Semi-Nonparametric Identification and Estimation of the Stochastic Frontier Model

Herman J. Bierens

Pennsylvania State University, USA

Hung-Pin Lai

National Chung Cheng University, Taiwan

The paper I am going to present is:

- Bierens, H. J., and H-P. Lai (2019), ”Semi-Nonparametric Identification and Estimation of the Stochastic Frontier Model”, working paper in progress, downloadable from http://www.personal.psu.edu/hxb11/SNP_SF_MODELS.PDF

This paper is incomplete, in that the actual estimation of the stochastic frontier model on simulated and actual data is still pending.

However, the theoretical parts are complete, and will be presented in this talk.

Introduction

The linear stochastic frontier (SF) model takes the form

$$Y_i = \alpha_0 + X_i'\beta_0 + V_i - W_i$$

for firm i, where V_i is an error term with the usual properties, X_i is the vector of regressors, and W_i is an unobserved non-negative random variable measuring the distance from the production frontier.

For example, suppose that the original model is a Cobb-Douglas production function, i.e.,

$$Q_i = \exp(-W_i)\exp(V_i)\exp(\alpha_0)K_i^{\beta_{0,1}}L_i^{\beta_{0,2}} = \exp(-W_i)\overline{Q}_i$$

where

$$\overline{Q}_i = \exp(V_i)\exp(\alpha_0)K_i^{\beta_{0,1}}L_i^{\beta_{0,2}}$$

is the production frontier of firm i, and

$$\exp(-W_i) = Q_i/\overline{Q}_i$$

represents the relative distance from the production frontier of firm i.

Taking logs, the linear SF model follows:

$$\begin{aligned} Y_i &= \ln(Q_i) = \alpha_0 + \beta_{0,1}\ln(K_i) + \beta_{0,2}\ln(L_i) + V_i - W_i \\ &= \alpha_0 + X_i'\beta_0 + V_i - W_i \end{aligned}$$

Throughout we assume the following standard conditions:

Assumption

- *The model variables* X_i, V_i *and* W_i *are independent of each other, and are jointly i.i.d. as* (X, V, W) *across observations . The variables* Y_i *and* X_i *are observed for* $i = 1, 2, ..., N$.
- $X \in \mathbb{R}^p$, $E[X'X] < \infty$ *and* $\Sigma = \text{Var}(X)$ *is nonsingular.*
- V *is distributed as* $\mathcal{N}(0, \sigma_0^2)$.
- $\Pr[W < 0] = 0$, $\Pr[W \leq w] > 0$ *for all* $w > 0$, *and* $E[W^2] < \infty$.

Now write the SF model as

$$\begin{aligned} Y_i &= (\alpha_0 - E[W]) + X_i'\beta_0 + V_i - (W_i - E[W]) \\ &= \mu_0 + X_i'\beta_0 + U_i,\ i = 1, 2, ..., N, \end{aligned}$$

where $\mu_0 = \alpha_0 - E[W]$ and $U_i = V_i - W_i + E[W]$.

As is well-known, under the aforementioned conditions the OLS estimators $\widehat{\beta}_N$ of β_0 and $\widehat{\mu}_N$ of μ_0 are strongly consistent, i.e.,

$$\Pr\left[\lim_{N\to\infty} \widehat{\beta}_N = \beta_0\right] = 1,\ \Pr\left[\lim_{N\to\infty} \widehat{\mu}_N = \mu_0\right] = 1$$

(also denoted by $\widehat{\beta}_N \overset{\text{a.s.}}{\longrightarrow} \beta_0$, $\widehat{\mu}_N \overset{\text{a.s.}}{\longrightarrow} \mu_0$, as $N \to \infty$, where ”a.s.” stands for ”almost surely”).

The main problems I will address in this talk are:

- Are the parameters α_0 and σ_0^2 and the unknown distribution of W semi-nonparametrically identified?
- If so, how can we estimate the parameters α_0 and σ_0^2 and the distribution function $F_0(w) = \Pr[W \leq w]$ of W?

Semi-nonparametric identification

Suppose that next to the SF model

$$Y = \alpha_0 + X'\beta_0 + V - W$$

there exists an alternative model

$$Y = \alpha + X'\beta_0 + V_* - W_*,$$

where $V_* \sim N(0, \sigma^2)$, W_* is a nonnegative random variable satisfying

$$\Pr[W_* \leq w] > 0 \text{ for all } w > 0,$$

and X, V_* and W_* are independent, such that both models are observational equivalent, in the sense that they generate the same conditional distribution of Y given X.

This is only possible if

$$\alpha_0 + V - W \sim \alpha + V_* - W_*,$$

which implies that for all $t > 0$,

$$E\left[\exp(-t(W - V - \alpha_0))\right] \equiv E\left[\exp(-t(W^* - V^* - \alpha))\right].$$

For $t > 0$ we can write, by the independence of V and W,
$E\left[\exp(-t(W - V - \alpha_0))\right] = \exp(t.\alpha_0)\exp(\sigma_0^2 t^2/2)E[\exp(-t.W)]$
where
$$\exp(\sigma_0^2 t^2/2) = E[\exp(t.V)]$$
is the well-known moment generating function of $V \sim N(0, \sigma_0^2)$,
and $E[\exp(-t.W)]$ is known as the Laplace transform of W.

Similarly, we have
$E\left[\exp(-t(W^* - V^* - \alpha))\right] = \exp(t.\alpha)\exp(\sigma^2 t^2/2)E[\exp(-t.W_*)]$

Thus, for $t > 0$,
$\exp(t.\alpha_0)\exp(\sigma_0^2 t^2/2)E[\exp(-t.W)] \equiv \exp(t.\alpha)\exp(\sigma^2 t^2/2)E[\exp(-t.W_*)]$

This equality is equivalent to
$E\left[\exp(-t.W)\right] \equiv \exp(t.(\alpha-\alpha_0)+(\sigma^2-\sigma_0^2)t^2/2).E\left[\exp(-t.W_*)\right],\ t > 0.$

If $\sigma^2 > \sigma_0^2$ then for arbitrary $K > 0$,

$$\begin{aligned}
&E\left[\exp(-t.W)\right] \\
&\equiv \exp(t.(\alpha - \alpha_0) + (\sigma^2 - \sigma_0^2)t^2/2).E\left[\exp(-t.W_*)\right] \\
&\geq \exp\left(t.(\alpha - \alpha_0) + (\sigma^2 - \sigma_0^2)t^2/2\right).E\left[\exp(-t.W_*).I(W_* \leq K)\right] \\
&\geq \exp\left(t.(\alpha - \alpha_0 - K) + (\sigma^2 - \sigma_0^2)t^2/2\right).\Pr\left[W_* \leq K\right],
\end{aligned}$$

where $I(.)$ is the well-known indicator function.

Since $\Pr\left[W_* \leq K\right] > 0$, the right hand side converges to ∞ as $t \to \infty$, so that there exists a $t_0 > 0$ such that

$$E[\exp(-t_0.W)] > 1.$$

But this is clearly impossible, and so is $\sigma^2 > \sigma_0^2$.

By a similar argument, $\sigma^2 < \sigma_0^2$ leads to the contradiction $E[\exp(-t_0.W_*)] > 1$ for some $t_0 > 0$, so that

$$\sigma^2 = \sigma_0^2.$$

Now we have
$E\left[\exp(-t.W)\right] \equiv \exp(t.(\alpha-\alpha_0)).E\left[\exp(-t.W_*)\right],\ t > 0.$

Suppose that $\alpha > \alpha_0$. Then for $0 < K < \alpha - \alpha_0$,

$$\begin{aligned} E\left[\exp(-t.W)\right] &\geq \exp(t.(\alpha-\alpha_0)).E\left[\exp(-t.W_*)I(W_* \leq K)\right] \\ &\geq \exp\left(t.(\alpha-\alpha_0-K)\right).\Pr\left[W_* \leq K\right]. \end{aligned}$$

Since $\alpha-\alpha_0-K > 0$, this inequality implies that there exists a $t_0 > 0$ such that $E[\exp(-t_0.W)] > 1$, so that $\alpha > \alpha_0$ is not possible, and by a similar argument, $\alpha < \alpha_0$ leads to the contradiction $E[\exp(-t_0.W_*)] > 1$ for some $t_0 > 0$, so that

$$\alpha = \alpha_0.$$

Consequently,

$$E\left[\exp(-t.W)\right] \equiv E\left[\exp(-t.W_*)\right],\ t \geq 0$$

which implies that

$$W \sim W^*.$$

The conclusion $W \sim W^*$ follows from the fact that distributions of nonnegative random variables are equal if and only their Laplace transforms are equal on an arbitrary open subset of $[0, \infty)$.

Summarizing, the following identification results hold.

Theorem. *Let $\mathcal{F}_+$ be the collection of all distribution functions F of nonnegative random variables satisfying $F(w) > 0$ for $w > 0$, and let $F_0(w)$ be the distribution function of the inefficiency variable W in the linear SF model .*
Suppose that the latter model is observational equivalent to the alternative model

$$Y = \alpha + X'\beta_0 + V_* - W_*,$$

where X, V_* and W_* are independent, $V_* \sim \mathcal{N}(0, \sigma^2)$, and the c.d.f. F_* of W_* belongs to $\mathcal{F}_+$.
Then $\alpha = \alpha_0$, $\sigma^2 = \sigma_0^2$ and $F_* = F_0$.

Note that without the condition that the c.d.f. F_* of W_* belongs to $\mathcal{F}_+$ we can only conclude that for some constant $c > 0$,

$$\alpha = c + \alpha_0,\ \sigma^2 = \sigma_0^2 \text{ and } W_* \sim c + W.$$

Of course this c cancels out in the alternative model.

Apart from the condition $F(w) > 0$ for $w > 0$ there are no further restrictions on $F \in \mathcal{F}_+$, so that this identification result applies to nonnegative discrete, continuous and mixed discrete-continuous distributions for W.

However, from now onwards it will be assumed that

Assumption. *The distribution of W is absolutely continuous with continuous density $f_0(w)$ and support $(0, \infty)$ or $[0, \infty)$.*

Hilbert spaces of functions

In order to model the unknown density $f_0(w)$ of W we need a little bit of Hilbert space theory.

A Hilbert space $\mathcal{H}$ is a vector space endowed with an innerproduct, denoted by $\langle x, y\rangle$ for $x, y \in \mathcal{H}$, and associated norm $||x|| = \sqrt{\langle x, x\rangle}$ and metric $||x - y||$, such that every Cauchy sequence in $\mathcal{H}$ takes a limit in $\mathcal{H}$.

Recall that a Cauchy sequence x_n is a sequence such that $\lim_{\min(k,m)\to\infty} ||x_m - x_k|| = 0$, which in a Hilbert space $\mathcal{H}$ implies that $\lim_{n\to\infty} ||x_n - x|| = 0$ for some $x \in \mathcal{H}$.

For example, the Euclidean space $\mathbb{R}^k$ is a vector space endowed with the innerproduct $\langle x, y\rangle = x'y$ and associated norm $||x|| = \sqrt{x'x}$ and metric $||x - y||$, and it is (or should be) well-known that any Cauchy sequence in $\mathbb{R}^k$ converges to a limit in $\mathbb{R}^k$.

It is often possible to define different innerproducts on the same vector space, but each version should mimic the properties of the innerproduct $\langle x, y\rangle = x'y$ in $\mathbb{R}^k$:

Definition. *An inner product on a real vector space* $\mathcal{V}$ *is a real function* $\langle x, y\rangle$: $\mathcal{V} \times \mathcal{V} \to \mathbb{R}$ *such that for all* x, y, z *in* $\mathcal{V}$ *and all* c *in* $\mathbb{R}$,

(*1*) $\langle x, y\rangle = \langle y, x\rangle$

(*2*) $\langle c.x, y\rangle = c\,\langle x, y\rangle$

(*3*) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$

(*4*) $\langle x, x\rangle > 0$ *if and only if* $x \neq 0$.

Hilbert spaces of functions

For a given density $w(x)$ on (a, b), where $-\infty \leq a < b \leq \infty$ and $w(x) > 0$ on (a, b), consider the space $L^2(w)$ of real functions $f(x)$ on (a, b) satisfying

$$\int_a^b f(x)^2 w(x)\mathrm{d}x < \infty.$$

Endow the space $L^2(w)$ with the innerproduct

$$\langle f, g\rangle = \int_a^b f(x)g(x)w(x)\mathrm{d}x$$

and associated norm $||f|| = \sqrt{\langle f, f\rangle}$ and metric $||f - g||$.

For any Cauchy sequence $f_n \in L^2(w)$ it can be shown (but this is not easy!) that there exists a function $f(x)$ satisfying $\int_a^b f(x)^2 w(x)\mathrm{d}x < \infty$, so that $f \in L^2(w)$, such that $\lim_{n\to\infty} \|f_n - f\| = 0$. Thus, $L^2(w)$ is a Hilbert space.

Complete orthonormal sequences

Let $\{\varphi_j(x)\}_{j=0}^{\infty}$ be an **orthonormal** sequence in $L^2(w)$:

$$\langle \varphi_i, \varphi_j \rangle = \int_a^b \varphi_i(x)\,\varphi_j(x)\,w(x)\mathrm{d}x = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The linear projection of $f(x)$ on $\{\varphi_j(x)\}_{j=0}^{n}$ is defined as

$$f_n(x) = \sum_{j=0}^{n} \gamma_j \varphi_j(x), \text{ where } ||f - f_n||^2 \text{ is minimal.}$$

Note that

$$\begin{aligned} ||f - f_n||^2 &= \int_a^b \left(f(x) - \sum_{j=0}^{n} \gamma_j \varphi_j(x) \right)^2 w(x)\mathrm{d}x \\ &= \int_a^b f(x)^2 w(x)\mathrm{d}x - 2\sum_{j=0}^{n} \gamma_j \langle f, \varphi_j \rangle + \sum_{j=0}^{n} \gamma_j^2 \end{aligned}$$

where the last equality is due to the orthonormality of $\{\varphi_j(x)\}_{j=0}^{\infty}$.

Hence, $||f - f_n||^2$ is minimal for

$$\gamma_j = \langle f, \varphi_j \rangle = \int_a^b f(x)\varphi_j\,(x)\,w(x)\mathrm{d}x,\ \sum_{j=0}^{n} \gamma_j^2 \leq \int_a^b f(x)^2 w(x)\mathrm{d}x < \infty.$$

The γ_j's involved are called the Fourier coefficients of $f(x)$.

An orthonormal sequence $\{\varphi_j(x)\}_{j=0}^{\infty}$ in $L^2\,(w)$ is called **complete** if for an arbitrary $f \in L^2\,(w)$,

$$\lim_{n\to\infty} ||f - f_n||^2 = \lim_{n\to\infty} \int_a^b (f(x) - f_n(x))^2\, w(x)\mathrm{d}x = 0,$$

where

$$f_n(x) = \sum_{j=0}^{n} \gamma_j \varphi_j(x) \text{ with } \gamma_j = \langle f, \varphi_j \rangle\,.$$

Then

$$f(x) = \lim_{n\to\infty} f_n(x) = \sum_{j=0}^{\infty} \gamma_j \varphi_j(x) \text{ a.e. on } (a, b),$$

where ”a.e.” stands for ”almost everywhere”.

The latter means that $f(x) = \lim_{n\to\infty} f_n(x)$ pointwise in $x \in (a, b)$, possibly except on a set $N_0 \subset (a, b)$ for which $\int_{N_0} 1.\mathrm{d}x = 0$.

A few examples of complete orthonormal sequences

Hermite polynomials

In the case that $w(x)$ is the standard normal density,

$$w_{N(0,1)}(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi},\ x \in \mathbb{R},$$

the Hermite polynomials form a complete orthonormal sequence in the corresponding Hilbert space $L^2(w_{N(0,1)})$.

The Hermite polynomials $\varphi_k\left(x\right)$, $k \geq 0$, on $\mathbb{R}$ can be generated recursively by the three terms recursive relation (TTRR)

$$\sqrt{k+1}\varphi_{k+1}(x) - x.\varphi_k(x) + \sqrt{k}\varphi_{k-1}(x) = 0, \;\; k \geq 1,$$

starting from $\varphi_0(x) = 1$, $\varphi_1(x) = x$.

Laguerre polynomials
In the case that $w(x)$ is the standard exponential density,

$$w_{\exp}(x) = \exp(-x) \text{ for } x \geq 0,\ w_{\exp}(x) = 0 \text{ for } x < 0,$$

the Laguerre polynomials form a complete orthonormal sequence in the corresponding Hilbert space $L^2(w_{\exp})$.

The Laguerre polynomials $\varphi_k\left(x\right),\, k \geq 0$, on $\left[0, \infty\right)$ can be generated recursively by the TTRR

$$(k{+}1)\varphi_{k+1}(x){+}(2k+1-x)\,\varphi_k(x){+}k.\varphi_{k-1}(x) = 0,\ k \geq 1,$$

starting from $\varphi_0(x) = 1$, $\varphi_1(x) = x - 1$.

However, the Laguerre polynomials also have closed form expressions:

$$\varphi_m(x) = \sum_{\ell=0}^{m} \binom{m}{\ell} \frac{(-1)^\ell}{\ell!} x^\ell.$$

More generally, for any density $w(x)$ on (a,b), where $-\infty \leq a < b \leq \infty$, and $w(x) > 0$ on (a,b), satisfying

$$\int_a^b |x|^n w(x) \mathrm{d}x < \infty \text{ for all } n \in \mathbb{N},$$

with expectation μ_w and variance σ_w^2, there exist unique sequences r_k and s_k, $k = 0, 1, 2, ...,$ such that the polynomials $\varphi_k(x)$ generated by the TTRR

$$r_{k+1}\varphi_{k+1}(x) + (s_k - x)\,\varphi_k(x) + r_k.\varphi_{k-1}(x) = 0, \quad k \geq 1,$$
$$\varphi_0(x) = 1, \ \varphi_1(x) = (x - \mu_w)/\sqrt{\sigma_w^2},$$

form a complete orthonormal sequence in $L^2(w)$.

SNP modeling of densities of nonnegative random variables

From now on we will focus on the exponential density $w_{\text{exp}}(x) = \exp(-x)$, $x \geq 0$, for w, for which the Laguerre polynomials

$$\varphi_0(x) \equiv 1,\ \varphi_m(x) = \sum_{\ell=0}^{m} \binom{m}{\ell} \frac{(-1)^\ell}{\ell!} x^\ell,\ x \geq 0 \text{ for } m \in \mathbb{N},$$

form a complete orthonormal sequence in $L^2(w_{\text{exp}})$.

For any density $f(x)$ on $[0, \infty)$ the function

$$g(x) = \sqrt{f(x)} / \sqrt{\exp(-x)},\ x \geq 0$$

is an element of the Hilbert space $L^2(w_{\text{exp}})$, because

$$\int_0^\infty g(x)^2 \exp(-x) \mathrm{d}x = \int_0^\infty f(x) \mathrm{d}x = 1.$$

As we have seen before,

$$g(x) = \sqrt{f(x)}/\sqrt{\exp(-x)}$$

can be approximated arbitrarily close by

$$g_n(x) = \sum_{m=0}^{n} \gamma_m \varphi_m(x),$$

$$\text{where } \gamma_m = \langle g, \varphi_m \rangle = \int_0^\infty g(x)\varphi_m(x)\exp(-x)\mathrm{d}x$$

in the sense that

$$\lim_{n\to\infty} ||g - g_n||^2 = \lim_{n\to\infty} \int_0^\infty (g(x) - g_n(x))^2 \exp(-x)\mathrm{d}x = 0.$$

This implies that

$$f(x) = \lim_{n\to\infty} g_n(x)^2 \exp(-x) = \left(\sum_{m=0}^{\infty} \gamma_m \varphi_m(x)\right)^2 \exp(-x) \text{ a.e. on } [0, \infty)$$

where $\sum_{m=0}^{\infty} \gamma_m^2 = 1$.

The condition $\sum_{m=0}^{\infty} \gamma_m^2 = 1$ follows from

$$
\begin{aligned}
1 &= \int_0^{\infty} f(x)\mathrm{d}x = \int_0^{\infty} \exp(-x) \left(\sum_{m=0}^{\infty} \gamma_m \varphi_m(x) \right)^2 \mathrm{d}x \\
&= \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \gamma_m \gamma_k \int_0^{\infty} \exp(-x)\varphi_m(x)\varphi_k(x)\mathrm{d}x = \sum_{m=0}^{\infty} \gamma_m^2
\end{aligned}
$$

where the last equality follows from the orthonormality of the Laguerre polynomials, i.e.,

$$
\int_0^{\infty} \exp(-x)\varphi_m(x)\varphi_k(x)\mathrm{d}x = I(m = k).
$$

The condition $\sum_{m=0}^{\infty} \gamma_m^2 = 1$ can be imposed by reparametrizing the Fourier coefficients γ_m as follows.

Since $\varphi_0(x) \equiv 1$, it follows that

$$\gamma_0 = \int_0^\infty \exp(-x)\sqrt{f(x)/\exp(-x)}\mathrm{d}x \in (0,1)$$

The condition $\sum_{m=0}^\infty \gamma_m^2 = 1$ can now be implemented by reparametrizing the γ_m's as

$$\gamma_0 = \frac{1}{\sqrt{1+\sum_{k=1}^\infty \delta_k^2}},\ \gamma_m = \frac{\delta_m}{\sqrt{1+\sum_{k=1}^\infty \delta_k^2}},\ m \in \mathbb{N},$$

where $\sum_{k=1}^\infty \delta_k^2 < \infty$.

Thus, every density $f(x)$ on $[0,\infty)$ has the SNP representation

$$f(x) = \exp(-x)\frac{\left(1+\sum_{m=1}^\infty \delta_m\varphi_m(x)\right)^2}{1+\sum_{k=1}^\infty \delta_k^2} \text{ a.e. on } [0,\infty)$$

where $\varphi_m(x)$ is the Laguerre polynomial of order m, and $\sum_{k=1}^\infty \delta_k^2 < \infty$.

This result can be more formally restated and extended as follows:

Theorem

(a) *Denote*

$$\Delta = \left\{ \boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty} \delta_m^2 < \infty \right\}.$$

Endow this space with the innerproduct

$$\langle \boldsymbol{\delta}_1, \boldsymbol{\delta}_2 \rangle = \sum_{m=1}^{\infty} \delta_{1,m}\delta_{2,m} \; for \; \boldsymbol{\delta}_i = \{\delta_{i,m}\}_{m=1}^{\infty} \in \Delta$$

and associated norm $||\boldsymbol{\delta}|| = \sqrt{\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle}$ *and metric* $||\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2||$.

Then Δ *becomes a Hilbert space.*

(b) *For any* $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} \in \Delta$, *denote*

$$f(w|\boldsymbol{\delta}) = \exp(-w)\frac{\left(1 + \sum_{m=1}^{\infty} \delta_m \varphi_m(w)\right)^2}{1 + \sum_{k=1}^{\infty} \delta_k^2},\ w \geq 0.$$

where

$$\varphi_m(w) = \sum_{\ell=0}^{m} \binom{m}{\ell} \frac{(-1)^{\ell}}{\ell!} w^{\ell},\ w \geq 0,$$

is the Laguerre polynomial of order m.

$(b.1)$ *For any density* $f(w)$ *on* $[0, \infty)$ *there exist possibly uncountable many* $\boldsymbol{\delta} \in \Delta$ *such that* $f(w) = f(w|\boldsymbol{\delta})$ *a.e. on* $[0, \infty)$, *hence* $\int_0^{\infty} |f(w) - f(w|\boldsymbol{\delta})| \mathrm{d}w = 0$.

$(b.2)$ *On the other hand, if* $f(w)$ *is continuous and everywhere positive on* $(0, \infty)$ *then such a* $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} \in \Delta$ *is unique, with*

$$\delta_m = \frac{\int_0^{\infty} \varphi_m(w) \exp(-w/2)\sqrt{f(w)}\mathrm{d}w}{\int_0^{\infty} \exp(-w/2)\sqrt{f(w)}\mathrm{d}w},\ m \in \mathbb{N}.$$

(c) *For any pair* $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2 \in \Delta$,

$$\int_0^{\infty} |f(w|\boldsymbol{\delta}_1) - f(w|\boldsymbol{\delta}_2)| \, \mathrm{d}w \leq 2||\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2||^2 + 4||\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2||.$$

Remarks.

Part (a) is a standard Hilbert space result.

For the proof of part (b), see Theorems 16 and 21 in:

- Bierens, H. J. (2014): "The Hilbert Space Theoretical Foundation of Semi-Nonparametric Modeling". Chapter 1 in: J. Racine, L. Su and A. Ullah (eds), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press.

For part (c), see Bierens and Lai (2019, Lemma 2).

The density

$$f(w|\boldsymbol{\delta}) = \exp(-w)\frac{\left(1+\sum_{m=1}^{\infty}\delta_m\varphi_m(w)\right)^2}{1+\sum_{k=1}^{\infty}\delta_k^2},\ w \geq 0,$$

with

$$\varphi_m(w) = \sum_{\ell=0}^{m}\binom{m}{\ell}\frac{(-1)^\ell}{\ell!}w^\ell,\ w \geq 0,$$

will now be used as the SNP specification of the density of an absolutely continuous distribution on $[0,\infty)$.

Moreover, since it was previously assumed that the density $f_0(w)$ of the actual inefficiency variable W in the SNP-SF model under review is continuous and positive on $(0,\infty)$, it follows from part $(b.2)$ of the aforementioned theorem that:

There exists a unique $\boldsymbol{\delta}^0 \in \Delta$ *such that for the true density* $f_0(w)$ *of* W, $f_0(w) = f(w|\boldsymbol{\delta}^0)$ *a.e. on* $[0,\infty)$.

In SNP estimation applications we cannot deal with infinite-dimensional parameters directly, and neither can we in the paper under review.

Therefore, the SNP density $f(w|\boldsymbol{\delta})$ will be used in truncated form, as

$$f(w|\pi_n\boldsymbol{\delta}) = \exp(-w)\frac{\left(1+\sum_{m=1}^{n}\delta_m\varphi_m(w)\right)^2}{1+\sum_{k=1}^{n}\delta_k^2},\ w \geq 0,$$

were π_n is the truncation operator, i.e.,

The operator π_n applied to $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty}$ as $\pi_n\boldsymbol{\delta}$ replaces all the δ_m's for $m > n$ by zeros.

Then by part (c) of the aforementioned theorem,

$$\lim_{n\to\infty}\int_0^{\infty}|f(w|\boldsymbol{\delta}) - f(w|\pi_n\boldsymbol{\delta})|\,\mathrm{d}w$$
$$\leq 2\lim_{n\to\infty}||\boldsymbol{\delta}-\pi_n\boldsymbol{\delta}||^2 + 4\lim_{n\to\infty}||\boldsymbol{\delta}-\pi_n\boldsymbol{\delta}|| = 0.$$

The test model and its parameters

In order check how our theoretical results work in practice, we have generated a random sample of size $N = 500$ from the data-generating process

$$Y = 1 + 0.8X_1 + 0.3X_2 + V - W,$$

which corresponds to the log of a Cobb-Douglas production function with increasing returns to scale.

The variables involved are generated as follows:

$$X_1 = \sqrt{5}U_1,\ X_2 = \frac{3U_2 - U_1}{\sqrt{2}},\ V = 0.5U_3,$$

where $(U_1, U_2, U_3)' \sim \mathcal{N}_3 0, I_3)$,

so that var(X_1) = var$(X_2) = 5$, cov$(X_1, X_2) = -2.5$, and corr$(X_1, X_2) = -0.5$.
The latter allows for a realistic amount of collinearity.

The distribution of W is specified as Gamma$(3, \gamma_0)$, i.e., the density of W is

$$f_0(w) = \frac{w^2 \exp(-w/\gamma_0)}{2\gamma_0^3}, \; \gamma_0 > 0, \; w \geq 0.$$

Note that W can be generated as

$$W \sim \gamma_0 \sum_{j=1}^{3} \ln(1/(1 - U_j^*)),$$

where the U_j^*'s are random drawings from the uniform $[0, 1]$ distribution.

The parameter γ_0 is chosen such that

$$\text{var}(W) = \text{var}(V) = \sigma_0^2 = 0.25,$$

which is the case for

$$\gamma_0 = 1/(2\sqrt{3}).$$

Moreover, it follows from the well-known moment generating function of the Gamma density $f_0(w)$ that its Laplace transform takes the form

$$\mathcal{L}_0(t) = \int_0^\infty \exp(-t.w) f_0(w) \mathrm{d}w = (1+\gamma_0 t)^{-3} = \left(1 + t/(2\sqrt{3})\right)^{-3}, \; t \geq 0.$$

Thus, the true parameters are

$$\beta_0 = (\beta_{0,1}, \beta_{0,2})' = (0.8, 0.3)', \; \alpha_0 = 1, \; \sigma_0^2 = \text{var}(V) = \text{var}(W) = 0.25.$$

Furthermore, $E[W] = 3.\gamma_0 = 0.5\sqrt{3}$, so that the true regression intercept is

$$\mu_0 = \alpha_0 - E[W] = 1 - 0.5\sqrt{3} \approx 0.1339746.$$

The OLS estimation results are not relevant for the current purpose, and will therefore not be displayed, except that $R^2 = 0.8584$.

As we have seen before, the Gamma density $f_0(w)$ of W has the series representation

$$f(w|\boldsymbol{\delta}^0) = \exp(-w)\frac{\left(1+\sum_{m=1}^{\infty}\delta_{0,m}\varphi_m(w)\right)^2}{1+\sum_{k=1}^{\infty}\delta_{0,k}^2} = f_0(w) \text{ a.e.},$$

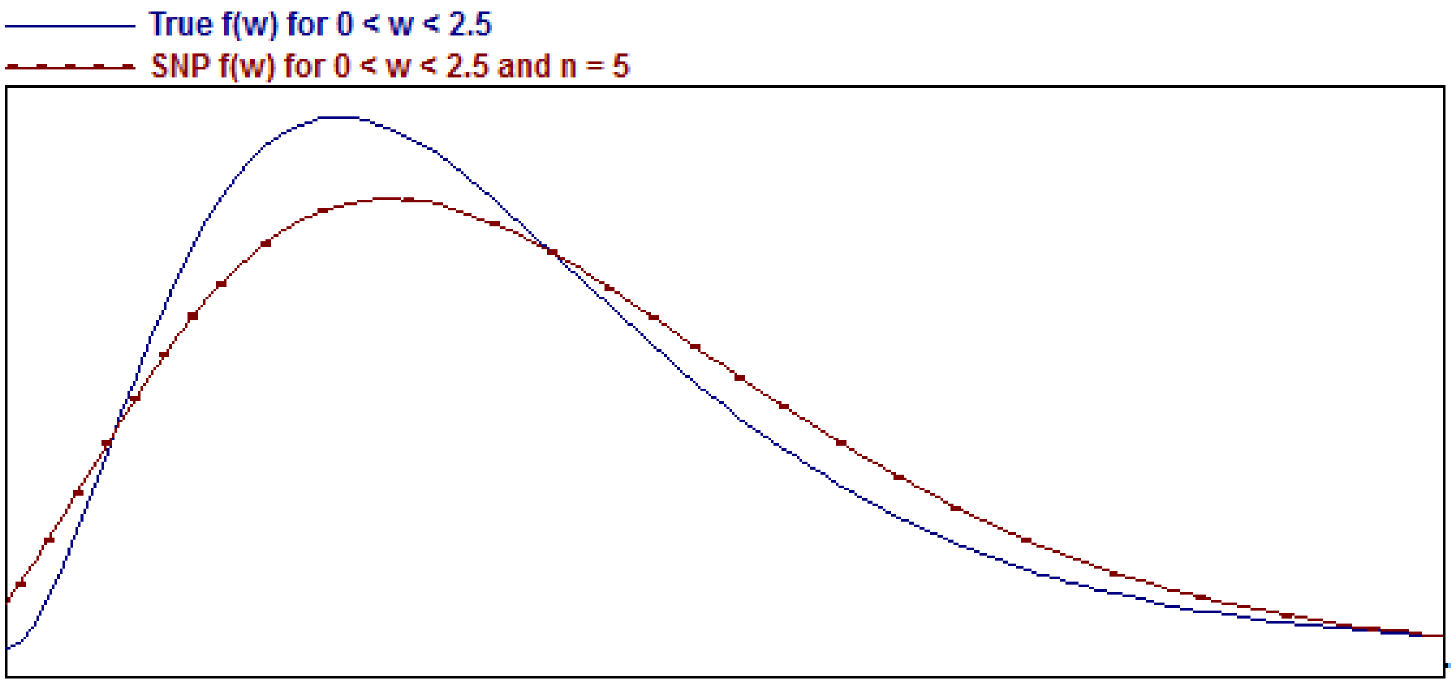
where the $\varphi_m(w)$'s are the Laguerre polynomials, and $\boldsymbol{\delta}^0 = \{\delta_{0,m}\}_{m=1}^{\infty} \in \Delta$ is unique.

In this case the $\delta_{0,m}$'s can be computed exactly as,

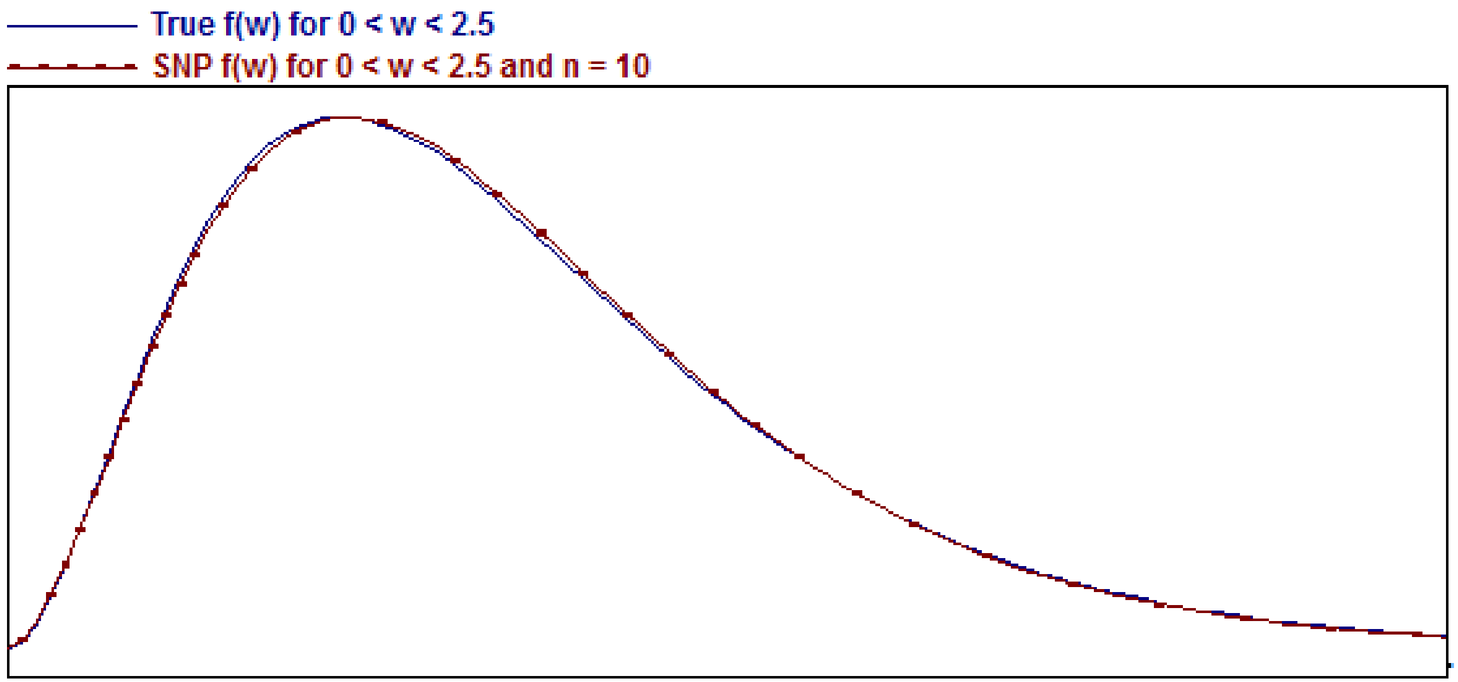
$$\begin{aligned}
\delta_{0,m} &= \sum_{\ell=0}^{m}(-1)^{\ell}\binom{m}{\ell}\frac{\ell+1}{\lambda_0^{\ell}},\ m \in \mathbb{N}, \\
\lambda_0 &= (1+1/\gamma_0)/2 = 0.5+\sqrt{3}.
\end{aligned}$$

The fit of the SNP density

The following two pictures compare the actual Gamma density $f_0(w)$ with $f(w|\pi_n\boldsymbol{\delta}^0)$, for $n = 5, 10$.



True f(w) for 0 < w < 2.5
SNP f(w) for 0 < w < 2.5 and n = 5


True f(w) for 0 < w < 2.5
SNP f(w) for 0 < w < 2.5 and n = 10

The empirical objective function

Recall from the linear SF model that

$$X_i'\widehat{\beta}_N - Y_i + ||X_i|| = X_i'(\widehat{\beta}_N - \beta_0) + ||X_i|| + W_i - V_i - \alpha_0.$$

The reason for adding the term $||X_i||$ to both sides of this equation is that

$$X_i'(\widehat{\beta}_N - \beta_0) + ||X_i|| \geq (1 - ||\widehat{\beta}_N - \beta_0||)||X_i|| \geq 0 \text{ a.s.}$$

for sufficient large N.

Then by the uniform strong law of large numbers and $||\widehat{\beta}_N - \beta_0|| \xrightarrow{\text{a.s.}} 0$,

$$\sup_{0 \leq t \leq c} \left| \frac{1}{N} \sum_{i=1}^{N} \exp\left(-t\left(X_i'\widehat{\beta}_N - Y_i + ||X_i||\right)\right) \right.$$
$$\left. - E\left[\exp\left(-t\left(||X|| + W - V - \alpha_0\right)\right)\right] \right| \xrightarrow{\text{a.s.}} 0, \text{ for any } c > 0$$

Since

$$E\left[\exp\left(-t\left(||X||+W-V-\alpha_0\right)\right)\right]$$
$$= E[\exp(-t.||X||)]\exp(t.\alpha_0)\exp(\sigma_0^2 t^2/2)E[\exp(-t.W)]$$

and

$$\sup_{0\le t\le c}\left|\frac{1}{N}\sum_{i=1}^{N}\exp\left(-t.||X_i||\right)-E[\exp(-t.||X||)]\right|\overset{\text{a.s.}}{\to}0$$

the first part of the following key result follows.

Theorem. *Denote* $Z_{i,N}=X_i'\widehat{\beta}_N-Y_i$ *and*

$$\Upsilon_N(t)=\ln\left(\sum_{i=1}^{N}\exp\left(-t.(Z_{i,N}+||X_i||)\right)\right)-\ln\left(\sum_{i=1}^{N}\exp(-t.||X_i||)\right).$$

Then for any constant $c>0$*,*

$$\sup_{t\in[0,c]}\left|\Upsilon_N(t)-\alpha_0 t-\sigma_0^2 t^2/2-\ln(\mathcal{L}_0(t))\right|\overset{\text{a.s.}}{\to}0,$$

where

$$\mathcal{L}_0(t)=E[\exp(-t.W)]$$

is the Laplace transform of W.

Moreover, it can be shown that for any constant $c > 0$*,*

$$\sup_{t\in[0,c]} \left|\Upsilon'_N(t) - \alpha_0 - \sigma_0^2 t - \mathrm{d}\ln(\mathcal{L}_0(t))/\mathrm{d}t\right| \xrightarrow{\text{a.s.}} 0,$$

$$\sup_{t\in[0,c]} \left|\Upsilon''_N(t) - \sigma_0^2 - \mathrm{d}^2\ln(\mathcal{L}_0(t))/(\mathrm{d}t)^2\right| \xrightarrow{\text{a.s.}} 0.$$

The first part of this theorem suggests to base our estimation approach on the empirical objective function

$$\begin{aligned} Q_N(\alpha,\sigma^2,\boldsymbol{\delta}|c) &= \int_0^c \left(\Upsilon_N(t) - \alpha t - \sigma^2 t^2/2 - \ln(\mathcal{L}(t|\boldsymbol{\delta}))\right)^2 \mathrm{d}t \\ &= \int_0^c \left(\Psi_N(t|\boldsymbol{\delta}) - \alpha t - \sigma^2 t^2/2\right)^2 \mathrm{d}t, \end{aligned}$$

for some chosen constant $c > 0$, where

$$\mathcal{L}(t|\boldsymbol{\delta}) = \int_0^\infty \exp(-t.w) f(w|\boldsymbol{\delta})\mathrm{d}w$$

is the Laplace transform of $f(w|\boldsymbol{\delta})$, and

$$\Psi_N(t|\boldsymbol{\delta}) = \Upsilon_N(t) - \ln(\mathcal{L}(t|\boldsymbol{\delta})).$$

Then pointwise in $\alpha, \sigma^2, \boldsymbol{\delta}$,

$$Q_N(\alpha, \sigma^2, \boldsymbol{\delta}|c) \stackrel{\text{a.s.}}{\to} Q(\alpha, \sigma^2, \boldsymbol{\delta}|c), \text{ where}$$

$$Q(\alpha, \sigma^2, \boldsymbol{\delta}|c) = \int_0^c \left(\Lambda(t|\boldsymbol{\delta}) - (\alpha - \alpha_0)t - (\sigma^2 - \sigma_0^2)t^2/2\right)^2 \mathrm{d}t,$$

with $\Lambda(t|\boldsymbol{\delta}) = \ln(\mathcal{L}(t|\boldsymbol{\delta}^0)) - \ln(\mathcal{L}(t|\boldsymbol{\delta}))$

Note that $\mathcal{L}(t|\boldsymbol{\delta}^0) = \mathcal{L}_0(t)$ is the Laplace transform of the true density $f_0(w) = f(w|\boldsymbol{\delta}^0)$ a.s. of the actual inefficiency variable W in the SF model.

It follows trivially from $\Lambda(t|\boldsymbol{\delta}^0) \equiv 0$ that $Q(\alpha, \sigma^2, \boldsymbol{\delta}^0|c) = 0$ if and only if $\alpha = \alpha_0$ and $\sigma^2 = \sigma_0^2$.

However, it follows from Theorem 3 in the paper under review that more generally, $Q(\alpha, \sigma^2, \boldsymbol{\delta}|c) = 0$ if and only if $\alpha = \alpha_0$, $\sigma^2 = \sigma_0^2$ **and** $\boldsymbol{\delta} = \boldsymbol{\delta}^0$.

The fit of $\Upsilon_N(t)$

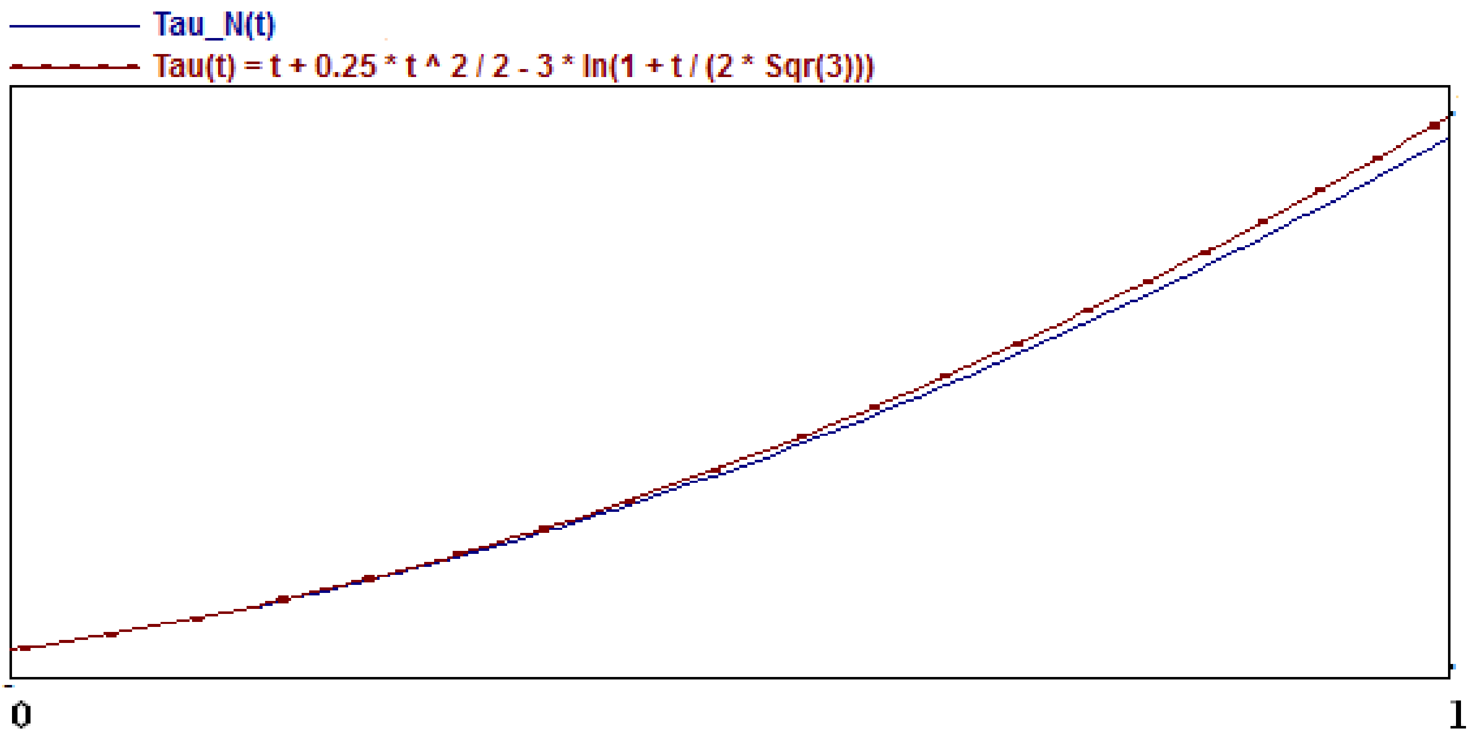
For the test data, $\alpha_0 = 1, \sigma_0^2 = 0.25$ and $\mathcal{L}_0(t) = \big(1 + t/(2\sqrt{3})\big)^{-3}$, so that in this case

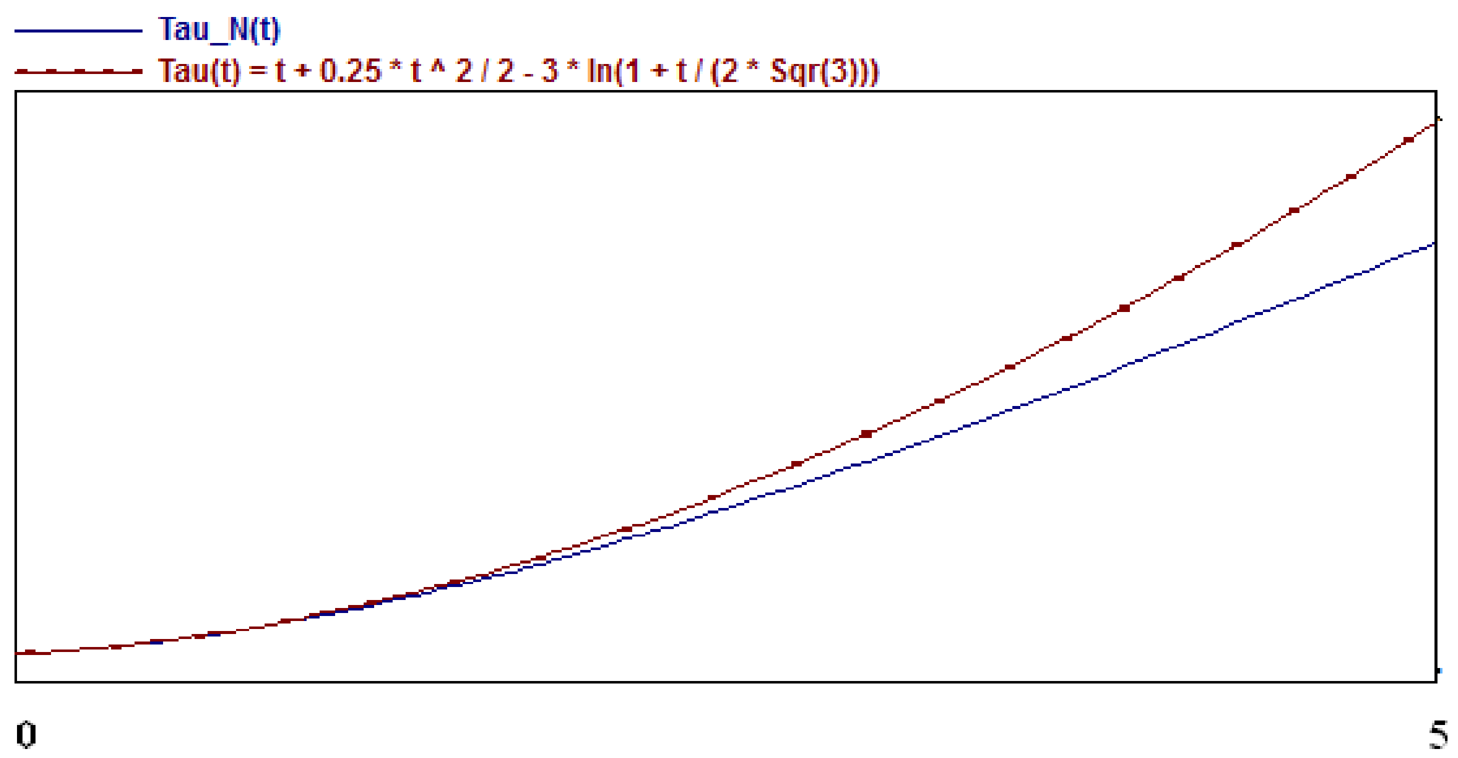
$$\sup_{t\in[0,c]} |\Upsilon_N(t) - \Upsilon(t)| \overset{\text{a.s.}}{\longrightarrow} 0$$

where

$$\Upsilon(t) = t + 0.25t^2/2 - 3.\ln\Big(1 + t/(2\sqrt{3})\Big)$$

To see how close $\Upsilon_N(t)$ gets to $\Upsilon(t)$ on $[0, c]$ for $N = 500$, we compare $\Upsilon_N(t)$ with $\Upsilon(t)$ for $c = 1$ and $c = 5$.

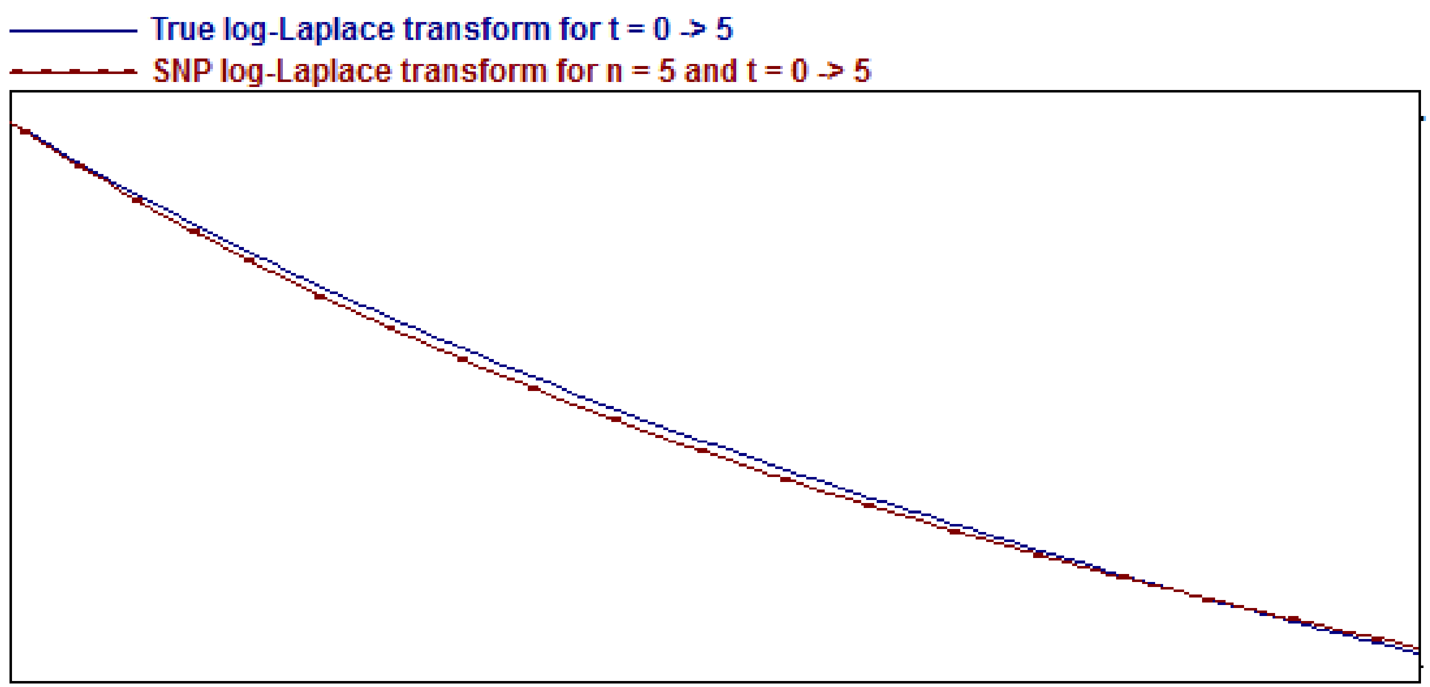

Tau_N(t)
Tau(t) = t + 0.25 * t ^ 2 / 2 - 3 * ln(1 + t / (2 * Sqr(3)))
0
1

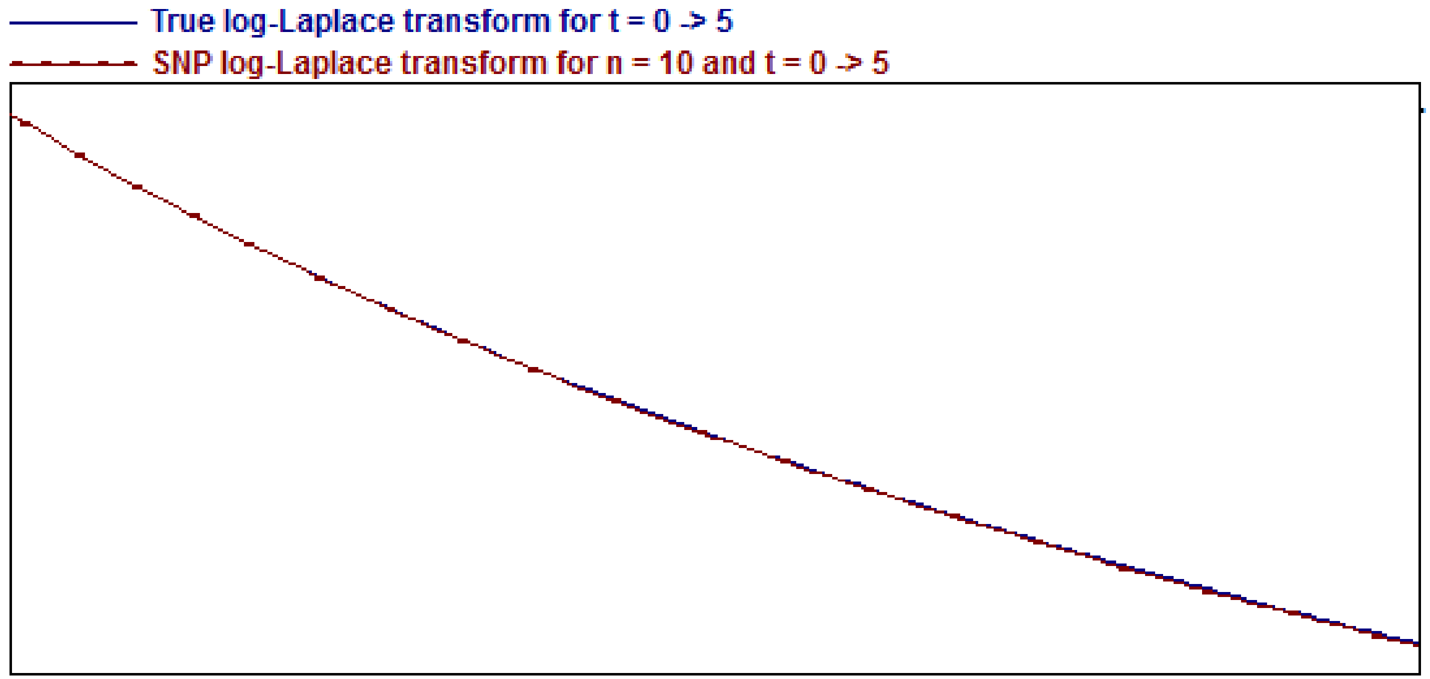


Tau_N(t)
Tau(t) = t + 0.25 * t ^ 2 / 2 - 3 * ln(1 + t / (2 * Sqr(3)))
0
5

The fit of $\ln(\mathcal{L}(t|\pi_n\boldsymbol{\delta}^0))$

Next, we compare $\ln(\mathcal{L}_0(t)) = 3.\ln(1+t/(2\sqrt{3}))$ with $\ln(\mathcal{L}(t|\pi_n\boldsymbol{\delta}^0))$ on $[0,5]$, for $n = 5, 10$.


True log-Laplace transform for t = 0 -> 5
SNP log-Laplace transform for n = 5 and t = 0 -> 5


True log-Laplace transform for t = 0 -> 5
SNP log-Laplace transform for n = 10 and t = 0 -> 5

Concentrating α and σ^2 out

There are various way to construct random functions $\alpha_N(\boldsymbol{\delta}|c)$ and $\sigma_N^2(\boldsymbol{\delta}|c)$ on Δ such that pointwise in $\boldsymbol{\delta}$,

$$\alpha_N(\boldsymbol{\delta}|c) \stackrel{\text{a.s.}}{\to} \alpha(\boldsymbol{\delta}|c), \text{ with } \alpha(\boldsymbol{\delta}^0|c) \;=\; \alpha_0,$$
$$\sigma_N^2(\boldsymbol{\delta}|c) \stackrel{\text{a.s.}}{\to} \sigma^2(\boldsymbol{\delta}|c), \text{ with } \sigma^2(\boldsymbol{\delta}^0|c) \;=\; \sigma_0^2.$$

Then instead of

$$Q_N(\alpha, \sigma^2, \boldsymbol{\delta}|c) = \int_0^c \left(\Psi_N(t|\boldsymbol{\delta}) - \alpha t - \sigma^2 t^2/2\right)^2 \mathrm{d}t,$$

we may use the concentrated objective function

$$\underline{Q}_N(\boldsymbol{\delta}|c) = Q_N(\alpha_N(\boldsymbol{\delta}|c), \sigma_N^2(\boldsymbol{\delta}|c), \boldsymbol{\delta}|c)$$

to estimate $\boldsymbol{\delta}^0$.

Given a consistent estimator $\widehat{\boldsymbol{\delta}}_N$ of $\boldsymbol{\delta}^0$ (to be determined shortly) we then can estimate α_0 and σ_0^2 consistently by $\alpha_N(\widehat{\boldsymbol{\delta}}_N|c)$ and $\sigma_N^2(\widehat{\boldsymbol{\delta}}_N|c)$, respectively.

One way to construct $\alpha_N(\boldsymbol{\delta}|c)$ and $\sigma_N^2(\boldsymbol{\delta}|c)$ is to solve

$$\left(\alpha_N(\boldsymbol{\delta}|c), \sigma_N^2(\boldsymbol{\delta}|c)\right)' = \arg \min_{(\alpha,\sigma^2)'\in\mathbb{R}^2} Q_N(\alpha, \sigma^2, \boldsymbol{\delta}|c),$$

with solutions

$$\begin{aligned}
\alpha_N(\boldsymbol{\delta}|c) &= 48c^{-3}\int_0^c t.\Psi_N(t|\boldsymbol{\delta})\mathrm{d}t - 60c^{-4}\int_0^c t^2\Psi_N(t|\boldsymbol{\delta})\mathrm{d}t, \\
\sigma_N^2(\boldsymbol{\delta}|c) &= 160c^{-5}\int_0^c t^2\Psi_N(t|\boldsymbol{\delta})\mathrm{d}t - 120c^{-4}\int_0^c t.\Psi_N(t|\boldsymbol{\delta})\mathrm{d}t,
\end{aligned}$$

where $\Psi_N(t|\boldsymbol{\delta}) = \Upsilon_N(t) - \ln(\mathcal{L}(t|\boldsymbol{\delta}))$, and pointwise limits

$$\begin{aligned}
\alpha(\boldsymbol{\delta}|c) &= \alpha_0 + 48c^{-3}\int_0^c t.\Lambda(t|\boldsymbol{\delta})\mathrm{d}t - 60c^{-4}\int_0^c t^2\Lambda(t|\boldsymbol{\delta})\mathrm{d}t, \\
\sigma^2(\boldsymbol{\delta}|c) &= \sigma_0^2 + 160c^{-5}\int_0^c t^2\Lambda(t|\boldsymbol{\delta})\mathrm{d}t - 120c^{-4}\int_0^c t.\Lambda(t|\boldsymbol{\delta})\mathrm{d}t,
\end{aligned}$$

respectively, where $\Lambda(t|\boldsymbol{\delta}) = \ln(\mathcal{L}(t|\boldsymbol{\delta}^0)) - \ln(\mathcal{L}(t|\boldsymbol{\delta}))$.

Since $\Lambda(t|\boldsymbol{\delta}^0) \equiv 0$, it follows that $\alpha(\boldsymbol{\delta}^0|c) = \alpha_0$ and $\sigma^2(\boldsymbol{\delta}^0|c) = \sigma_0^2$.

How well, or bad, do $\alpha_N(\boldsymbol{\delta}^0|c)$ and $\sigma_N^2(\boldsymbol{\delta}^0|c)$ fit?

For the test data we have compared $\alpha_N(\pi_n\boldsymbol{\delta}^0|c)$ and $\sigma_N^2(\pi_n\boldsymbol{\delta}^0|c)$ with $\alpha_0 = 1$ and $\sigma_0^2 = 0.25$, respectively.

Table 1. Test of $\alpha_N(\pi_n\boldsymbol{\delta}^0|c)$ and $\sigma_N^2(\pi_n\boldsymbol{\delta}^0|c)$ for $N = 500$

$c = 1$			$c = 2$		
n	$\alpha_N(\pi_n\boldsymbol{\delta}^0\|c)$	$\sigma_N^2(\pi_n\boldsymbol{\delta}^0\|c)$	n	$\alpha_N(\pi_n\boldsymbol{\delta}^0\|c)$	$\sigma_N^2(\pi_n\boldsymbol{\delta}^0\|c)$
5	0.93079083	−0.25351275	5	0.90788127	−0.20027573
10	1.00113264	−0.37074911	10	0.94305218	−0.22958482
15	1.00182725	−0.37190678	15	0.94339948	−0.22987424
20	1.00183129	−0.37191351	20	0.94340150	−0.22987592
$c = 3$			$c = 4$		
n	$\alpha_N(\pi_n\boldsymbol{\delta}^0\|c)$	$\sigma_N^2(\pi_n\boldsymbol{\delta}^0\|c)$	n	$\alpha_N(\pi_n\boldsymbol{\delta}^0\|c)$	$\sigma_N^2(\pi_n\boldsymbol{\delta}^0\|c)$
5	0.88443425	−0.16865993	5	0.85968683	−0.14494912
10	0.90788152	−0.18168619	10	0.87727229	−0.15227639
15	0.90811305	−0.18181482	15	0.87744594	−0.15234875
20	0.90811440	−0.18181557	20	0.87744695	−0.15234917

Although the results for $\alpha_N(\pi_n\boldsymbol{\delta}^0|c)$ are not too bad, it is clear that $\sigma_N^2(\pi_n\boldsymbol{\delta}^0|c)$ is useless in samples of size $N = 500$.

One of the reasons why $\sigma_N^2(\pi_n\boldsymbol{\delta}^0|c)$ performs so badly is that for the test data, $\Upsilon(t) = \lim_{N\to\infty}\Upsilon_N(t)$ and $\Upsilon_N(t)$ veer apart with t, where $\Upsilon_N(t) < \Upsilon(t)$, so that

$$\Psi_N(t|\boldsymbol{\delta}^0) = \Upsilon_N(t) - \ln(\mathcal{L}(t|\boldsymbol{\delta}^0)) < \Upsilon(t) - \ln(\mathcal{L}(t|\boldsymbol{\delta}^0)) = t + 0.25t^2/2,$$

where the gap between $t + 0.25t^2/2$ and $\Psi_N(t|\boldsymbol{\delta}^0)$ increases with t.

Moreover, in view of the plots of $\Upsilon_N(t)$ and the fact that $\ln(\mathcal{L}(t|\boldsymbol{\delta}^0) < 0$, we may conclude that for the test data,

$$\Psi_N(t|\boldsymbol{\delta}^0) > 0.$$

Now recall that

$$
\begin{aligned}
\sigma_N^2(\boldsymbol{\delta}^0|c) &= 160c^{-5}\int_0^c t^2\Psi_N(t|\boldsymbol{\delta}^0)\mathrm{d}t - 120c^{-4}\int_0^c t.\Psi_N(t|\boldsymbol{\delta}^0)\mathrm{d}t \\
&= 160c^{-2}\int_0^1 (u-3/4)u.\Psi_N(c.u|\boldsymbol{\delta}^0)\mathrm{d}u \\
&= -160c^{-2}\int_0^{3/4} |u-3/4|.u.\Psi_N(c.u|\boldsymbol{\delta}^0)\mathrm{d}u \\
&\quad +160c^{-2}\int_{3/4}^1 |u-3/4|u.\Psi_N(c.u|\boldsymbol{\delta}^0)\mathrm{d}u
\end{aligned}
$$

Apparently, for the test data the first integral dominates the second, which makes $\sigma_N^2(\boldsymbol{\delta}^0|c)$ negative, and the large factor 160 makes things worse.

Nevertheless, the asymptotic result $\sigma_N^2(\boldsymbol{\delta}^0|c) \stackrel{\text{a.s.}}{\to} \sigma_0^2$ is correct, but we need a much larger sample than $N = 500$ to get $\sigma_N^2(\boldsymbol{\delta}^0|c)$ satisfactory close to σ_0^2.

An alternative approach to concentrate α and σ^2 out

An alternative approach can be based on the results

$$\sup_{t\in[0,c]} \left|\Upsilon_N'(t) - \alpha_0 - \sigma_0^2 t - \mathrm{d}\ln(\mathcal{L}_0(t))/\mathrm{d}t\right| \xrightarrow{\text{a.s.}} 0,$$

$$\sup_{t\in[0,c]} \left|\Upsilon_N''(t) - \sigma_0^2 - \mathrm{d}^2\ln(\mathcal{L}_0(t))/(\mathrm{d}t)^2\right| \xrightarrow{\text{a.s.}} 0,$$

Using the notations

$\Psi_N(t|\boldsymbol{\delta}) = \Upsilon_N(t) - \ln(\mathcal{L}(t|\boldsymbol{\delta}))$, $\Lambda(t|\boldsymbol{\delta}) = \ln(\mathcal{L}(t|\boldsymbol{\delta}^0)) - \ln(\mathcal{L}(t|\boldsymbol{\delta}))$,

these results imply that for any small $\tau \in (0, c)$,

$$\sup_{t\in[\tau,c]} \left|\Psi_N'(t|\boldsymbol{\delta}) - \alpha_0 - \sigma_0^2 t - \Lambda'(t|\boldsymbol{\delta})\right| \xrightarrow{\text{a.s.}} 0,$$

$$\sup_{t\in[\tau,c]} \left|\Psi_N''(t|\boldsymbol{\delta}) - \sigma_0^2 - \Lambda''(t|\boldsymbol{\delta})\right| \xrightarrow{\text{a.s.}} 0.$$

Combining these results yield

$$\sup_{t\in[\tau,c]} |(\Psi'_N(t|\boldsymbol{\delta}) - t.\Psi''_N(t|\boldsymbol{\delta})) - \alpha_0 - (\Lambda'(t|\boldsymbol{\delta}) - t.\Lambda''(t|\boldsymbol{\delta}))| \overset{\text{a.s.}}{\longrightarrow} 0,$$

$$\sup_{t\in[\tau,c]} \left|\Psi''_N(t|\boldsymbol{\delta}) - \sigma_0^2 - \Lambda''(t|\boldsymbol{\delta})\right| \overset{\text{a.s.}}{\longrightarrow} 0,$$

The reason for this τ is to keep t away from zero, because

$$\Psi''_N(0|\boldsymbol{\delta}) \overset{\text{a.s.}}{\longrightarrow} \sigma_0^2 + \text{var}(W) - \left(\int_0^\infty w^2 f(w|\boldsymbol{\delta})\mathrm{d}w - \left(\int_0^\infty w f(w|\boldsymbol{\delta})\mathrm{d}w\right)^2\right),$$

where the expression between brackets is the variance of the density $f(w|\boldsymbol{\delta})$.

However if $f(w|\boldsymbol{\delta})$ is a density with infinite first or second moments then this variance is infinite, so that then

$$\Psi''_N(0|\boldsymbol{\delta}) \overset{\text{a.s.}}{\longrightarrow} -\infty.$$

Similarly,

$$\Psi'_N(0|\boldsymbol{\delta}) \xrightarrow{\text{a.s.}} \alpha_0 - E[W] + \int_0^\infty w f(w|\boldsymbol{\delta})\mathrm{d}w$$

so that

$$\Psi'_N(0|\boldsymbol{\delta}) \xrightarrow{\text{a.s.}} \infty \text{ if } \int_0^\infty w f(w|\boldsymbol{\delta})\mathrm{d}w = \infty.$$

Recall that for the test data we have used the values $\alpha_0 = 1$ and $\sigma_0^2 = 0.25$.

In order to check how close

$$\Psi''_N(t|\pi_n\boldsymbol{\delta}^0) \text{ and } \Psi'_N(t|\pi_n\boldsymbol{\delta}^0) - t.\Psi''_N(t|\pi_n\boldsymbol{\delta}^0)$$

are to $\sigma_0^2 = 0.25$ and $\alpha_0 = 1$, respectively, we have computed the maxima and minima of these functions over 100 grid points, $t_i = i/100$, $i = 1, ..., 100$, and for $n = 5, 10, 15, 20$.

Table 2. Range of $\Psi_N''(t|\pi_n\boldsymbol{\delta}^0)$ for $N = 500$, $\sigma_0^2 = 0.25$

n	$\max_{0.01\le t\le 1} \Psi_N''(t\|\pi_n\boldsymbol{\delta}^0)$	$\min_{0.01\le t\le 1} \Psi_N''(t\|\pi_n\boldsymbol{\delta}^0)$
5	0.19454618	−0.74815102
10	0.23227388	0.16484178
15	0.23060300	0.19117993
20	0.23053105	0.19109597

These results are based on the choice $c = 1$. We have done some further experiments for larger c, up to $c = 4$, but the results are the same as in this table.

Thus, it seems that $\max_{t\in[\tau,c]} \Psi_N''(t|\boldsymbol{\delta}^0)$ invariant for $c > 1$.

These results suggests to use for some integer $K > 2$,

$$\overline{\sigma}^2_{N,K}(\boldsymbol{\delta}) = \max_{i=1,2,\ldots,K} \Psi''_N(t_i|\boldsymbol{\delta}), \text{ where } t_t = i/K,$$

as an alternative for the previous $\sigma^2_N(\boldsymbol{\delta}|c)$.

Then pointwise in $\boldsymbol{\delta} \in \Delta$,

$$\overline{\sigma}^2_{N,K}(\boldsymbol{\delta}) \stackrel{\text{a.s.}}{\longrightarrow} \overline{\sigma}^2_K(\boldsymbol{\delta})$$

where

$$\overline{\sigma}^2_K(\boldsymbol{\delta}) = \sigma^2_0 + \max_{i=1,2,\ldots,K} \Lambda''(t_i|\boldsymbol{\delta}), \ \overline{\sigma}^2_K(\boldsymbol{\delta}^0) = \sigma^2_0.$$

Moreover, both $\overline{\sigma}^2_{N,K}(\boldsymbol{\delta})$ and $\overline{\sigma}^2_K(\boldsymbol{\delta})$ are continuous in each $\boldsymbol{\delta} \in \Delta$.

Table 3. Range of $\Phi_N(t|\pi_n\boldsymbol{\delta}^0)$ for $N = 500$, $\alpha_0 = 1,$ where $\Phi_N(t|\boldsymbol{\delta}) = \Psi'_N(t|\boldsymbol{\delta}) - t.\Psi''_N(t|\boldsymbol{\delta})$

n	$\max_{0.01\le t\le 1}\Phi_N(t\|\pi_n\boldsymbol{\delta}^0)$	$\min_{0.01\le t\le 1}\Phi_N(t\|\pi_n\boldsymbol{\delta}^0)$
5	1.1130532	1.0560260
10	1.0283804	0.9977377
15	1.0263746	0.9964439
20	1.0263778	0.9964403

These maxima and minima are surprisingly close to $\alpha_0 = 1$, even for $n = 5$, but the minima are slightly closer.

Therefore, we propose to use

$$\overline{\alpha}_{N,K}(\boldsymbol{\delta}) = \min_{i=1,2,\ldots,K}\left(\Psi'_N(t_i|\boldsymbol{\delta}) - t_i.\Psi''_N(t_i|\boldsymbol{\delta})\right), \text{ where } t_i = i/K,$$

as an alternative for the previous $\alpha_N(\boldsymbol{\delta}|c)$.

Then pointwise in $\boldsymbol{\delta} \in \Delta$,

$$\overline{\alpha}_{N,K}(\boldsymbol{\delta}) \xrightarrow{\text{a.s.}} \overline{\alpha}_K(\boldsymbol{\delta}),$$

where

$$\overline{\alpha}_K(\boldsymbol{\delta}) = \alpha_0 + \min_{i=1,2,\ldots,K} \left(\Lambda'(t_i|\boldsymbol{\delta}) - t_i.\Lambda''(t_i|\boldsymbol{\delta})\right), \; \overline{\alpha}_K(\boldsymbol{\delta}^0) = \alpha_0.$$

Again, both $\overline{\alpha}_{N,K}(\boldsymbol{\delta})$ and $\overline{\alpha}_K(\boldsymbol{\delta})$ are continuous in each $\boldsymbol{\delta} \in \Delta$.

Integrated method of moments sieve estimation

Recall that the original objective function is

$$Q_N(\alpha, \sigma^2, \boldsymbol{\delta}|c) = \int_0^c \left(\Psi_N(t|\boldsymbol{\delta}) - \alpha.t - \sigma^2 t^2/2\right)^2 \mathrm{d}t$$

For notational convenience I will assume that we have chosen $c = 1$, and write

$$Q_N(\alpha, \sigma^2, \boldsymbol{\delta}) = \int_0^1 \left(\Psi_N(t|\boldsymbol{\delta}) - \alpha.t - \sigma^2 t^2/2\right)^2 \mathrm{d}t$$

For the same reason I will suppress the dependence of $\overline{\alpha}_{N,K}(\boldsymbol{\delta})$ and $\overline{\sigma}^2_{N,K}(\boldsymbol{\delta})$ and their limits on K and write

$$\overline{\alpha}_N(\boldsymbol{\delta}) = \overline{\alpha}_{N,K}(\boldsymbol{\delta}),\ \overline{\sigma}^2_N(\boldsymbol{\delta}) = \overline{\sigma}^2_{N,K}(\boldsymbol{\delta}),\ \overline{\alpha}(\boldsymbol{\delta}) = \alpha_K(\boldsymbol{\delta}),\ \overline{\sigma}^2(\boldsymbol{\delta}) = \sigma^2_K(\boldsymbol{\delta})$$

Plugging in $\overline{\alpha}_N(\boldsymbol{\delta})$ for α and $\overline{\sigma}_N^2(\boldsymbol{\delta})$ for σ^2 yields the concentrated objective function

$$\underline{Q}_N(\boldsymbol{\delta}) = Q_N(\overline{\alpha}_N(\boldsymbol{\delta}), \overline{\sigma}_N^2(\boldsymbol{\delta}), \boldsymbol{\delta})$$

with pointwise limit

$$\underline{Q}(\boldsymbol{\delta}) = \int_0^1 \left(\Lambda(t|\boldsymbol{\delta}) + (\alpha_0 - \overline{\alpha}(\boldsymbol{\delta}))\,.t + \left(\sigma_0^2 - \overline{\sigma}^2(\boldsymbol{\delta})\right) t^2/2\right)^2 \mathrm{d}t$$

Note that $\underline{Q}(\boldsymbol{\delta}^0) = 0$, and that both $\underline{Q}_N(\boldsymbol{\delta})$ and $\underline{Q}(\boldsymbol{\delta})$ are continuous on Δ.

One might think of estimating $\boldsymbol{\delta}^0$ by

$$\arg\min_{\boldsymbol{\delta}\in\Delta} \underline{Q}_N(\boldsymbol{\delta}).$$

However, since Δ is infinite dimensional, this problem has infinitely many solutions, and none of these solutions will be consistent.

A solution for this problem is *sieve estimation*.

Sieve estimation

The idea of sieve estimation is to construct an increasing sequence of finite dimensional compact subsets Δ_n, $n \in \mathbb{N}$, of Δ, called sieve spaces, such that

$$\overline{\cup_{n=1}^{\infty}\Delta_n} = \Delta,$$

where the bar denotes the closure.

For example, let

$$\Delta_n = \left\{ \boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{n} \delta_m^2 \leq M_n,\ \delta_m = 0 \text{ for } m > n \right\},$$

where M_n is an a priori chosen strictly increasing positive sequence converging (slowly) to ∞ as $n \to \infty$.

Next, let $n_N \in \mathbb{N}$ be *any* nondecreasing subsequence of the sample size N satisfying

$$\lim_{N\to\infty} n_N = \infty,\ \lim_{N\to\infty} n_N/N = 0.$$

Then

$$\widehat{\boldsymbol{\delta}}_N = \arg \min_{\boldsymbol{\delta} \in \Delta_{n_N}} \underline{Q}_N(\boldsymbol{\delta})$$

is a sieve estimator of $\boldsymbol{\delta}^0$.

Consistency of sieve estimators

The question arises: Is $\widehat{\boldsymbol{\delta}}_N$ a consistent estimator of $\boldsymbol{\delta}^0$, in the sense that $p \lim_{N \to \infty} ||\widehat{\boldsymbol{\delta}}_N - \boldsymbol{\delta}^0|| = 0$?

The consistency of sieve estimators is well-established in the sieve estimation literature. See

- Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models." In Heckman, J. & E. Leamer (eds.), *Handbook of Econometrics, Vol. 6*, Ch. 76. Elsevier,

for a review.

The sieve consistency proof in Chen (2007, Theorem 3.1) employs the condition that, in the present case,

$$p\lim_{N\to\infty}\sup_{\boldsymbol{\delta}\in\Delta_{n_N}}\left|\underline{Q}_N(\boldsymbol{\delta})-\underline{Q}(\boldsymbol{\delta})\right|=0.$$

However, this condition is difficult to verify.
Moreover, under this condition the consistency of $\widehat{\boldsymbol{\delta}}_N$ follows almost trivially, so that in essence the consistency of $\widehat{\boldsymbol{\delta}}_N$ is assumed rather than proved.

In

- Bierens, H. J. (2014), "Consistency and Asymptotic Normality of Sieve ML Estimators Under Low-Level Conditions", *Econometric Theory* 30, 1021-1076, Theorem 4.1,

I have provided an alternative consistency proof, which requires that there exists an infinite dimensional *compact* subset Δ_c of Δ containing $\boldsymbol{\delta}^0$ in its interior such that

$$\lim_{N\to\infty}\Pr[\widehat{\boldsymbol{\delta}}_N\in\Delta_c]=1.$$

Also this condition is difficult to verify, but this problem can be solved by reducing Δ to an infinite dimensional compact parameter space, which is possible.

In the SNP-SF case under review the consistency of $\widehat{\boldsymbol{\delta}}_N$ can be proved much easier, and without additional conditions, due to the particular form of $Q_N(\alpha, \sigma^2, \boldsymbol{\delta})$ as an integral of a squared difference, by which it can be shown that

Lemma. *For any* $(\alpha, \sigma^2, \boldsymbol{\delta}) \in \mathbb{R} \times [0, \infty) \times \Delta$,

$$\begin{aligned} Q_N(\alpha, \sigma^2, \boldsymbol{\delta}) &\leq \sqrt{Q(\alpha, \sigma^2, \boldsymbol{\delta})} \left(\sqrt{Q(\alpha, \sigma^2, \boldsymbol{\delta})} + R_N \right), \\ &\geq \sqrt{Q(\alpha, \sigma^2, \boldsymbol{\delta})} \left(\sqrt{Q(\alpha, \sigma^2, \boldsymbol{\delta})} - R_N \right), \\ \textit{where } R_N &= 2\sqrt{Q_N(\alpha_0, \sigma_0^2, \boldsymbol{\delta}^0)} \stackrel{\text{a.s.}}{\to} 0. \end{aligned}$$

This result plays a key role in proving the consistency of $\widehat{\boldsymbol{\delta}}_N$.

Theorem. *The sieve estimator* $\widehat{\boldsymbol{\delta}}_N = \arg\min_{\boldsymbol{\delta}\in\Delta_{n_N}} \underline{Q}_N(\boldsymbol{\delta})$ *is weakly consistent, in the sense that*

$$p\lim_{N\to\infty} ||\widehat{\boldsymbol{\delta}}_N - \boldsymbol{\delta}^0|| = 0.$$

This result implies that

$$\widehat{\alpha}_N = \overline{\alpha}_N(\widehat{\boldsymbol{\delta}}_N) \xrightarrow{\text{p}} \alpha_0,\ \widehat{\sigma}^2_N = \overline{\sigma}^2_N(\widehat{\boldsymbol{\delta}}_N) \xrightarrow{\text{p}} \sigma^2_0,$$

and

$$\int_0^\infty |f(w|\widehat{\boldsymbol{\delta}}_N) - f_0(w)|\mathrm{d}w \leq 2||\boldsymbol{\delta}^0 - \widehat{\boldsymbol{\delta}}_N||^2 + 4||\boldsymbol{\delta}^0 - \widehat{\boldsymbol{\delta}}_N|| \xrightarrow{\text{p}} 0.$$

Numerical issues

Computation of the objective function

The objective function $Q_N(\alpha, \sigma^2, \boldsymbol{\delta})$ in the case $c = 1$ involves three types of integrals, namely

- $\int_0^1 \Psi_N(t|\boldsymbol{\delta})^2 \mathrm{d}t$
- $\int_0^1 t\Psi_N(t|\boldsymbol{\delta}) \mathrm{d}t$,
- $\int_0^1 t^2\Psi_N(t|\boldsymbol{\delta}) \mathrm{d}t$,

where $\Psi_N(t|\boldsymbol{\delta}) = \Upsilon_N(t) - \ln(\mathcal{L}(t|\boldsymbol{\delta}))$.

Due to the logs in $\Upsilon_N(t)$ and $\ln(\mathcal{L}(t|\boldsymbol{\delta}))$ these integrals do not have closed form expressions.

Therefore we propose to compute these integrals by the Legendre-Gauss quadrature.

The Legendre-Gauss quadrature

One of the standard numerical integration procedures is the Legendre-Gauss quadrature [LGQ hereafter] to approximate integrals of the type $\int_{-1}^{1} \phi(x)\mathrm{d}x$, for square integrable real functions $\phi(x)$ on $[-1, 1]$, by

$$\int_{-1}^{1} \phi(x)\mathrm{d}x \approx \sum_{\ell=1}^{L} w_\ell \phi(x_\ell)$$

for some $L \geq 2$, where the abscissas x_ℓ and the weights $w_\ell > 0$ depend on L but not on ϕ.

Therefore, for each L the pairs (w_ℓ, x_ℓ) can be tabulated. These tables are given, for example, in the link

https://pomax.github.io/bezierinfo/legendre-gauss.html

for $L = 2, 3, ..., 64$.

Moreover, the LGQ approximation is exact if $\phi(x)$ itself is a polynomial of order $2L-1$ or less.

The latter implies that $\sum_{\ell=1}^{L} w_\ell = \int_{-1}^{1} 1.\mathrm{d}x = 2.$

Since

$$\int_0^1 \phi(u)\mathrm{d}u = 0.5 \int_{-1}^{1} \phi((x+1)/2)\mathrm{d}x,$$

the LGQ approximation also applies to square integrable real functions on the unit interval, as

$$\int_0^1 \phi(u)\mathrm{d}u \approx 0.5 \sum_{\ell=1}^{L} w_\ell \phi((x_\ell+1)/2) = \sum_{\ell=1}^{L} b_\ell \phi(a_\ell),$$

say, for some fixed but not too small an L, where

$$a_\ell = (x_\ell+1)/2,\ b_\ell = w_\ell/2,\ \ell = 1, 2, ..., L.$$

Thus,

- $\int_0^1 \Psi_N(t|\boldsymbol{\delta})^2 \mathrm{d}t \approx \sum_{\ell=1}^{L} b_\ell \Psi_N(a_\ell|\boldsymbol{\delta})^2$,
- $\int_0^1 t\Psi_N(t|\boldsymbol{\delta})\mathrm{d}t \approx \sum_{\ell=1}^{L} b_\ell a_\ell \Psi_N(a_\ell|\boldsymbol{\delta})$,
- $\int_0^1 t^2\Psi_N(t|\boldsymbol{\delta})\mathrm{d}t \approx \sum_{\ell=1}^{L} b_\ell a_\ell^2 \Psi_N(a_\ell|\boldsymbol{\delta})$

As an example of how accurate the LGQ approximation can be, consider the Laplace transform of the χ_2^2 distribution, i.e., $\mathcal{L}(t|\chi_2^2) = 1/(1+2t)$.

It is not hard to verify that

$$\int_0^1 \ln\left(\mathcal{L}(t|\chi_2^2)\right) \mathrm{d}t = 1 - 1.5\ln(3) = -0.647918433002165.$$

Next, let us compute this integral via the LGQ approach, as $\sum_\ell^L b_\ell \ln(1/(1+2a_\ell))$, for $L = 5, 10, 20, 30, 40, 50$.
The results are presented in Table 4.

Table 4. Accuracy of the LGQ approximation

$$
\begin{aligned}
\int_0^1 \ln\left(\mathcal{L}(t|\chi_2^2)\right) \mathrm{d}t &= -0.647918433002165\\
\text{LGQ } (L = 05) &= -0.532160589540430\\
\text{LGQ } (L = 10) &= -0.647918433002425\\
\text{LGQ } (L = 20) &= -0.647918433002165\\
\text{LGQ } (L = 30) &= -0.647918433002164\\
\text{LGQ } (L = 40) &= -0.647918433002164\\
\text{LGQ } (L = 50) &= -0.647918433002165
\end{aligned}
$$

These results are quite amazing, except in the case $L = 5$, for which the percentage error is about -18%.

Our guess is that the favorable results for $L \geq 10$ are due to the monotonicity of the integrand $\ln(1/(1+2t))$.

Efficient computation of $f(w|\pi_n\delta)$ and its Laplace transform

Note that we can write $f(w|\pi_n\boldsymbol{\delta})$ as a quadratic form in the parameters, as,

$$f(w|\pi_n\boldsymbol{\delta}) = \frac{(\pi_+^n\boldsymbol{\delta})' B_n(w)(\pi_+^n\boldsymbol{\delta})}{(\pi_+^n\boldsymbol{\delta})'(\pi_+^n\boldsymbol{\delta})}$$

where

- $\pi_+^n\boldsymbol{\delta} = (1, \delta_1, \delta_1, ..., \delta_n)'$
- $B_n(w)$ is the $(n+1)\times(n+1)$ matrix-valued function with elements

$$\begin{aligned} b_{k,m}(w) &= \exp(-w)\varphi_k(w)\varphi_m(w) \\ &= \sum_{i=0}^{k}\sum_{j=0}^{m}\binom{k}{i}\binom{m}{j}\frac{(-1)^{i+j}}{i!j!}w^{i+j}\exp(-w) \end{aligned}$$

  for $k, m = 0, 1, ..., n.$

Now the Laplace transform of $f(w|\pi_n\boldsymbol{\delta})$ can be written as

$$\mathcal{L}(t|\pi_n\boldsymbol{\delta}) \stackrel{\text{def.}}{=} \int_0^\infty \exp(-t.w) f(w|\pi_n\boldsymbol{\delta})\mathrm{d}w = \frac{(\pi_+^n\boldsymbol{\delta})' A_n(t)(\pi_+^n\boldsymbol{\delta})}{(\pi_+^n\boldsymbol{\delta})'(\pi_+^n\boldsymbol{\delta})},$$

where

$$A_n(t) = \int_0^\infty \exp(-t.w) B_n(w)\mathrm{d}w$$

with elements

$$\begin{aligned}
a_{k,m}(t) &= \sum_{i=0}^{k}\sum_{j=0}^{m}\binom{k}{i}\binom{m}{j}\frac{(-1)^{i+j}}{i!j!}\int_0^\infty w^{i+j}\exp(-(1+t)w)\mathrm{d}w \\
&= \sum_{i=0}^{k}\sum_{j=0}^{m}\binom{k}{i}\binom{m}{j}\frac{(-1)^{i+j}}{i!j!}\left(\frac{1}{1+t}\right)^{i+j+1}\int_0^\infty w^{i+j}\exp(-w)\mathrm{d}w \\
&= \sum_{i=0}^{k}\sum_{j=0}^{m}\binom{k}{i}\binom{m}{j}\binom{i+j}{i}(-1)^{i+j}\left(\frac{1}{1+t}\right)^{i+j+1}
\end{aligned}$$

for $k, m = 0, 1, ..., n.$

The last equality follows from $\int_0^\infty w^n \exp(-w)\mathrm{d}w = n!$

Therefore, in computing the integrals

$$\int_0^1 \Psi_N(t|\pi_n\boldsymbol{\delta})^2 \mathrm{d}t \approx \sum_{\ell=1}^{L} b_\ell \left(\Upsilon_N(a_\ell) - \ln(\mathcal{L}(a_\ell|\pi_n\boldsymbol{\delta}))\right)^2$$

$$\int_0^1 t\Psi_N(t|\pi_n\boldsymbol{\delta}) \mathrm{d}t \approx \sum_{\ell=1}^{L} b_\ell a_\ell \left(\Upsilon_N(a_\ell) - \ln(\mathcal{L}(a_\ell|\pi_n\boldsymbol{\delta}))\right)$$

$$\int_0^1 t^2\Psi_N(t|\pi_n\boldsymbol{\delta}) \mathrm{d}t \approx \sum_{\ell=1}^{L} b_\ell a_\ell^2 \left(\Upsilon_N(a_\ell) - \ln(\mathcal{L}(a_\ell|\pi_n\boldsymbol{\delta}))\right)$$

for different $\boldsymbol{\delta}$'s, we need to compute the vector

$$(\Upsilon_N(a_1), ..., \Upsilon_N(a_L))'$$

only once, and to compute the Laplace transforms $\mathcal{L}(a_\ell|\pi_n\boldsymbol{\delta}))$ for different $\boldsymbol{\delta}$'s we need to compute the matrices $A_n(a_\ell)$ for $\ell = 1, 2, ..., L$ only once.

The shape of the concentrated objective function and its limit

Recall that the concentrated objective function takes the form

$$\underline{Q}_N(\boldsymbol{\delta}) = Q_N(\overline{\alpha}_N(\boldsymbol{\delta}), \overline{\sigma}^2_N(\boldsymbol{\delta}), \boldsymbol{\delta})$$

with pointwise limit

$$\underline{Q}(\boldsymbol{\delta}) \ = \ \int_0^1 \left(\Lambda(t|\boldsymbol{\delta}) + (\alpha_0 - \overline{\alpha}(\boldsymbol{\delta})) .t + \left(\sigma_0^2 - \overline{\sigma}^2(\boldsymbol{\delta})\right) t^2/2\right)^2 \mathrm{d}t$$

$$\underline{Q}(\boldsymbol{\delta}^0) \ = \ 0$$

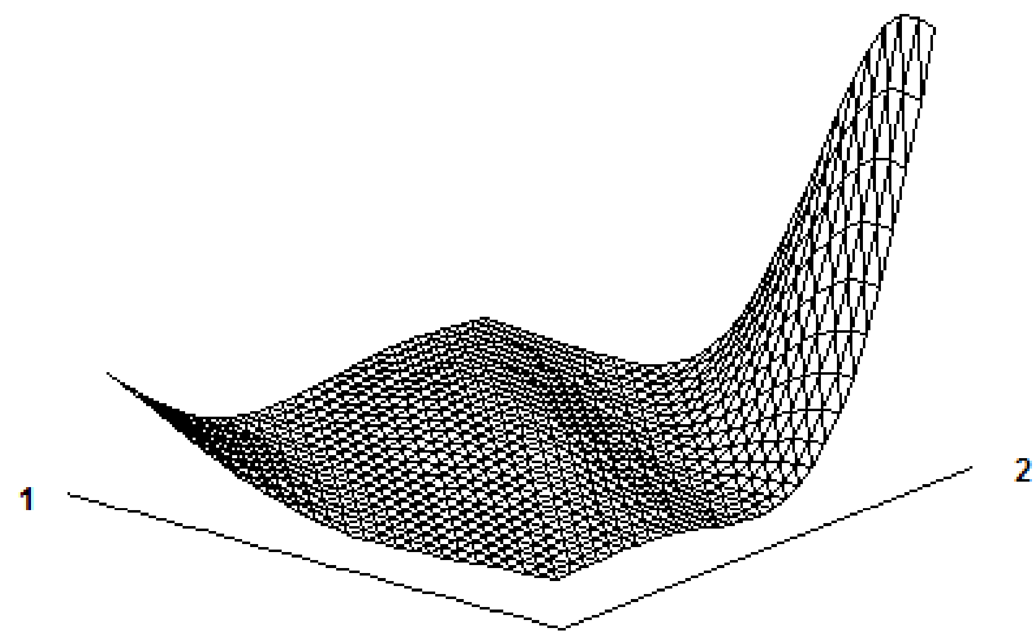
In order to check how $\underline{Q}_N(\boldsymbol{\delta})$ looks like in a neighborhood of $\boldsymbol{\delta}^0 = \{\delta_{0,m}\}_{m=1}^{\infty}$ for the test data, we have plotted $\underline{Q}_N(\pi_{20}\boldsymbol{\delta})$ for

$$(\delta_1, \delta_2) \ \in \ [\delta_{0,1} - 1, \delta_{0,1} + 1] \times [\delta_{0,2} - 1, \delta_{0,2} + 1]$$

$$\delta_m \ = \ \delta_{0,m} \text{ for } m = 3, 4, 5, ..., 20$$

using 29 grid points for δ_1 and δ_2.

Axis 1: delta(1) = .10396305 -+ 1
Axis 2: delta(2) =-.18991224 -+ 1



The case $(\delta_1, \delta_2) = (\delta_{0,1}, \delta_{0,2})$ corresponds roughly to a point in the middle of the picture, which however lies on top of a ridge.

The points (δ_1, δ_2) for which $\underline{Q}_N(\pi_{20}\boldsymbol{\delta})$, with $N = 500$, is minimal, compared with $(\delta_{0,1}, \delta_{0,2})$ and its value $\underline{Q}_N(\pi_{20}\boldsymbol{\delta}^0)$, are:

δ_1	-0.82460838	$\delta_{0,1}$	0.10396305
δ_2	-0.11848367	$\delta_{0,2}$	-0.18991224
$\underline{Q}_N(\pi_{20}\boldsymbol{\delta})$	0.00507953	$\underline{Q}_N(\pi_{20}\boldsymbol{\delta}^0)$	0.03024638

Thus, (δ_1, δ_2) lies in the flat valley on the right of the picture, close to axis 2.

Recall that $\underline{Q}_N(\boldsymbol{\delta}) \to \underline{Q}(\boldsymbol{\delta})$ a.s. as $N \to \infty$, with $\underline{Q}(\boldsymbol{\delta}^0) = 0$ and $\underline{Q}(\boldsymbol{\delta}) > 0$ whenever $||\boldsymbol{\delta} - \boldsymbol{\delta}^0|| > 0$.

The function $\underline{Q}(\pi_{20}\boldsymbol{\delta})$ for

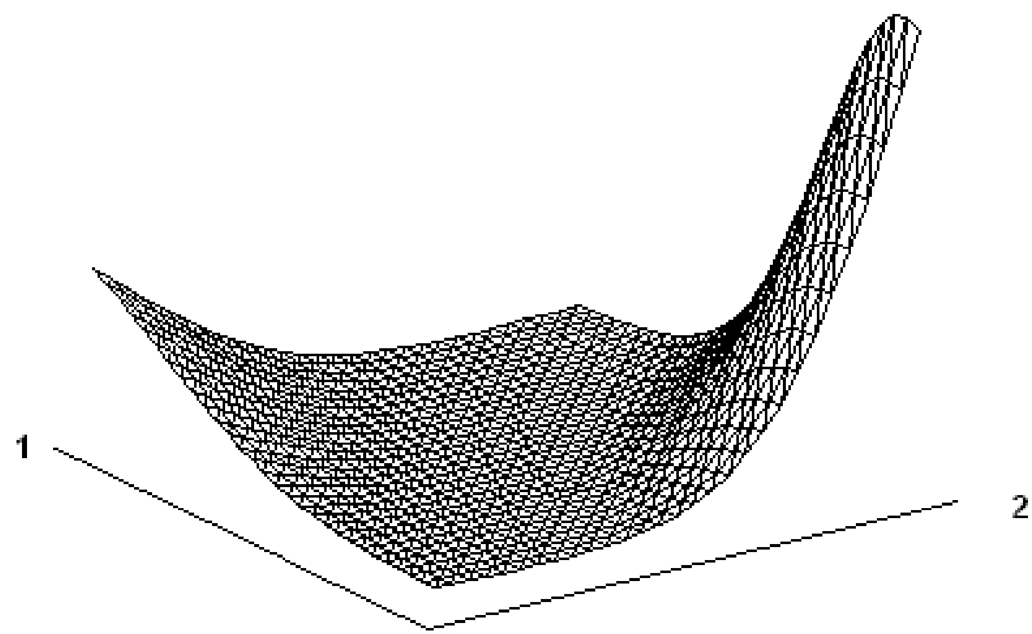
$$\begin{aligned} (\delta_1, \delta_2) &\in [\delta_{0,1} - 1, \delta_{0,1} + 1] \times [\delta_{0,2} - 1, \delta_{0,2} + 1] \\ \delta_m &= \delta_{0,m} \text{ for } m = 3, 4, 5, ..., 20 \end{aligned}$$

using 29 grid points for δ_1 and δ_2, as before, is displayed in the following plot.

As expected, the minimum of $\underline{Q}(\pi_{20}\boldsymbol{\delta})$ is attained in $(\delta_1, \delta_2) = (\delta_{0,1}, \delta_{0,2})$, with function value zero.

However, this minimum point is located in a rather flat valley.

Axis 1: delta(1) = .10396305 -+ 1
Axis 2: delta(2) =-.18991224 -+ 1



Recall that for the test data the fit of

$$\Upsilon_N(t) = \ln\left(\sum_{i=1}^{N} \exp\left(-t.(Z_{i,N} + ||X_i||)\right)\right) - \ln\left(\sum_{i=1}^{N} \exp(-t.||X_i||)\right)$$

to its a.s. limit $\Upsilon(t) = t + 0.25t^2/2 - 3.\ln\left(1 + t/(2\sqrt{3})\right)$ is far from perfect.

This may explain why $\underline{Q}_N(\boldsymbol{\delta})$ is not close to its limit $\underline{Q}(\boldsymbol{\delta})$.

Therefore, the conclusion is that for our SNP sieve estimation approach to work satisfactory we need a much larger sample than $N = 500$.

Moreover, the flatness of $\underline{Q}(\boldsymbol{\delta})$ near $\boldsymbol{\delta}^0$ is a numerical challence for computing its sieve estimator, even for large samples.

Thus, much more work has to be done before our approach is operational.

Thank you!