Estimating the weight function in biased sample

Wolfgang K. Härdle

Ya'acov Ritov

April 9, 2007

Jump

A model is considered to be biased if it can be naturally represented as a sample from a density with respect to some dominating measure μ given by

$$p(x) = \frac{q(x)w(x)}{\int q(x')w(x')d\mu(x')},$$

where q is some 'natural' pdf for the problem, representing in some sense the 'true' underlined distribution, while w is a given weight function that biased the sample.

Vardi (1985) was the first to analyze systematically these type of models. Asymptotic theory was develop in Gill, Vardi and Wellner (1088). 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).

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).

$$p(x; f, h, g) = \frac{w(u, z; f)g(u)h(z)}{\int \int w(\tilde{u}, \tilde{z}; f)g(\tilde{u})h(\tilde{z})\, \mu_{.}\, (\tilde{u})\, d\nu(\tilde{z})}.$$

$w(u, z; f)$ is some parametric weight function, f is the parameter to be estimated. h is unknown, and typically too complicated to be estimated. g is known (and it is part of the design).

We will consider the weight function with $u \in \mathbb{R}$ and $z \in \mathbb{R}^p$

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

The likelihood equation for f is then:

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

If Z has at least one continuous component, the GMLE estimator of its distribution is discrete with a point mass at each observation, so that we obtain

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

where $\hat{h}_j$ is the estimated mass at the point Z_j.

However, the likelihood equation for h is clearly

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

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

$$0 = \dot{\ell}(f) = \frac{1}{n} \sum_i U_i Z_i - \sum_j \frac{\int u z_j e^{uf^{\mathsf{T}} z_j} g(u)\, d\mu(u)}{\int e^{uf^{\mathsf{T}} z_j} g(u)\, d\mu(u)}$$

$$= \frac{1}{n} \sum_i Z_i \{ U_i - \hat{E}_f \ (U | f^{\mathsf{T}} Z_i) \}$$

$E_f(U|z)$ is function of the known $g(\cdot)$ and $w(\cdot;\cdot)$.

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

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

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

$$0 = \dot{\ell}(f) = \frac{1}{n} \sum_i Z_i \{U_i - \hat{E}_f \ (U | f^\mathsf{T} Z_i)\}$$

Generally, the derivative of $\hat{E}_f \ (U | f^\mathsf{T} Z_i)\}$ is

$$- \operatorname{Var}_f \ (U | f^\mathsf{T} Z_i) Z_i Z_i^\mathsf{T}.$$

Hence ℓ is concave in f.

Hence the maximizer of ℓ is simple to find and is asymptotically normal with asymptotic covariance function given by $\mathrm{E}\{\operatorname{Var}_f \ (U | f^\mathsf{T} z) Z Z^\mathsf{T}\}$

The next n pages are taken from the talk "Empirical Pricing Kernels and Investor Preferences" of K. Detlefsen, W. K. Härdle1, and R. A. Moro1.

An investor observes the stock price and forms his subjective opinion about the future evolution.

An opinion on the future value S_t can be described by a subjective density p (historical or physical density).

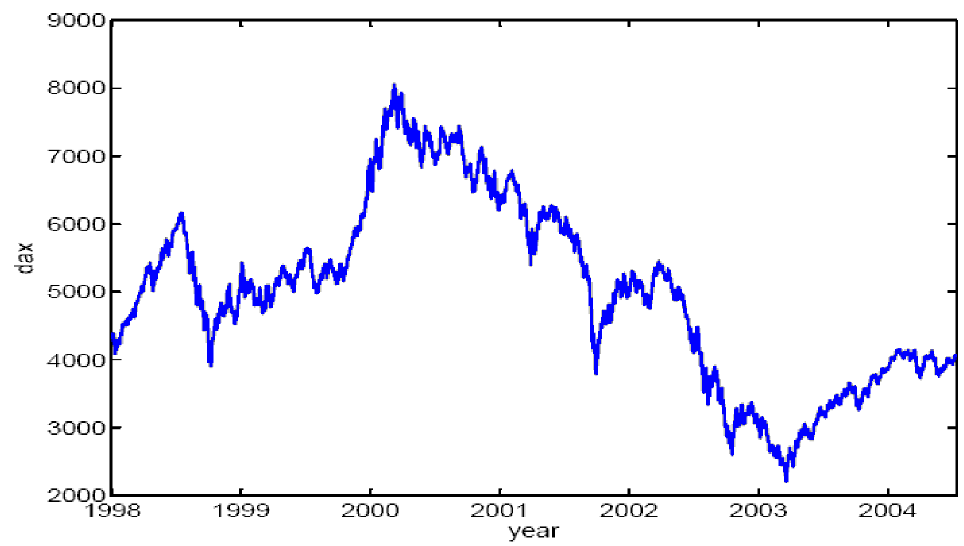


Figure 1: DAX, 1998 – 2004. Daily observations.

There is also a state-price density (SPD) q implied by the market prices of options.

The SPD (a.k.a. risk-neutral density) differs from p because it corresponds to replication strategies (martingale risk neutral measure).

A person alone does not use in general a replication strategy but thinks in terms of his p density.

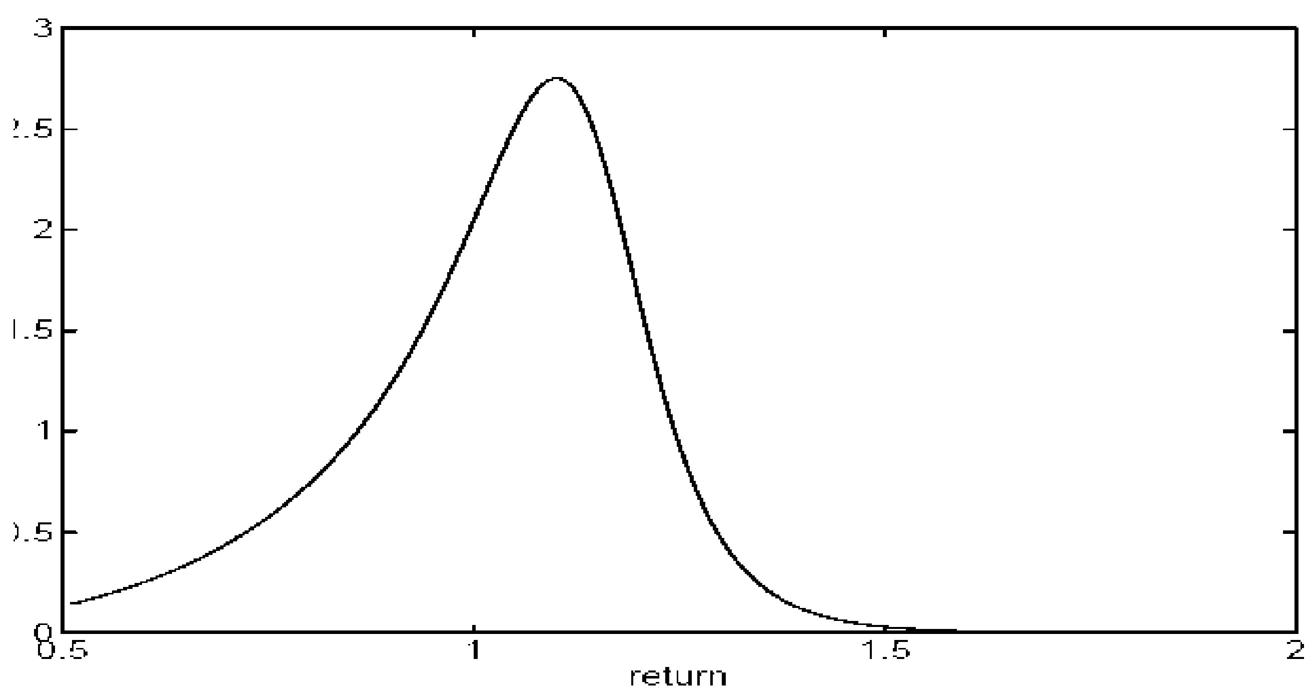
For SPD estimation a Heston continuous stochastic volatility model is used, which is an industry standard for option pricing models:

$$\frac{dS_t}{S_t} = r\,dt + \sqrt{V_t}dW_t^1$$

where the volatility process is modeled by a square-root process:

$$dV_t = \xi(\eta - V_t)dt + \theta\sqrt{V_t}dW_t^2,$$

and W^1 and W^2 are Wiener processes with correlation ρ.


3
2.5
2
1.5
1
0.5
0
0.5
1
1.5
2
return

The pricing kernel $K(x)$ is defined as:

$$K(x) = \frac{q(x)}{p(x)}$$

An estimate of the pricing kernel is called empirical pricing kernel (EPK). We use the estimate:

$$\hat{K}(x) = \frac{\hat{q}(x)}{\hat{p}(x)}$$

where $\hat{q}$ and $\hat{p}$ are the estimated risk-neutral and subjective densities.

Since . . . are equivalent . . . the pricing kernel is:

$$K(S_T) = \frac{q(S_T)}{p(S_T)} = \frac{U'(S_T)}{U'(S_0)}$$

The aggregate return in the perceptible state u is given by:

$$R_A(u) = \int U^{-1}(u, z) f(z) dz$$

In oder to solve this for $f(\cdot)$:

$$\arg\min_{f \in \mathcal{F}} \int \left\{ R_f^A(u) - U_M^{-1}(u) \right\}^2 \tilde{P}(du),$$

where $U_M^{-1}(u)$ is the inverse of the estimated market utility function, $\tilde{P}$ is the distribution of utility levels.

End of Cut and paste section

It is assumed in that the observed density of the form

$$p(x) = cq_\nu(x)w(x; f)$$

where q_ν is assumed to follow a given parametric function and c is a normalization factor. The weight function is theoretically derived to be given by

$$w(x; f) = 1/U'(x),$$

where U is the market utility function.

The market utility function itself is assumed to be a function of the mixture of the individual investors, such that

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

where $g(\cdot;\cdot)$ is the inverse utility function and it is considered known, and $f = f(\cdot)$ is a probability density function. A subject with inverse utility function $g(\cdot;\xi)$ has 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 therefore the representation:

$$p(x) = c q_\nu(x) \int \frac{\partial}{\partial u} g(u;\xi) f(\xi)\, d\xi,$$

$$\text{with } x = \int g(u;\xi) f(\xi)\, d\xi.$$

Like DHM we want to solve for f

$$\int g(u;\xi)f(\xi)\,d\xi = \psi(u),$$

• q_ν is estimated as a parametric density.

• p can be estimated at a standard non-parametric rate based on sample from p.

• We assume therefore that $\psi = cp/q$ 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 $\alpha_i > 0$.

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 the approximation as

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

This looks like a standard linear model, and indeed we suggest to estimate f by solving it.

However, some basic assumptions of regression are violated.

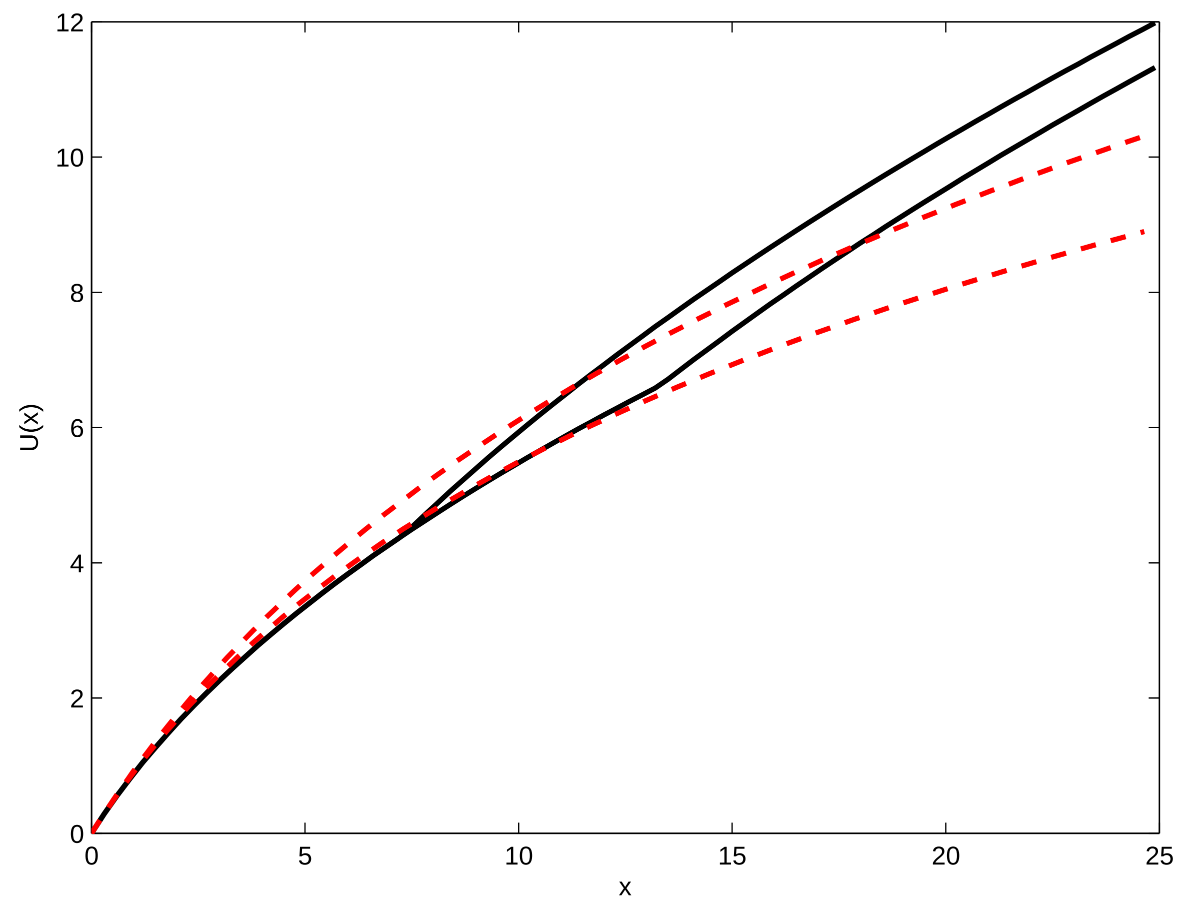
How fast can f be estimated? We are going to present toys examples similar to those of DHM. These examples show that in very similar models can have completely different behavior type:

(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.

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



Consider the individual utility function:

$$U(x;\xi) = \alpha_2 c^{1-1/\alpha_2} [x-\xi]_+^{1/\alpha_1} \vee (x+c)^{1/\alpha_2} - \alpha_2 c, \quad \xi > 0$$

where $\alpha_1, \alpha_2 > 1$ and $c > 0$ are given.

Then

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

where $\beta = \alpha_2^{-1} c^{-1+1/\alpha_2}$. To simplify and generalize a little, we consider the general case where $\xi \in \mathbb{R}$ and

$$g(u;\xi) = \begin{cases} g_2(u) & u \le h(\xi) \\ g_1(u) + \xi & u > h(\xi) \end{cases}$$

where

$$h = g_2 - g_1 \nearrow .$$

Then

$$\psi(u) = \int_0^{h^{-1}(u)} \xi f(\xi)\, d\xi + g_2(u) F\{h^{-1}(u)\} + g_2(u)\Big\{1 - F\{h^{-1}(u)\}\Big\}$$

where F is the cdf corresponding to the pdf f. Changing variables:

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

Taking a derivative:

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

Hence $f(\cdot)$ can be estimated in the same rate as the rate of the estimation of second derivative of ψ, which 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 s bounded derivative, then f can be estimated with an $\mathcal{O}_p(n^{-(s-2)/(2s+1)})$ error.

Modest changes in the inverse utility function may create situations in which f can hardly estimated, or even not at all.

Suppose $g(u;\xi) = (u\xi)^{-1}\{(u+\xi)^{\alpha} - 1\}$ for $\xi \in \mathbb{R}^{+}$ and known $\alpha > 1$. If α is integer then $\psi(\cdot)$ is only a function of the first α moments of f, and hence there is no consistent estimator of f.

Seemingly as $\alpha \to \infty$ more and more moments will be revealed.

The limiting form of the inverse utility function, as $\alpha \to \infty$ and $\alpha\zeta \to \xi$ is given by

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

The density f is now identified.

For example, all its moments can be estimated, e.g., by $\int \xi^i f(\xi)\, d\xi = \psi^{(i+1)}(0)$. Really?!

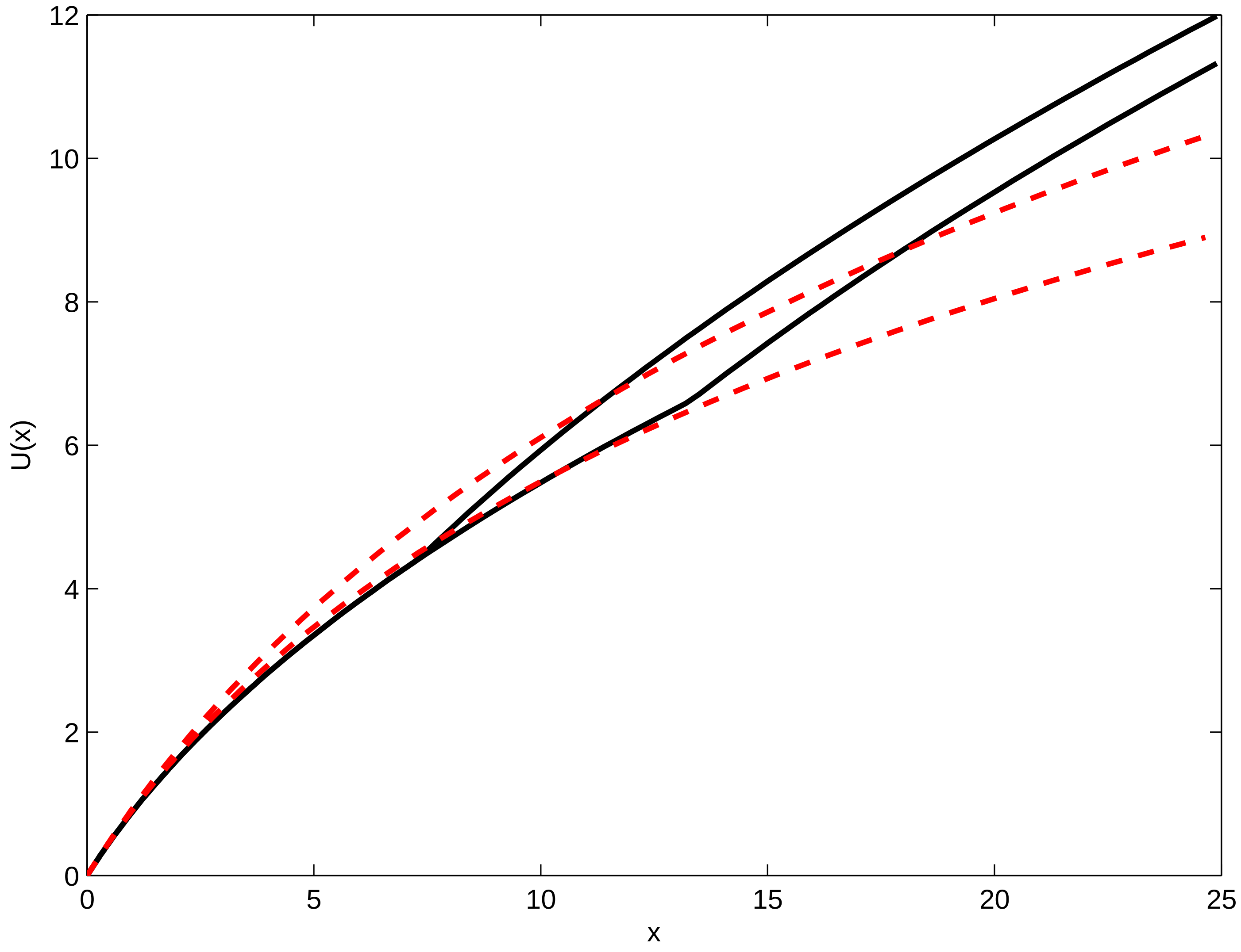


Figure 1: The switching utility function $U(\cdot;\xi)$ ($\alpha_1 = 2$, $\alpha_2 = 2.25$, $c = 2$) for two different values of ξ (solid lines), and of the logarithmic utility for two values of ξ (broken lines).

We will argue now that if $f(\cdot)$ is assumed to have two bounded derivatives then its value at a point can be estimated, but in a very slow rate, slower than any polynomial rate. To be more exact,we argue that there is an estimator such that $\hat{f}(s) - f(s) = \mathcal{O}_p(n^{-\alpha \log\log n/\log n})$ for some α, and that there is not any other estimator $\tilde{f}(s)$, such that $\tilde{f}(s) - f(s) = \mathcal{O}_p(n^{-\alpha/\log\log n})$ for some $\alpha > 0$.

The idea is standard: pick $\tilde{\Delta}_n(\cdot)$ such that

$$\int g(u;\xi)\tilde{\Delta}_n(\xi)\,d\xi = o(n^{-1/2}).$$

Then one cannot test between $f_0(\cdot)$ and $f_n(\cdot) = f_0(\cdot) + \tilde{\Delta}_n(\cdot)$. However if $n^\alpha\tilde{\Delta}(c) \to \infty$, then the rate n^α cannot be achieved.

The perturbation should be as non-smooth as permitted by by the three restrictions: $f_0+\tilde{\Delta}_n$ should (i) integrated to 1; (ii) be positive; and (iii) the 2nd derivative should be uniformly bounded; and (iv) $\tilde{\Delta}$ should have compact support.

We wanted to work with $\tilde{\Delta} = \varphi^{(m)}$, a high derivative of the normal pdf — a nice smooth function with known derivatives. However, its support is not compact. So approximations are in need.

Let

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

Note that for $k \leq m$:

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

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

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

The a_i series is bimodal, hence the sum is of the order of the largest a_i.

$$\|\pi_m^{(2k)}\|_\infty \le \max a_i \le C(2k)!\{c_2\frac{m}{k}\}^{2k}.$$

Let

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

We take $m = \lceil c_1 k \rceil$. Then $\Delta_{m,k}^{(2)}$ is uniformly bounded, while

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

for some $c_4 > 1$.

$$\int \xi^{-1} \left(e^{u\xi} - 1 \right) \left(f(\xi) + \xi \Delta_{m,k}(\xi) \right) d\xi$$
$$= \cdots$$
$$= \psi(u) + (-1)^k \left(1 + \mathrm{o}(1) \right) \frac{\sqrt{2\pi} d^{2k+1}}{(2k+2)! m^{1/2} c_2^{2k}} u^{2k} e^{uc}.$$

Hence if

$$\frac{d^{2k+1}}{(2k+2)! m^{1/2} (c_1 c_2)^{2k}} = o(n^{-1/2}),$$

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 $k = \log n / \log \log \log n$. However, then $n^\alpha \Delta_{m,k}(c) \to \infty$ for any $\alpha > 0$. This proves that f can be estimated in any n^α, $\alpha > 0$ rate.

The practical way would be standard least squares, but then rates are difficult to evaluate.

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

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

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

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

Formally:

$$\sqrt{\frac{m}{2\pi d^2}} \sum_{k=0}^{m} \binom{m}{k} \{\frac{-1}{d^2}\}^k \psi_s^{(2k)}(u)$$

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

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

$$\to s^{-1} f(s).$$

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

$$\hat{f}(s) = s\sqrt{\frac{m}{2\pi d^2}} \sum_{k=0}^{m} \binom{m}{k} \left(\frac{-1}{d^2}\right)^k \int K(u) \hat{\psi}_s^{(2k)}(u)\, du$$

$$= s\sqrt{\frac{m}{2\pi d^2}} \sum_{k=0}^{m} \binom{m}{k} \left(\frac{-1}{d^2}\right)^k \int K^{(2k)}(u) \hat{\psi}_s(u)\, du$$

$$= \int \bar{K}(u) \hat{\psi}_s(u)\, du$$

where

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

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 $\gamma_m = 1 + o(1)$. Hence

$$\|\bar{K}\|_\infty \leq s\frac{m}{2\pi\sigma d} \sum_{k=0}^{m} \binom{m}{k} \left(\frac{\alpha m}{k}\right)^{2k} (2k)! = O(\alpha^m m^m).$$

It can be seen by comparing the results that if ψ_s can be estimated at a standard polynomial rate then the balance between variance and bias is achieved when

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

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

$$\hat{f}(s) - f(s) = O\left(n^{-\alpha \log \log n / \log n}\right),$$

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

Since f cannot be really estimated, one may ask whether the model of logarithmic useless? Surprisingly not.

Although f cannot be estimated per-se many of its functionals can be estimated quite easily.

For example, as mentioned, its moments. Similarly $\psi(u)$ considered as the a simple linear functional can be estimated quite easily.

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. In fact, these are two aspects of one phenomena. The shape of the observable ψ hardly depends on f, and essentially depends only on a few aspects of f, which 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 f the less you are about ψ.

Suppose that f is supported by some compact interval (a, b). Then one can approximate

$$\psi(u) = \sum_{i=1}^{m} \beta_i^0 u^i + R_m(u)$$

$$\beta_j^0 = \int \xi^{j-1} f(\xi)\, d\xi / j!$$

$$0 \leq R_m(u) = \frac{1}{(m+1)!} \psi^{m+1}(\tilde{u}) \leq \frac{b^m e^{ub}}{(m+1)!}.$$

That is, very few βs are needed, but they reveal very little.

We assume that in our disposal there is an estimate $\hat{\psi} = \hat{\psi}_n$. Let $\Sigma_{ij}(u_1, \dots, u_k) = \operatorname{cov}\{\hat{\psi}(u_i), \hat{\psi}(u_j)\}$ for any $u_1, \dots, u_k > 0$. We assume

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 $\max_i |E\psi(u_i) - \psi(u_i)|^2 = \log n/n$.

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 assumption is valid assuming 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 relationship 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 the vector $Y = \{\hat{\psi}(u_1), \ldots, \hat{\psi}(u_k)\}^{\mathsf{T}}$ with 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 of $\tilde{\psi}$ is almost the parametric rate even if $\hat{\psi}$ achieves only a non-parametric rate.

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

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

$$k^{-1}\mathrm{E}\sum_{i=1}^{k}\{\tilde{\psi}(u_i) - \psi(u_i)\}^2 = k^{-1}\mathrm{E}\{\varepsilon^{\mathsf{T}} Z(Z^{\mathsf{T}}Z)^{-1}Z^{\mathsf{T}}\varepsilon\}$$

$$= k^{-1}\,\mathrm{trace}\{Z(Z^{\mathsf{T}}Z)^{-1}Z^{\mathsf{T}}\mathrm{E}(\varepsilon\varepsilon^{\mathsf{T}})\}.$$

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

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

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

A1. Assume that for some c, d and each ε there are $h_{\varepsilon,1}, \dots, h_{\varepsilon,M(\varepsilon)}$ 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 $\beta_j = \int \gamma_j(\xi) f(\xi)\, d\xi$, $j = 1, \dots, M(\varepsilon)$, such that $\max_{c<u<d} |\psi(u) - \sum_{j=1}^{M(\varepsilon)} \beta_j h_j(u)| < \varepsilon$.

Theorem *Suppose Assumptions* A*1* *and* A*1* *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 and 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. Will it work?

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 A1 , and by a similar argument, for a vector β on the simplex

$$\sup_u \left| \sum_{j=1}^{M(\varepsilon)} \beta_j g(u; \xi_j) - \sum_{j=1}^{M(\varepsilon)} \beta_j \sum_{l=1}^{M(\varepsilon)} \gamma_l(\xi_j) h_l(u) \right| \leq \varepsilon.$$

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

The starting equation for the DAX problem can be rewritten as

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

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

$$p(x) = \frac{q(x) \int \frac{\partial}{\partial u} g(U(x;f);\xi) f(\xi)\, d\mu(\xi)}{\int q\{y\}\left(\int \frac{\partial}{\partial u} g(U(y;f);\xi) f(\xi)\, d\mu(\xi)\right\} dy}$$

$$= \frac{q(x)\psi_f'\{\psi_f^{-1}(x)\}}{\int q(y)\psi_f'\{\psi_f^{-1}(y)\}\, dy}$$

The statistical model assumed by DHM is that we obtain a simple random sample from p, parametrized by the high dimensional parameter f. A natural approach is to estimate f by the MLE or a variant of it.

The density of the U_i can be easily found:

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

The model of random sample from the density p can be well approximated as $\sigma \longrightarrow 0$ by a measurement error model: $U_1, \ldots, U_n$ are sampled from r_f, and observed only indirectly through

$$X_i = \psi_f(U_i) + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

independent.

Now, the log-likelihood of the joint density of $(X_1, Y_1), \ldots, (X_n, Y_n)$ 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 to replacing U_i by $\psi_f^{-1}(X_i)$.

The EM algorithm, therefore, iterates therefore between

The E step:

$$U_i \leftarrow \psi_f^{-1}(X_i), \quad i = 1, \ldots, n,$$

and

The M step:

$$f \leftarrow \arg\max \left[\sum_{i=1}^{n} \left\{ \log q\{\psi_f(U_i)\} + 2\log\{\psi'_f(U_i)\} \right\} - nC_f \right].$$

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:

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

We consider the following approximation of the EM:

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

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

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