = 1, \ldots, n$, where $\varepsilon_1, \ldots, \varepsilon_n$ is a random sample from $N(0, \sigma^2)$ independent from the random sample $U_1, \ldots, U_n$ taken from the density r_f. Now, the log-likelihood of the joint density is given by

$$\ell_f = \sum_{i=1}^{n}\Big[\log q\{\psi_f(U_i)\} + 2\log\{\psi_f'(U_i)\}\Big] - nC_f - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \psi_f(U_i))^2,$$

where

$$C_f = \log\int q\{\psi_f(v)\}\{\psi_f'(v)\}^2\,dv.$$

By a well known formula for the Bayes estimator in the Gaussian measurement error model, under the above model, the distribution of $\psi_f(U_i) - X_i$ given X_i is normal with mean $\sigma^2 f_X'(X_i)/f_X(X_i)$ and a second moment equal to $\sigma^4 f_X''(X_i)/f_X(X_i) + \sigma^2$, where f_X is the marginal density of X_i. At the limit as $\sigma^2 \to 0$, the conditional expectation of the log-likelihood given the X_is amounts therefore to replacing U_i by $\psi_f^{-1}(X_i)$. We conclude that the limiting EM algorithm iterates therefore between

$$\text{The E step:} \qquad U_i \leftarrow \psi_f^{-1}(X_i), \quad i = 1, \ldots, n,$$

and

$$\text{The M step: } f \leftarrow \arg\max\left[\sum_{i=1}^{n}\Big\{\log q\{\psi_f(U_i)\} + 2\log\{\psi_f'(U_i)\}\Big\} - nC_f\right].$$

Let $\mathbf{U} = (U_1, \ldots, U_n)$, $\mathbf{X} = (X_1, \ldots, X_n)$, and denote the E-step by $\mathbf{U} = \psi_f^{-1}(\mathbf{X})$. The M-step can be solved by solving the likelihood equation:

$$\begin{aligned} 0 &= \dot{\ell}_f^M(\xi; \mathbf{U}) \\ &= \sum_{i=1}^{n}\left[\frac{q'\{\psi_f(U_i)\}}{q\{\psi_f(U_i)\}}g(U_i;\xi) + \frac{2}{\psi_f'(U_i)}\frac{\partial}{\partial u}g(U_i,\xi) - \dot{C}_f(\xi)\right], \end{aligned} \tag{7}$$

for all $\xi \in \operatorname{supp} f$, where

$$\begin{aligned}\dot{C}_f(\xi) &= \frac{\int \big[\frac{q'\{\psi_f(v)\}}{q\{\psi_f(v)\}} g(v;\xi) + \frac{2}{\psi'_f(v)} \frac{\partial}{\partial u} g(v,\xi)\big] q\{\psi_f(v)\}\{\psi'_f(v)\}^2 \, dv}{\int q\{\psi_f(v)\}\{\psi'_f(v)\}^2 \, dv} \\ &= \mathrm{E}_f\big[\frac{q'\{\psi_f(U)\}}{q\{\psi_f(U)\}} g(U;\xi) + \frac{2}{\psi'_f(U)} \frac{\partial}{\partial u} g(U,\xi)\big] \\ &= \mathrm{E}_f\{T_f(U;\xi)\}, \quad \text{say.}\end{aligned}$$

An approximate M-step (which is enough, since all we need in the M-step is that the likelihood will be strictly increased) is obtained by considering an approximate Newton-Raphson solution of (7), where $\mathcal{O}_p(\sqrt{n})$ terms in the Hessian of the log-likelihood are discarded. That is the term

$$\sum_{i=1}^{n} \{\nabla_f T_f(U_i;\xi) - \mathrm{E}_f \nabla_f T_f(U;\xi)\}.$$

We consider therefore the algorithm:

$$f_{i+1} = f_i + H_{f_i}^{-1} \ell_{f_i}^M \{\cdot; \psi_{f_i}^{-1}(\mathbf{X})\}, \quad i = 1, 2, \ldots,$$

where $H_f : L_2(\mu) \to L_2(\mu)$ is the operator given by:

$$H_f(\xi,\zeta) = \operatorname{cov}_f\{T_f(U;\xi)\ ,\ T_f(U;\zeta)\}.$$

4 EPK: Rates of convergence

In the previous section we considered the MLE estimate of f. In this section we consider simple estimators of the type suggested by DHM. Using these estimators we will be able to discuss possible minimax rates of convergence. In essence, we start with some naive nonparametric estimator and improve it or demix it for f. One simple method for demixing the EPK is to start with (2) which can be written as

$$\begin{aligned}1 &= c \int \frac{\partial}{\partial u} g(u;\xi) f(\xi) \, d\xi \frac{q}{p}\{\int g(u;\xi) f(\xi) \, d\xi\} \\ &= c \frac{\partial}{\partial u} \frac{q}{p}\{\int g(u;\xi) f(\xi) \, d\xi\}.\end{aligned}$$

Hence

$$\frac{q}{p}\{\int g(u;\xi) f(\xi) \, d\xi\} = \alpha + \beta u$$

for some α and β, or

$$\int g(u;\xi)f(\xi)\ d\xi = \left(\frac{p}{q}\right)^{-1}(\alpha+\beta u).$$

The utility function of an individual is defined up to affine transformation. To assure that it is well defined, we assume that that utility at a return of 1 is 0, and its derivative there is 1. In other words we assume that $g(0,\xi) \equiv \frac{\partial}{\partial u}g(0,\xi) \equiv 1$. Hence

$$\alpha = \frac{p(1)}{q(1)}$$
$$\beta = \frac{p'(1)}{q(1)} - \frac{p(1)}{q(1)}\frac{q'(1)}{q(1)}.$$

We therefore want to solve

$$\int g(u;\xi)f(\xi)\,d\xi = \psi(u), \tag{8}$$

for some ψ. Since q is estimated as a parametric density (based on a much larger sample), and p can be estimated at a standard non-parametric rate based on a direct sample from p, ψ can as well be estimated at a regular density estimation rate. The analysis of this section starts with (8). We assume that ψ and its relevant derivatives can be estimated in some polynomial rate $\|\hat{\psi}^{(i)} - \psi^i\|_\infty = \mathcal{O}_p(n^{-\alpha_i})$ for some $\alpha_i > 0$. The natural estimator suggested by DHM is given by the inverse function of a weighed density estimator. Under strict monotonicity and boundness, the inverse function inherits most properties from the density kernel estimator.

Note that model (8) looks like a linear model. For example if f is approximated by a finite distribution with point mass at $\xi_1,\ldots,\xi_m$, and we consider the equation at the k points $u_1,\ldots,u_k$ then we can write (8) as

$$\hat{\psi}(u_i) = \sum_{j=1}^{m}\beta_j g(u_i;\xi_j) + \varepsilon_i, \qquad i = 1,\ldots,k. \tag{9}$$

(9) looks like a standard linear model, and indeed we suggest to estimate f by solving it. However, it is not. Most linear model assumptions are violated. E.g., $\varepsilon_1,\ldots,\varepsilon_k$ are not i.i.d. , they are not independent of the random $u_1,\ldots,u_k$, which are in fact an estimated function of the observed values $x_1,\ldots,x_k$.

The basic idea of this section is as follow. We assume that we have some naive nonparametric estimator of ψ. We then proceed to use the pseudo

linear model (9) to to estimate the mixing distribution and to improve the estimate of ψ itself. We show that this method yields the minimax rates.

How fast can f be estimated? In the rest of the section we present simple examples following DHM. These examples show that in a very similar models very different type of behavior can be obtained. It can be that (*i*) There is no consistent estimator of f; (*ii*) f can be estimated at a regular nonparametric rate of $n^{-\alpha}$; (*iii*) f can be estimated but at a very slow rate. Thus one can suspect that any optimistic result of demixing depends too heavily on assumptions, and are *a priori* not robust (at least in the minimax sense). In particular, any result should be checked to stand against different changes in the model.

4.1 Switching between two utilities

Following DHM assume that for $x, \xi > 0$:

$$U(x;\xi) = \alpha_2(1-c)^{1-1/\alpha_2}\{[x-\xi]_+^{1/\alpha_1} \vee (x-c)^{1/\alpha_2}\} - \alpha_2(1-c), \qquad (10)$$

where $\alpha_2 > \alpha_1 > 1$ are given, $c < 0$, and $[x]_+ = x\mathbf{1}(x > 0)$. See Figure 2 . Then

$$g(u;\xi) = \min\{\beta^{\alpha_2}\{u + \alpha_2(1-c)\}^{\alpha_2} + c,\ \beta^{\alpha_1}\{u + \alpha_2(1-c)\}^{\alpha_1} + \xi\},$$

where $\beta = \alpha_2^{-1}(1-c)^{-1+1/\alpha_2}$. To simplify and generalize the discussion, we consider the slightly more general case:

$$g(u;\xi) = \begin{cases} g_2(u) & -\infty < u \le h(\xi) \\ g_1(u) + \xi & \infty > u > h(\xi) \end{cases}, \quad \xi > 0,$$

where

$$h^{-1} = g_2 - g_1 \qquad (11)$$

is a strictly increasing function. Note that $g(u;\xi)$ is continuous in ξ. Then (8) is translated to

$$\psi(u) = \int^{h^{-1}(u)} \xi f(\xi)\, d\xi + g_2(u)F\big\{h^{-1}(u)\big\} + g_2(u)\Big\{1 - F\big\{h^{-1}(u)\big\}\Big\},$$

where F is the cdf corresponding to the pdf f. Changing variables and considering (11)

$$\psi\big\{h(s)\big\} = \int^{s} \xi f(\xi)\, d\xi - sF(s) + g_2\big\{h(s)\big\}.$$

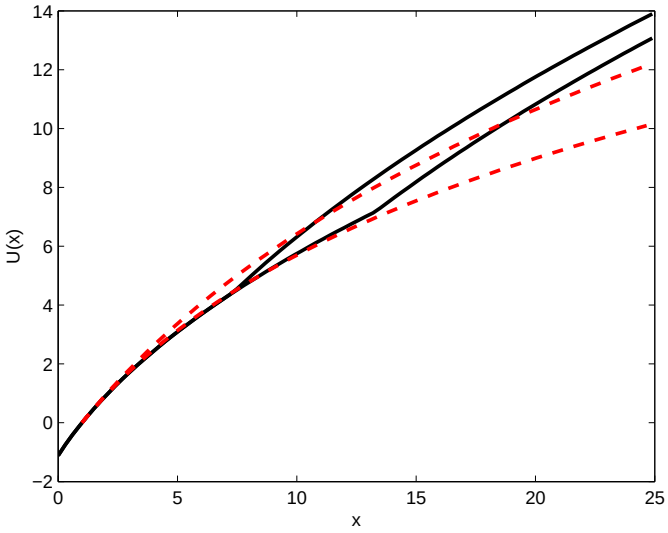


Figure 2: The utility function $U(\cdot;\xi)$ of (10) ($\alpha_1 = 2$, $\alpha_2 = 2.25$, $c = 2$) for two different values of ξ (solid lines), and of (12) for two values (broken lines).

Taking derivative:

$$F(s) = h'(s)\Big\{g_2'\big\{h(s)\big\} - \psi'\big\{h(s)\big\}\Big\}.$$

Hence estimating F at s is equivalent to the estimation of ψ' at $h(s)$. In other words, $f(\cdot)$ can be estimated at the same rate as the rate of the estimation of second derivative of ψ, which in turn is essentially governed by the rate of estimation of the second derivative of p, which depends on the level of smoothness assumption we are willing to accept. Thus if we assume that p has s bounded derivatives, then f can be estimated with an $\mathcal{O}_p(n^{-(s-2)/(2s+1)})$ error. See Silverman (1986) for a general review of density estimation.

4.2 Polynomial and exponential inverse utility function

The previous example was a relatively optimistic example. However, modest changes in the inverse utility function may create situations in which f can hardly estimated, or even not at all.

Suppose a CRRA (constant relative risk aversion) utility:

$$g(u;\zeta) = (\alpha\zeta^{\alpha-1})^{-1}\left\{(u+\zeta)^{\alpha} - \zeta^{\alpha}\right\} + 1, \quad u \in \mathbb{R}, \zeta \in \mathbb{R}^{+},$$

where α is known. Note that g is scaled such that both its value and derivative at 0 are equal to 1. That is, we consider only one branch of (10). If α is an integer, then $\psi(\cdot)$ is a function of only the first α moments of f, and hence there is no consistent estimator of f.

Seemingly, more and more moments are revealed as $\alpha \to \infty$. However, it is not clear that they could be estimated effectively. The limiting form of the inverse utility function, as $\alpha \to \infty$ and $\alpha/\zeta \to \xi$ is given by

$$g(u;\xi) \equiv \xi^{-1}(e^{u\xi} - 1) + 1. \tag{12}$$

The density f is now identified. For example, all its moments can be estimated, e.g., by $\int \xi^{i} f(\xi)\, d\xi = \psi^{(i+1)}(0)$. We are going now to analyze this model in some detail. We will argue now that if $f(\cdot)$ is assumed to have two bounded derivatives, then its value at a point can indeed be estimated, but this can be done only in a very slow convergence rate, slower than any polynomial rate. To be more exact,we argue

Theorem 4.1 *Assume that g is given by* (12) *and f is bounded and has two bounded derivatives. Suppose the minimax rate of estimation of ψ is n^{γ}, $\gamma \in (0, 1/2)$. Then there is an estimator $\hat{f}$ such that $\hat{f}(s) - f(s) = \mathcal{O}_p(n^{-\alpha \log\log n/\log n})$ for some α, and there is no estimator $\tilde{f}(s)$, such that $\tilde{f}(s) - f(s) = \mathcal{O}_p(n^{-\alpha/\log\log n})$ for some α.*

The proof is given in Appendix A .

4.3 Smoothing the empirical estimate and an uncertainty principle

As in the previous subsections we start with a nonparametric $\hat{\psi}$. The purpose of this subsection is to show that a simple projection of this initial estimator yields a considerably better estimator.

We argued in Subsection 4.2 there that there is no reasonable estimator of f for g given in (12). Is this model useless? The surprising answer is no. Although f cannot be estimated per-se, many of its functionals can be estimated quite easily and quite well. For example, as mentioned, its moments. Similarly $\psi(u)$ considered as a simple linear functional can be

estimated quite easily. Suppose that f is supported on some compact interval $[a, b]$. Then one can approximate

$$\psi(u) = \sum_{i=1}^{m} \beta_i u^i + R_m(u)$$

where for some $\tilde{u} \in (0, u)$:

$$\begin{aligned} 0 \le R_m(u) &= \frac{1}{(m+1)!}\psi^{m+1}(\tilde{u}) \\ &= \frac{1}{(m+1)!}\int_a^b \xi^m e^{\tilde{u}\xi} f(\xi)\, d\xi \\ &\le \frac{b^m e^{ub}}{(m+1)!}. \end{aligned} \tag{13}$$

Generally speaking, the faster the coefficients β are converging to 0, the easier it is to estimate ψ and the harder it is to estimate the mixing density g. As (13) shows, we need only very few terms to approximate ψ quite well. In fact we show that in this smooth case, where on one hand f can be hardly estimated, ψ can be estimated almost at the parametric rate. This is not an accident — these are two faces of one phenomena. The shape of the observable ψ hardly depends on f, and essentially depends only on a few aspects of f. These aspects can be estimated well (and hence ψ can be estimated well too). The other aspects can hardly be estimated and hence f cannot be estimated in a reasonable rate. This yields an uncertainty principle — the more you are certain about ψ the less you are about f.

Recall that a function g is called completely monotone if $(-1)^k g^{(k)} \ge 0$, and it is called Bernstein function if its first derivative is completely monotone. It is well known (Feller, 1996) g is completely monotone if, and only if $g(u) = \int_0^\infty e^{-u\xi}\, dF(\xi)$. In other words, ψ is a Bernstein function. Nonparametric maximum likelihood estimation for exponential mixture (and hence completely monotone density) was discussed in Jewel (1982). Balabdaoui and Wellner (2007) discussed the estimation of a k-monotone density.

We assume that at our disposal there is an estimate $\hat{\psi} = \hat{\psi}_n$. For any $u_1, \ldots, u_k > 0$, let $\Sigma(u_1, \ldots, u_k) \in \mathbb{R}^{k\times k}$, where

$$\Sigma_{ij}(u_1, \ldots, u_k) = \operatorname{cov}\big\{\hat{\psi}(u_i)\,,\, \hat{\psi}(u_j)\big\}.$$

Consider the following assumption:

A1. For any n there is $k = k_n$, and $u_1, \dots, u_k \in (c, d)$, $0 < c < d$, such that the spectral radius of $\Sigma(u_1, \dots, u_k)$ is $\mathcal{O}(k/n)$ and $\max_i |\mathrm{E}\,\psi(u_i) - \psi(u_i)|^2 = \mathcal{O}(\log n/n)$.

Assumption A1 is satisfied by many nonparametric density and regression estimators when they are strictly under-smooth. We care much more about bias than about variance of the original estimator $\hat{\psi}$. Thus, we have in mind a kernel estimator with bandwidth of order $n^{-1/4+\varepsilon}$. The spectral radius is based on the assumptions that the estimator at points that are a multiply of the bandwidth apart are (almost) independent, for example this is trivially the case with kernel estimator with compact support. The relationships in the assumption are derived from assuming that the bias of the estimator is $\mathcal{O}(\sigma^2)$, the variance is $\mathcal{O}(1/n\sigma)$, and $k = \mathcal{O}(\sigma^{-1})$.

Consider now the least squares regression of $Y = \{\hat{\psi}(u_1), \dots, \hat{\psi}(u_k)\}^{\mathsf{T}}$ on the design matrix $Z \in \mathbb{R}^{k \times m}$, $Z_{ij} = u_i^j$. That is $\hat{\beta} = (Z'Z)^{-1}Z'Y$, where $\hat{\beta} \in \mathbb{R}^m$. Finally let $\tilde{\psi}(u) = \sum_{j=1}^m \hat{\beta}_j u^j$, $u > 0$. We argue that the error achieved by $\tilde{\psi}$ is almost the parametric rate even although $\hat{\beta}$ achieves can be estimated in strictly lower rate.

Theorem 4.2 *Suppose $g(u;\xi) \equiv \xi^{-1}(e^{u\xi} - 1)$ and that f is supported on a compact interval. Assume* A1 *holds and $m = m_n = \log n / \log\log n$. Then $k^{-1}\sum_{i=1}^k \{\tilde{\psi}(u_i) - \psi(u_i)\}^2 = \mathcal{O}_p\{(\log n)^2/n\}$.*

Proof. Let β^0 be the true value $\beta_j^0 = \int \xi^{j-1} f(\xi)\, d\xi / j!$. Write $Y = Z\beta + \varepsilon$, where ε include both the random error and the bias terms due to both the estimator and the truncation. The latter term is given in (13). By standard least squares results

$$\begin{aligned} k^{-1}\mathrm{E}\sum_{i=1}^k \{\tilde{\psi}(u_i) - \psi(u_i)\}^2 &= k^{-1}\mathrm{E}\{\varepsilon^{\mathsf{T}} Z (Z^{\mathsf{T}}Z)^{-1} Z^{\mathsf{T}} \varepsilon\} \\ &= k^{-1}\,\mathrm{trace}\{Z(Z^{\mathsf{T}}Z)^{-1}Z^{\mathsf{T}}\mathrm{E}(\varepsilon\varepsilon^{\mathsf{T}})\}. \end{aligned}$$

Since $Z(Z^{\mathsf{T}}Z)^{-1}Z^{\mathsf{T}}$ is a projection matrix on a m-dimensional space, the RHS is bounded by the largest eigenvalue of $\mathrm{E}(\varepsilon\varepsilon^{\mathsf{T}})$ times m/k. This has three sources (variance and two biases) and hence

$$k^{-1}\mathrm{E}\sum_{i=1}^k \{\tilde{\psi}(u_i) - \psi(u_i)\}^2 = \mathcal{O}\Big[\frac{m}{k}\Big\{\frac{k}{n} + k\frac{\log n}{n} + k\Big(\frac{b^m}{m!}\Big)^2\Big\}\Big].$$

The factor k before the last two terms is due to the norm of the unit vector in $\mathbb{R}^k$, and, the last term is by (13). The theorem follows by taking $m = \log n / \log\log n$ □

A more general result can be based on an assumption like the following

A2. Assume that for some c, d and each ε there are $h_{\varepsilon,1}, \ldots, h_{\varepsilon,M(\varepsilon)}$ such that

$$\sup_{\xi} \min_{\gamma} \max_{c<u<d} \Big| g(u;\xi) - \sum_{j=1}^{M(\varepsilon)} \gamma_j h_j(u) \Big| < \varepsilon$$

Note that clearly the assumption ensures the existence of $\gamma(\cdot)$ such that $\max_{c<u<d} |g(u;\xi) - \sum_{j=1}^{M(\varepsilon)} \gamma_j(\xi) h_j(u)| < \varepsilon$, but then there are also $\beta_j = \int \gamma_j(\xi) f(\xi)\, d\xi$, $j = 1, \ldots, M(\varepsilon)$, such that $\max_{c<u<d} |\psi(u) - \sum_{j=1}^{M(\varepsilon)} \beta_j h_j(u)| < \varepsilon$.

The following theorem can be proved similarly to Theorem 4.2 :

Theorem 4.3 *Suppose assumptions* A1 *and* A2 *hold. Define* ε_n *by*

$$\varepsilon_n = \arg\min_{\varepsilon} \{M(\varepsilon)/n + \varepsilon\}.$$

Let $\tilde{\psi}$ *be the least squares estimate of the regression of* $\hat{\psi}$ *on* $h_{\varepsilon_n,1}, \ldots, h_{\varepsilon_n,M(\varepsilon_n)}$. *Then* $k^{-1} \sum_{i=1}^{k} \{\tilde{\psi}(u_i) - \psi(u_i)\}^2 = \mathcal{O}_p(\varepsilon_n)$.

In practice, Theorems 4.2 and 4.3 may seem to be of a limited use — a knowledge of the structure of the span of the individual utility functions is needed, and the regression is based on an identified efficient base, which may be not natural. For example, we used a polynomial base for the exponential utility function. The practical approach is an histogram or discrete approximation of f. We want now to discuss when such a procedure does yield an effective estimator, an estimator which is both statistically speaking efficient, but at the same time easy to compute and can be be used in off-the-shelf manner.

This is indeed the case. Let $\xi_1, \ldots, \xi_{M(\varepsilon)}$ be reasonably spaced points in the support of f. With the notation introduced after Assumption A2 , and by a similar argument, for a vector β on the simplex

$$\sup_{u} \Big| \sum_{j=1}^{M(\varepsilon)} \beta_j g(u;\xi_j) - \sum_{j=1}^{M(\varepsilon)} \beta_j \sum_{l=1}^{M(\varepsilon)} \gamma_l(\xi_j) h_l(u) \Big| \le \varepsilon.$$

Hence, one can use the base function $g(\cdot;\xi_1), \ldots, g(\cdot;\xi_{M(\varepsilon)})$ as well.

A Appendix: Proof of Theorem 4.1 .

We start the proof with the negative result. The proof is standard. We exhibit a small perturbation that cannot be detected. The perturbed density should remain a probablity density function with bounded second derivative. It should be however very wiggly so that the exponential mixing would smooth it out to make it hardly detectable through ψ. Very convenient candidates could be high derivatives of the normal density, but the supports of these functions are not bounded, while the support of f is bounded at least from below. We therefore use derivatives of approximations of the normal density. Here are the details.

Consider

$$\pi_m(\xi) = \pi_m(\xi; c, d) = \left\{1 - \left(\frac{\xi - c}{d}\right)^2\right\}^m \mathbf{1}\{\xi \in (c - d, c + d)\}$$

for some c, d, where $\mathbf{1}$ denotes the indicator function. π_m is approximately the normal pdf normalized improperly, cf. (24) below. Note that for $k \leq m$:

$$\int_{c-d}^{c+d} e^{u\xi} \pi_m^{(k)}(\xi)\, d\xi = (-1)^k u^k \int e^{u\xi} \pi_m(\xi)\, d\xi. \tag{14}$$

and

$$\pi_m^{(2k)}(c) = (-1)^k d^{-2k} \binom{m}{k} (2k)! \tag{15}$$

Write

$$\pi_m(\xi) = (1 - \frac{\xi - c}{d})^m (1 + \frac{\xi - c}{d})^m$$

and taking the derivative of the RHS:

$$\begin{aligned} &\pi_m^{2k}(\xi) \\ &= d^{-2k} \sum_{i=0}^{2k} \binom{2k}{i} (-1)^i \frac{m!}{(m-i)!} (1 - \tilde{\xi})^{m-i} \frac{m!}{(m-2k+i)!} (1 + \tilde{\xi})^{m-2k+i} \\ &= d^{-2k} \sum_{i=0}^{k} (-1)^i a_i, \qquad \text{say.} \end{aligned} \tag{16}$$

For simplicity we write $\tilde{\xi} = (\xi - c)/d$. Note that

$$\frac{a_{i+1}}{a_i} = \frac{2k - i}{i + 1} \frac{m - i}{m - 2k + i + 1} \frac{1 + \tilde{\xi}}{1 - \tilde{\xi}}$$

Hence the sum in the RHS of (16) is of unimodal terms with alternating signs. Let l be the index where the maximum is achieved:

$$\frac{2k-l}{l+1}\frac{m-l}{m-2k+l+1} = \{1+\circ(1)\}\frac{1-\tilde{\xi}}{1+\tilde{\xi}}. \tag{17}$$

Then

$$\begin{aligned} a_l &\geq a_l - \sum_{j=1}(a_{l+2j-1} - a_{l+2j}) - \sum_{j=1}(a_{l-2j+1} - a_{l-2j}) \\ &= (-1)^l \sum_{i=0}^{2k}(-1)^i a_i \\ &= a_l - a_{l+1} + \sum_{j=1}(a_{l+2j} - a_{l+2j+1}) - a_{l-1} + \sum_{j=1}(a_{l-2j} - a_{l-2j-1}) \\ &\geq -a_l \end{aligned} \tag{18}$$

where, if necessarily, the sequences are padded by zeros at the ends. But then for some $C = \mathcal{O}(1)$, C may vary from line to line:

$$\begin{aligned} a_l &= (2k)!\binom{m}{l}\binom{m}{2k-l}(1-\tilde{\xi})^{m-l}(1+\tilde{\xi})^{m-2k+l} \\ &\leq C(2k)!\frac{m^{2m}(1-\tilde{\xi})^{m-l}(1+\tilde{\xi})^{m-2k+l}}{l^l(m-l)^{m-l}(2k-l)^{2k-l}(m-2k+l)^{m-2k+l}} \\ &= C(2k)!\ (1-\tilde{\xi}^2)^{m-2k}\Big\{\frac{k(1-\tilde{\xi})}{2k-l}\Big\}^{2k-l}\ \Big\{\frac{k(1+\tilde{\xi})}{l}\Big\}^{l}\ \{\frac{m}{k}\}^{2k} \\ &\qquad \times \{1+\frac{l}{m-l}\}^{m-l}\{1+\frac{2k-l}{m-2k+l}\}^{m-2k+l} \end{aligned} \tag{19}$$

To deal with the following terms of the RHS of (19) we assume that $0 < l \leq k$. The case $2k > l \geq k$ is dealt similarly. The cases of $l \in \{0, 2k\}$ are simple

$$\begin{aligned} \Big\{\frac{k(1-\tilde{\xi})}{2k-l}\Big\}^{2k-l}\ \Big\{\frac{k(1+\tilde{\xi})}{l}\Big\}^{l} &= \Big\{\frac{k(1-\tilde{\xi})}{2k-l}\Big\}^{2(k-l)}\Big\{\frac{1+\tilde{\xi}}{1-\tilde{\xi}}\frac{2k-l}{l}\Big\}^{l} \\ &\leq 2^k\Big\{\frac{1+\tilde{\xi}}{1-\tilde{\xi}}\frac{2k-l}{l+1}\Big\}^{l} \\ &\leq 3^k\Big\{\frac{m-2k+l+1}{m-l}\Big\}^{l}, \qquad \text{by (17)} \\ &\leq 3^k \end{aligned} \tag{20}$$

The next bound is easy,

$$\{1+\frac{l}{m-l}\}^{m-l}\{1+\frac{2k-l}{m-2k+l}\}^{m-2k+l}<e^{2k}, \tag{21}$$

since $(1+1/x)^x<e^x$ for any $x>0$. We conclude from (16), (18), (20), and (21):

$$\|\pi_m^{(2k)}\|_\infty \le a_l \le C(2k)!\{c_2\frac{m}{k}\}^{2k} \tag{22}$$

for $c_2>1$.

Let

$$\Delta_{m,k}(\xi)=\frac{d^{2k}}{(2k+2)!(c_1c_2)^{2k}}\pi_m^{(2k)}(\xi)$$

where we take $m=\lceil c_1k\rceil$. Note that by (22) $\Delta_{m,k}^{(2)}$ is uniformly bounded, while by (15)

$$\Delta_{m,k}(c)\ge c_3k^{-2}(c_1c_2)^{-2k}\binom{m}{k}\ge c_4^{-k} \tag{23}$$

for some $c_4>1$. However by (14)

$$\begin{aligned}
&\int \xi^{-1}\{e^{u\xi}-1\}\{f(\xi)+\xi\Delta_{m,k}(\xi)\}\,d\xi\\
&\quad=\psi(u)+\frac{d^{-2(m-k)}}{(2k+2)!c_2^{2k}}u^{2k}\int_a^b \pi_m(\xi)e^{u\xi}\,d\xi\\
&\quad=\psi(u)+(-1)^k\{1+o(1)\}\frac{d^{2k}}{(2k+2)!c_2^{2k}}u^{2k}\int_a^b e^{-m(\xi-c)^2/d^2}e^{u\xi}\,d\xi\\
&\quad=\psi(u)+(-1)^k\{1+o(1)\}\frac{\sqrt{2\pi}d^{2k+1}}{(2k+2)!m^{1/2}c_2^{2k}}u^{2k}e^{uc}.
\end{aligned}$$

Hence if

$$\frac{d^{2k+1}}{(2k+2)!m^{1/2}(c_1c_2)^{2k}}=o(n^{-1/2}),$$

or $k\log k-\log n\to\infty$, then one would not be able to test between f to $f+\xi\Delta_{m,k}$. In particular this happens when $k=\log n/\log\log n$. However, then, by (23), $n^\alpha\Delta_{m,k}(c)\to\infty$ for any $\alpha>0$. This proves that f can be estimated in any n^α, $\alpha>0$ rate.

We move now to the positive result. We suggest an estimator of the mixing density f whose rate of convergence is easy to evaluate. Of course, the practical way would be the standard least squares as discussed in Subsection 4.3 , but then rates are difficult to evaluate. We suggest therefore in the proof a kernel estimator of g given by $\int \hat{\psi}(u)\bar{K}(u)\,du$ for some $\bar{K}$ given below. Here are the details.

If $\psi(u) = \int g(u;\xi)\, f(\xi)\, d\xi$, let $\psi_s = \psi_s(u) = e^{-us}(\psi(u)\text{-}1)$. Assume for simplicity that by assumption $f(\xi) = 0$ for $\xi \notin (s_0 - d, s_0 + d)$. Then since

$$\psi_s(u) = \int e^{u(\xi-s)}\xi^{-1} f(\xi)\, d\xi - e^{-us}\int \xi^{-1} f(\xi)\, d\xi$$
$$\psi_s^{(k)}(u) = \int (\xi - s)^k e^{u(\xi-s)}\xi^{-1} f(\xi)\ d\xi - (-1)^k s^k e^{-us}\int \xi^{-1} f(\xi)\, d\xi,$$

then formally:

$$\sqrt{\frac{m}{2\pi d^2}}\sum_{k=0}^{m}\binom{m}{k}\{\frac{-1}{d^2}\}^k\psi_s^{(2k)}(u)$$
$$= \sqrt{\frac{m}{2\pi d^2}}\int \pi_m(\xi; s, d)e^{u(\xi-s)}\xi^{-1} f(\xi)\, d\xi$$
$$- \sqrt{\frac{m}{2\pi d^2}}\pi_m(s; 0, d)e^{-us}\int \xi^{-1} f(\xi)\, d\xi$$

where $\pi_m(\cdot) = \pi_m(\cdot; s, d)$. Note that for any smooth bounded function h with two bounded derivatives:

$$\begin{aligned}\sqrt{\frac{m}{2\pi d^2}}\int \pi_m(\xi; s, d)h(\xi)\, d\xi &\approx \sqrt{\frac{m}{d^2}}\int \varphi\big\{\sqrt{m}(\xi - s)/d\big\}h(\xi)\, d\xi \\ &\approx h(s) + \mathcal{O}(m^{-1}),\end{aligned} \tag{24}$$

where φ is the standard normal density. Hence

$$\sqrt{\frac{m}{2\pi d^2}}\sum_{k=0}^{m}\binom{m}{k}\Big(\frac{-1}{d^2}\Big)^k\psi_s^{(2k)}(u) \to s^{-1}\xi(s) \quad \text{as } m \to \infty. \tag{25}$$

Let $\hat{\psi}_s$ be an estimator of ψ_s. Let K be a smooth kernel of order $2m$, integrated to 1, and with bounded support kernel. Then by (25) we can

estimate $f(s)$ by

$$\begin{aligned}\hat{f}(s) &= s\sqrt{\frac{m}{2\pi d^2}}\sum_{k=0}^{m}\binom{m}{k}\Big(\frac{-1}{d^2}\Big)^k\int K(u)\hat{\psi}_s^{(2k)}(u)\,du \\ &= s\sqrt{\frac{m}{2\pi d^2}}\sum_{k=0}^{m}\binom{m}{k}\Big(\frac{-1}{d^2}\Big)^k\int K^{(2k)}(u)\hat{\psi}_s(u)\,du \\ &= \int \bar{K}(u)\hat{\psi}_s(u)\,du\end{aligned} \tag{26}$$

where

$$\bar{K}(u) \equiv s\sqrt{\frac{m}{2\pi d^2}}\sum_{k=0}^{m}\binom{m}{k}\Big(\frac{-1}{d^2}\Big)^k K^{(2k)}(u)$$

Since we have already developed the machinery we pick

$$K(u) = \gamma_m\sqrt{\frac{2m}{2\pi\sigma^2}}\pi_{2m}(u;u_0,\sigma)$$

where $\gamma_m = 1+\mathfrak{o}(1)$. Hence by (22)

$$\|\bar{K}\|_\infty \le s\frac{m}{2\pi\sigma d}\sum_{k=0}^{m}\binom{m}{k}\Big(\frac{cm}{k}\Big)^{2k}(2k)! = \mathcal{O}(c^m m^m). \tag{27}$$

If ψ_s can be estimated at a standard polynomial rate, $\hat{\psi}-\psi = \mathcal{O}_p(n^{-\gamma})$, then we obtain from (26) and (27) that $\hat{\psi}$ induce an error of $\mathcal{O}(c^m m^m/n^\gamma)$. To this we have to add the bias of $\mathcal{O}(m^{-1})$ as given by (24). The minimization of the error estimate is obtained therefore of the order of the value at m when these two terms are equal:

$$m\log m - \gamma\log n = \log m.$$

By taking $m = m_n = \alpha\log n/\log\log n$ we achieve the rate of

$$\hat{f}(s) - f(s) = \mathcal{O}_p\big(n^{-\alpha\log\log n/\log n}\big),$$

for any $\alpha < 1$. We have shown that the optimal rate of convergence is n^{α_n} for some $\alpha_n \to 0$ slowly.

REFERENCES

Ait-Sahalia, Y. and Lo, A. (2000). Nonparametric risk-management and implied risk aversion. *Journal of Econometrics*, **94**.

Balabdaoui, F. and Wellner, J. A. (2007). Estimation of a k-monotone density: limit distribution theory and the spline connection. Manuscript.

Detlefsen K., Härdle, ,W.K., and Moro, R. A. (2007) ,Empirical Pricing Kernels and Investor Preferences. Manuscript

Feller, W. (1966): *An introcustion to Probability Theory and Its applications, vol. II* Wiley, New-York.

Friedman, M. and Savage, L. P. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, **56**, 279304.

Gilbert, P. B. (2000): Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. *The Annals of Statistics* **28**, 151–194.

Gilbert, P. B., Lele, S. R., and Vardi,Y.(1999). Maximum likelihood estimation in semipara- metric selection bias models with application to AIDS vaccine trials. *Biometrika* **86** 27–43.

Jewell, N. P. (1982). Mixtures of exponential distributions. *Ann. of Statis.* **10**, 479–482.

Ji, W, H., Veitch, C. R., and Craig, J. L. (1999). An evaluation of the efficiency of rodent trapping methods: The effect of trap arrangement, cover type, and bait, *New Zealand Journal Of Ecology* **23**, 45–51.

Gill, R. D., Vardi, Y., and Wellner, J. A. (1988).Large sample theory of empirical distributions in biased sampling models. *Ann.Statist.* **16** 1069 1112.

Silverman, B., (1986). *Density Estimation*. Chapman and Hall, London.

Vardi, Y. (1985). Empirical distributions in selection bias models. *Ann.Statist.* **13**, 178-203.

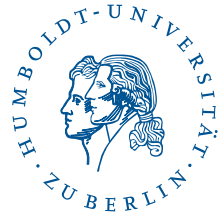
SFB 649 Discussion Paper Series 2007

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Trade Liberalisation, Process and Product Innovation, and Relative Skill Demand" by Sebastian Braun, January 2007.
002 "Robust Risk Management. Accounting for Nonstationarity and Heavy Tails" by Ying Chen and Vladimir Spokoiny, January 2007.
003 "Explaining Asset Prices with External Habits and Wage Rigidities in a DSGE Model." by Harald Uhlig, January 2007.
004 "Volatility and Causality in Asia Pacific Financial Markets" by Enzo Weber, January 2007.
005 "Quantile Sieve Estimates For Time Series" by Jürgen Franke, Jean-Pierre Stockis and Joseph Tadjuidje, February 2007.
006 "Real Origins of the Great Depression: Monopolistic Competition, Union Power, and the American Business Cycle in the 1920s" by Monique Ebell and Albrecht Ritschl, February 2007.
007 "Rules, Discretion or Reputation? Monetary Policies and the Efficiency of Financial Markets in Germany, 14th to 16th Centuries" by Oliver Volckart, February 2007.
008 "Sectoral Transformation, Turbulence, and Labour Market Dynamics in Germany" by Ronald Bachmann and Michael C. Burda, February 2007.
009 "Union Wage Compression in a Right-to-Manage Model" by Thorsten Vogel, February 2007.
010 "On σ−additive robust representation of convex risk measures for unbounded financial positions in the presence of uncertainty about the market model" by Volker Krätschmer, March 2007.
011 "Media Coverage and Macroeconomic Information Processing" by Alexandra Niessen, March 2007.
012 "Are Correlations Constant Over Time? Application of the CC-TRIG$_t$-test to Return Series from Different Asset Classes." by Matthias Fischer, March 2007.
013 "Uncertain Paternity, Mating Market Failure, and the Institution of Marriage" by Dirk Bethmann and Michael Kvasnicka, March 2007.
014 "What Happened to the Transatlantic Capital Market Relations?" by Enzo Weber, March 2007.
015 "Who Leads Financial Markets?" by Enzo Weber, April 2007.
016 "Fiscal Policy Rules in Practice" by Andreas Thams, April 2007.
017 "Empirical Pricing Kernels and Investor Preferences" by Kai Detlefsen, Wolfgang Härdle and Rouslan Moro, April 2007.
018 "Simultaneous Causality in International Trade" by Enzo Weber, April 2007.
019 "Regional and Outward Economic Integration in South-East Asia" by Enzo Weber, April 2007.
020 "Computational Statistics and Data Visualization" by Antony Unwin, Chun-houh Chen and Wolfgang Härdle, April 2007.
021 "Ideology Without Ideologists" by Lydia Mechtenberg, April 2007.
022 "A Generalized ARFIMA Process with Markov-Switching Fractional Differencing Parameter" by Wen-Jen Tsay and Wolfgang Härdle, April 2007.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".



023 "Time Series Modelling with Semiparametric Factor Dynamics" by Szymon Borak, Wolfgang Härdle, Enno Mammen and Byeong U. Park, April 2007.

024 "From Animal Baits to Investors' Preference: Estimating and Demixing of the Weight Function in Semiparametric Models for Biased Samples" by Ya'acov Ritov and Wolfgang Härdle, May 2007.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".