Statistica Sinica 20(2010): Supplement,

INVESTORS' PREFERENCE: ESTIMATING AND DEMIXING OF THE WEIGHT FUNCTION IN SEMIPARAMETRIC MODELS FOR BIASED SAMPLES.

Ya'acov Ritov and Wolfgang K. Härdle

The Hebrew University of Jerusalem and Humboldt-Universität zu Berlin

Supplementary Material

A Appendix: Proof of Theorem 3.3.

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 probability density function with a 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 ϑ is bounded at least from below. We therefore use derivatives of approximations of the normal density. Here are the details.

Consider

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

for some c, d, where $\mathbf{1}$ denotes the indicator function. π_m is approximately the normal pdf normalized improperly, cf. (11) below. Note that for $k \le m$:

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

and

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

Write

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

and taking the derivative of the RHS:

$$
\begin{aligned}
&\pi_m^{2k}(\xi) \\
&= d^{-2k} \sum_{i=0}^{2k} \binom{2k}{i} (-1)^i \frac{m!}{(m-i)!} (1-\tilde{\xi})^{m-i} \frac{m!}{(m-2k+i)!} (1+\tilde{\xi})^{m-2k+i} \\
&= d^{-2k} \sum_{i=0}^{k} (-1)^i a_i, \qquad \text{say.}
\end{aligned} \tag{3}
$$

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

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

It follows that the RHS of (3) is a sum of unimodal terms with alternating signs (i.e., there is an l such that $a_1, \ldots, a_l$ is an increasing sequence, while $a_l, \ldots, a_k$ is a decreasing one), where l is defined by:

$$
\frac{2k-l}{l+1} \frac{m-l}{m-2k+l+1} = \{1 + \mathcal{o}(1)\} \frac{1-\tilde{\xi}}{1+\tilde{\xi}}. \tag{4}
$$

Then

$$
\begin{aligned}
a_l &\geq a_l - \sum_{j=1} (a_{l+2j-1} - a_{l+2j}) - \sum_{j=1} (a_{l-2j+1} - a_{l-2j}) \\
&= (-1)^l \sum_{i=0}^{2k} (-1)^i a_i \\
&= a_l - a_{l+1} + \sum_{j=1} (a_{l+2j} - a_{l+2j+1}) - a_{l-1} + \sum_{j=1} (a_{l-2j} - a_{l-2j-1}) \\
&\geq -a_l,
\end{aligned} \tag{5}
$$

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:

$$
\begin{aligned}
a_l &= (2k)! \binom{m}{l} \binom{m}{2k-l} (1-\tilde{\xi})^{m-l} (1+\tilde{\xi})^{m-2k+l} \\
&\leq C (2k)! \frac{m^{2m} (1-\tilde{\xi})^{m-l} (1+\tilde{\xi})^{m-2k+l}}{l^l (m-l)^{m-l} (2k-l)^{2k-l} (m-2k+l)^{m-2k+l}} \\
&= C (2k)!\, (1-\tilde{\xi}^2)^{m-2k} \Big\{ \frac{k(1-\tilde{\xi})}{2k-l} \Big\}^{2k-l} \Big\{ \frac{k(1+\tilde{\xi})}{l} \Big\}^{l} \Big\{ \frac{m}{k} \Big\}^{2k} \\
&\qquad \times \Big\{ 1 + \frac{l}{m-l} \Big\}^{m-l} \Big\{ 1 + \frac{2k-l}{m-2k+l} \Big\}^{m-2k+l}
\end{aligned} \tag{6}
$$

To deal with the following terms of the RHS of (6) we assume that $0 < l \le k$. The case $2k > l \ge k$ is dealt similarly. The case of $l \in \{0, 2k\}$ is simple:

$$\begin{aligned}\Big\{\frac{k(1-\tilde{\xi})}{2k-l}\Big\}^{2k-l}\Big\{\frac{k(1+\tilde{\xi})}{l}\Big\}^{l} &= \Big\{\frac{k(1-\tilde{\xi})}{2k-l}\Big\}^{2(k-l)}\Big\{\frac{1+\tilde{\xi}}{1-\tilde{\xi}}\frac{2k-l}{l}\Big\}^{l} \\ &\le 2^k\Big\{\frac{1+\tilde{\xi}}{1-\tilde{\xi}}\frac{2k-l}{l+1}\Big\}^{l} \\ &\le 3^k\Big\{\frac{m-2k+l+1}{m-l}\Big\}^{l}, \qquad \text{by (4)} \\ &\le 3^k\end{aligned} \tag{7}$$

The next bound is easy,

$$\big\{1+\frac{l}{m-l}\big\}^{m-l}\big\{1+\frac{2k-l}{m-2k+l}\big\}^{m-2k+l} < e^{2k}, \tag{8}$$

since $(1+1/x)^x < e^x$ for any $x > 0$. We conclude from (3), (5), (7), and (8):

$$\|\pi_m^{(2k)}\|_\infty \le a_l \le C(2k)!\big\{c_2\frac{m}{k}\big\}^{2k} \tag{9}$$

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 $m = \lceil c_1 k \rceil$. Note that by (9) $\Delta_{m,k}^{(2)}$ is uniformly bounded, while by (2)

$$\Delta_{m,k}(c) \ge c_3 k^{-2}(c_1c_2)^{-2k}\binom{m}{k} \ge c_4^{-k} \tag{10}$$

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

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

Hence if

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

or $k\log k - \log n \to \infty$, then one would not be able to test between ϑ to $\vartheta + \xi\Delta_{m,k}$. In particular this happens when $k = \log n/\log\log n$. However, then, by (10), $n^\alpha \Delta_{m,k}(c) \to \infty$ for any $\alpha > 0$. This proves that ϑ cannot be estimated in any n^α, $\alpha > 0$ rate.

We move now to the positive result. We suggest an estimator of the mixing density ϑ whose rate of convergence is easy to evaluate. Of course, the practical way would be the standard least squares as discussed in Subsection 3.3, 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 $\psi(u) = \int g(u;\xi)\,\vartheta(\xi)\,d\xi$, let $\psi_s = \psi_s(u) = e^{-us}(\psi(u)\text{-}1)$. Assume for simplicity (but wlog) that by assumption $\vartheta(\xi) = 0$ for $\xi \notin (s_0 - d, s_0 + d)$. Since

$$\begin{aligned}
\psi_s(u) &= \int e^{u(\xi-s)}\xi^{-1}\vartheta(\xi)\,d\xi - e^{-us}\int \xi^{-1}\vartheta(\xi)\,d\xi \\
\psi_s^{(k)}(u) &= \int (\xi-s)^k e^{u(\xi-s)}\xi^{-1}\vartheta(\xi)\;d\xi - (-1)^k s^k e^{-us}\int \xi^{-1}\vartheta(\xi)\,d\xi,
\end{aligned}$$

then formally:

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

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

$$\begin{aligned}
&\sqrt{\frac{m}{2\pi d^2}}\int \pi_m(\xi;s,d)h(\xi)\,d\xi \\
&\qquad = \sqrt{\frac{m}{d^2}}\int \varphi\{\sqrt{m}(\xi-s)/d\}h(\xi)\,d\xi + \mathcal{O}(m^{-1}), \\
&\qquad = h(s) + \mathcal{O}(m^{-1}),
\end{aligned} \tag{11}$$

where φ is the standard normal density. Hence

$$\sqrt{\frac{m}{2\pi d^2}} \sum_{k=0}^{m} \binom{m}{k} \Big(\frac{-1}{d^2}\Big)^k \psi_s^{(2k)}(u) \to s^{-1}\xi(s) \quad \text{as } m \to \infty. \tag{12}$$

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 (12) $\vartheta(s)$ can be estimated by

$$\begin{aligned} \hat{\vartheta}(s) &= s\sqrt{\frac{m}{2\pi d^2}} \sum_{k=0}^{m} \binom{m}{k} \Big(\frac{-1}{d^2}\Big)^k \int K(u)\hat{\psi}_s^{(2k)}(u)\,du \\ &= s\sqrt{\frac{m}{2\pi d^2}} \sum_{k=0}^{m} \binom{m}{k} \Big(\frac{-1}{d^2}\Big)^k \int K^{(2k)}(u)\hat{\psi}_s(u)\,du \\ &= \int \bar{K}(u)\hat{\psi}_s(u)\,du \end{aligned} \tag{13}$$

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 $\gamma_m = 1 + \mathcal{o}(1)$. Hence by (9)

$$\|\bar{K}\|_\infty \le s\frac{m}{2\pi\sigma d} \sum_{k=0}^{m} \binom{m}{k} \Big(\frac{cm}{k}\Big)^{2k} (2k)! = \mathcal{O}(c^m m^m). \tag{14}$$

If ψ_s can be estimated at a standard polynomial rate, $\hat{\psi} - \psi = \mathcal{O}_p(n^{-\gamma})$, then, by (13) and (14)m $\hat{\psi}$ induce an error of $\mathcal{O}(c^m m^m / n^\gamma)$. To this we the bias of $\mathcal{O}(m^{-1})$ as given by (11) should be added. 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 $m = m_n = \alpha \log n / \log\log n$ the rate of

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

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