# Supplement to Designing to detect heteroscedasticity in a regression model

Lanteri, A.[1*], Leorato, S.[2], López-Fidalgo, J.[3], and Tommasi, C.[4]

[1,2,4]Department of Economics, Management and Quantitative Methods, University of Milan, Italy.
[3]Institute of Data Science and Artificial Intelligence, University of Navarre, Spain

[*]Corresponding author contacts: alessandro.lanteri@unimi.it
Via Conservatorio, 7. 20122 Milano, Italy.

## Proofs and technical results

**Theorem 1.**

*Proof.* The $\mathrm{D}_s$-optimum design for $\boldsymbol{\gamma}$ maximizes the following criterion (see for instance, Atkinson et al. (2007), Sect.10.3):

$$\Phi_{\mathrm{D}_s}(\xi;\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma}) = \frac{|\mathcal{I}(\xi;\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma})|}{|\mathcal{I}_{11}(\xi;\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma})|} = \frac{|\mathcal{I}(\xi;\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma})|}{\frac{1}{2\sigma^4}\,|\boldsymbol{M}(\xi;\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma})|}. \tag{1}$$

Since the Fisher information matrix can be partitioned as follows:

$$\mathcal{I}(\xi;\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma}) = \begin{bmatrix} \boldsymbol{M}(\xi;\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma}) & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{2\sigma^4}\boldsymbol{V}(\xi;\sigma^2,\boldsymbol{\gamma}) \end{bmatrix}, \tag{2}$$

where

$$\boldsymbol{V}(\xi;\sigma^2,\boldsymbol{\gamma}) = \sum_{i=1}^{k} \boldsymbol{f}(\boldsymbol{x}_i;\sigma^2,\boldsymbol{\gamma})\boldsymbol{f}(\boldsymbol{x}_i;\sigma^2,\boldsymbol{\gamma})^T\xi(\boldsymbol{x}_i), \quad \boldsymbol{f}(\boldsymbol{x};\sigma^2,\boldsymbol{\gamma}) = \begin{bmatrix} 1 \\ \sigma^2\,\nabla\log h(\boldsymbol{x};\boldsymbol{\gamma}) \end{bmatrix},$$

the $\mathrm{D}_s$-criterion (1) becomes,

$$\Phi_{\mathrm{D}_s}(\xi;\sigma^2,\boldsymbol{\gamma}) = \left(\frac{1}{2\sigma^4}\right)^s |\boldsymbol{V}(\xi;\sigma^2,\boldsymbol{\gamma})| = \left(\frac{1}{2}\right)^s |\widetilde{\boldsymbol{V}}(\xi;\boldsymbol{\gamma})| \tag{3}$$

where

$$\widetilde{\boldsymbol{V}}(\xi;\boldsymbol{\gamma}) = \sum_{i=1}^{k} \tilde{\boldsymbol{f}}(\boldsymbol{x}_i;\boldsymbol{\gamma})\tilde{\boldsymbol{f}}(\boldsymbol{x}_i;\boldsymbol{\gamma})^T\,\xi(\boldsymbol{x}_i), \quad \tilde{\boldsymbol{f}}(\boldsymbol{x};\boldsymbol{\gamma}) = \begin{bmatrix} 1 \\ \nabla \log h(\boldsymbol{x};\boldsymbol{\gamma}) \end{bmatrix},$$

which proves the theorem. □

**Corollary 1.**

*Proof.* The proof follows immediately from Theorem 1. Since $\gamma$ is scalar,

$$|\tilde{\boldsymbol{V}}| = \sum_{i=1}^{k} \left[\frac{\nabla h(x_i;\gamma)}{h(x_i;\gamma)}\right]^2 \xi(\boldsymbol{x}_i) - \left[\sum_{i=1}^{k} \frac{\nabla h(x_i;\gamma)}{h(x_i;\gamma)}\,\xi(\boldsymbol{x}_i)\right]^2 = \mathrm{Var}\left(\nabla \log h(x_i;\gamma)\right),$$

where $\mathrm{Var}(\cdot)$ denotes the variance with respect to $\xi$. Therefore, a $\mathrm{D}_1$-optimal design for $\gamma$ is equally weighted at $x_l \in \{x : \operatorname{argmin}_x \nabla \log h(x;\gamma)\}$ and $x_u \in \{x : \operatorname{argmax}_x \nabla \log h(x;\gamma)\}$. □

**Theorem 2.**

*Proof.* Let $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_0 + \boldsymbol{\lambda}/\sqrt{n}$ for a specific value of $\boldsymbol{\lambda}$. If $n$ is large enough to guarantee $\boldsymbol{\gamma}_0 + \boldsymbol{\lambda}/\sqrt{n} = \boldsymbol{\gamma}_1 \in \mathcal{B}(\boldsymbol{\gamma}_0, r)$, assumptions

$$\sup_{x\in\mathcal{X}} \|\nabla h(\boldsymbol{x},\boldsymbol{\gamma_0})\| \le M_1 < \infty;$$

$$\sup_{x\in\mathcal{X}} \sup_{\boldsymbol{\gamma}\in\mathcal{B}(\boldsymbol{\gamma}_0,r)} \|h''(\boldsymbol{x},\boldsymbol{\gamma})\|_2 \le M_2 < \infty,$$

allow us to compute a first order Taylor expansion of the function $h(\boldsymbol{x};\boldsymbol{\gamma}_1)$ at $\boldsymbol{\gamma}_0$, and thus,

$$I_{12}(\xi;\boldsymbol{\gamma}_1) = 1 + \log A_h - \log G_h, \tag{4}$$

becomes

$$\begin{aligned}
I_{12}(\xi;\boldsymbol{\gamma}_1) &= 1 + \log\left\{\sum_{i=1}^{k}\Big[h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0) + \nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_0)\right.\\
&+ \left.\frac{1}{2}(\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_0)^T h''(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})(\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_0)\Big]\xi(\boldsymbol{x}_i)\right\}\\
&- \sum_{i=1}^{k}\xi(\boldsymbol{x}_i)\log\Big[h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0) + \nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T(\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_0)\\
&+ \frac{1}{2}(\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_0)^T h''(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})(\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_0)\Big]\\
&= 1+\log\left\{1 + \sum_{i=1}^{k}\left[\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\frac{\boldsymbol{\lambda}}{\sqrt{n}} + \frac{1}{2\,n}\boldsymbol{\lambda}^T h''(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right]\xi(\boldsymbol{x}_i)\right\}\\
&- \sum_{i=1}^{k}\xi(\boldsymbol{x}_i)\log\left[1 + \nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\frac{\boldsymbol{\lambda}}{\sqrt{n}} + \frac{1}{2\,n}\boldsymbol{\lambda}^T h''(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right],
\end{aligned} \tag{5}$$

where $h^{''}(\boldsymbol{x}_i;\boldsymbol{\gamma})$ denotes the Hessian matrix of $h(\boldsymbol{x}_i;\boldsymbol{\gamma})$ and $\bar{\boldsymbol{\gamma}}$ is a value of $\boldsymbol{\gamma}$ such that $||\boldsymbol{\gamma}_0 - \bar{\gamma}|| \leq ||\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}_1||$. From the Taylor expansion of $\log(1+z)$ up to the second order, the previous expression becomes:

$$
\begin{aligned}
I_{12}(\xi;\boldsymbol{\gamma}_1) &= 1+\sum_{i=1}^{k}\left[\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\frac{\boldsymbol{\lambda}}{\sqrt{n}}+\frac{1}{2\,n}\boldsymbol{\lambda}^T h^{''}(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right]\xi(\boldsymbol{x}_i)\\
&- \frac{1}{2}\left\{\sum_{i=1}^{k}\left[\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\frac{\boldsymbol{\lambda}}{\sqrt{n}}+\frac{1}{2\,n}\boldsymbol{\lambda}^T h^{''}(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right]\xi(\boldsymbol{x}_i)\right\}^2+O\left(\frac{||\boldsymbol{\lambda}||^3}{n^{\frac{3}{2}}}\right)\\
&- \sum_{i=1}^{k}\xi(\boldsymbol{x}_i)\left\{\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\frac{\boldsymbol{\lambda}}{\sqrt{n}}+\frac{1}{2\,n}\boldsymbol{\lambda}^T h^{''}(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right.\\
&- \left.\frac{1}{2}\left[\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\frac{\boldsymbol{\lambda}}{\sqrt{n}}+\frac{1}{2\,n}\boldsymbol{\lambda}^T h^{''}(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right]^2+O\left(\frac{||\boldsymbol{\lambda}||^3}{n^{\frac{3}{2}}}\right)\right\}
\end{aligned}
$$

The first terms of the two developments simplify and thus:

$$
\begin{aligned}
I_{12}(\xi;\boldsymbol{\gamma}_1) &= 1-\frac{1}{2}\left[\sum_{i=1}^{k}\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\frac{\boldsymbol{\lambda}}{\sqrt{n}}\,\xi(\boldsymbol{x}_i)\right]^2\\
&+ \frac{1}{2}\sum_{i=1}^{k}\xi(\boldsymbol{x}_i)\left[\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\frac{\boldsymbol{\lambda}}{\sqrt{n}}\right]^2+O\left(\frac{||\boldsymbol{\lambda}||^3}{n^{\frac{3}{2}}}\right)\\
&= 1-\frac{1}{2\,n}\left[\boldsymbol{\lambda}^T\sum_{i=1}^{k}\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)\,\xi(\boldsymbol{x}_i)\right]\left[\sum_{i=1}^{k}\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\,\xi(\boldsymbol{x}_i)\boldsymbol{\lambda}\right]\\
&+ \frac{1}{2\,n}\sum_{i=1}^{k}\xi(\boldsymbol{x}_i)\left[\boldsymbol{\lambda}^T\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)\right]\left[\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\boldsymbol{\lambda}\right]+O\left(\frac{||\boldsymbol{\lambda}||^3}{n^{\frac{3}{2}}}\right)
\end{aligned}
$$

Rearranging the terms and taking into account that

$$
\begin{aligned}
\zeta(\xi;\boldsymbol{\lambda};\boldsymbol{\gamma}_0) &= \frac{1}{2}\boldsymbol{\lambda}^T\left(\sum_{i=1}^{k}\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\xi(\boldsymbol{x}_i)\right.\\
&- \left.\sum_{i=1}^{k}\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)\xi(\boldsymbol{x}_i)\,\sum_{i=1}^{k}\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T\xi(\boldsymbol{x}_i)\right)\boldsymbol{\lambda}.
\end{aligned}
$$

we obtain the thesis. □

Next Lemma is used for the proof of Theorem 3

**Lemma 1.** *Let $h_i$, $i=1,\ldots,n$ be $n$ positive quantities such that*

$$0<\underline{h}=h_1\leq h_2\leq\cdots\leq h_n=\overline{h}<\infty,$$

*and let* $A_n = n^{-1}\sum_i h_i$ *and* $G_n = \sqrt[n]{\prod_i h_i}$ *the arithmetic and geometric means of* $\{h_i\}_{i\le n}$*. We have the following bounds for* $D_n = \log A_n - \log G_n$*:*

$$\frac{2}{n}\left[\log\frac{\underline{h}+\overline{h}}{2} - \frac{1}{2}(\log\underline{h} + \log\overline{h})\right] \le D_n \le -\log H + H - 1, \tag{6}$$

*where*

$$H = \underline{h}\cdot\frac{\log\overline{h} - \log\underline{h}}{\overline{h}-\underline{h}}. \tag{7}$$

*The lower bound is achieved when* $h_1 = \underline{h}$, $h_n = \overline{h}$ *and* $h_i = (\overline{h}+\underline{h})/2$ *for* $i = 2,\dots,n-1$*. Further, the upper bound is reached for* $h_1 = \cdots = h_{r^*} = \underline{h}$ *and* $h_{r^*+1} = \cdots = h_n = \overline{h}$*, where* $r^* = \left\lfloor n\left(\frac{\overline{h}}{\overline{h}-\underline{h}} - \frac{1}{\log\overline{h}-\log\underline{h}}\right)\right\rfloor$ *and* $\lfloor\cdot\rfloor$ *denotes the integer part of a number.*

*Proof.* **a) Lower Bound** To find the lower bound in (6), we use a recursive argument similar to Tung (1975). Given the geometric and arithmetic means of $n-1$ terms, $A_{n-1}$ and $G_{n-1}$, we can define the difference $D_n$ as a function of an "additional" term $a$:

$$D_n(a) = \log\left[n^{-1}((n-1)A_{n-1} + a)\right] - n^{-1}\log\left(G_{n-1}^{n-1}\cdot a\right) \tag{8}$$

then we solve

$$D_n'(a) = \frac{1}{(n-1)A_{n-1}+a} - \frac{1}{n\,a} = 0 \tag{9}$$

that gives $a = A_{n-1}$. Since $D''(A_{n-1}) = \frac{n-1}{n\,A_{n-1}} > 0$ then $D_n \ge D_n(A_{n-1}) = \frac{n-1}{n}D_{n-1}$ is the minimum. By applying recursively, we get:

$$D_n \ge \frac{n-1}{n}D_{n-1} \ge \cdots \ge \frac{2}{n}\mathrm{D}_2 = \frac{2}{n}\left[\log\left(\frac{\underline{h}+\overline{h}}{2}\right) - \frac{1}{2}\log(\underline{h}\cdot\overline{h})\right].$$

It can be easily checked that $D_n = \frac{2}{n}\mathrm{D}_2$ if $h_1 = \underline{h}$, $h_n = \overline{h}$ and all other terms are equal to $A_2 = \frac{\underline{h}+\overline{h}}{2}$.

**b) Upper Bound** From left hand-side of (9), we have that $D_n'(a) \ge 0$ for all $a \ge A_{n-1}$ and thus, $D_n(a)$ reaches its maximum in one of the extremes $\underline{h}$ or $\overline{h}$. Let then $r$ be an integer such that

$$\underline{h} = h_1 = \ldots = h_r < h_{r+1} = \ldots = \overline{h}.$$

We can then define the function $D_n(r)$:

$$\begin{aligned} D_n(r) &= \log\frac{r\underline{h} + (n-r)\overline{h}}{n} - \frac{r}{n}\log\underline{h} - \frac{n-r}{n}\log\overline{h} \\ &= \log\frac{r + (n-r)b}{n} - \frac{n-r}{n}\log b \end{aligned}$$

where $b = \overline{h}/\underline{h}$.

By extending the function $D_n(r)$ to the whole interval $[1, n]$, we can solve:

$$D'_n(r) = \frac{1-b}{r+(n-r)b} + \frac{1}{n \log b} = 0,$$

that gives the maximum point

$$r^* = n\left(\frac{b}{b-1} - \frac{1}{\log b}\right),$$

with

$$D_n(r^*) = \frac{\log b}{b-1} - 1 - \log\frac{\log b}{b-1} = H - 1 - \log H,$$

where $H$ is defined in (7).

If $r^*$ is not integer, then the maximum of $D_n(r)$ over $[1, n] \cap \mathbb{N}$ must be attained at either $\lfloor r^* \rfloor$ or $\lfloor r^* \rfloor + 1$. By writing $\lfloor r^* \rfloor = n\left(\frac{b}{b-1} - \frac{1}{\log b}\right) - \mu$, where $\mu \in (0, 1)$ is the fractional part and by comparing $D_n(\lfloor r^* \rfloor)$ and $D_n(\lfloor r^* \rfloor + 1)$, we easily see that

$$D_n(\lfloor r^* \rfloor) - D_n(\lfloor r^* \rfloor + 1) = \frac{b-1}{n} - \log b \geq 0$$

which concludes the proof.

□

**Theorem 3.**

*Proof.* The maximization of (4) corresponds to finding the maximum possible value for the difference between the logarithms of the arithmetic and geometric means of $h(\boldsymbol{x}_i)$, $i = 1, \ldots, k$. The proof follows immediately from part b) of the proof of Lemma 1. □

**Theorem 4.**

*Proof.* From Theorem 3, by setting $b = \overline{h}/\underline{h}$, we easily obtain

$$\omega = \frac{b \log b - b + 1}{(b-1) \log b}.$$

Taking the limit for $\boldsymbol{\gamma}_1 \to \boldsymbol{\gamma}_0$, that is, for $b \to 1$, and by applying de l'Höpital rule twice,

$$\begin{aligned} \lim_{b \to 1} \frac{b \log b - b + 1}{(b-1) \log b} &= \lim_{b \to 1} \frac{\log b}{\log b + (b-1)/b} = \lim_{b \to 1} \frac{b \log b}{b \log b + b - 1} \\ &= \lim_{b \to 1} \frac{\log b + 1}{\log b + 2} = \frac{1}{2} \end{aligned}.$$

In order to complete the proof, it is enough to show that the function $\omega(\boldsymbol{\gamma}_1) - 1/2 > 0$, or equivalently, that the function

$$\frac{b}{b-1} - \frac{1}{\log(1+(b-1))} - 1/2 > 0$$

for all $b > 1$. By denoting $v = b - 1$, we can rewrite the above inequality as

$$\log(1+v) > \frac{2v}{v+2}.$$

Since we already proved that the function $g(v) = \log(1+v) - \frac{2v}{v+2} \to 0$ for $v \to 0$, it is enough to show that, for $v > 0$, its derivative is positive:

$$g'(v) = \frac{1}{v+1} - \frac{2v+4-2v}{(v+2)^2} = \frac{v^2}{(v+1)(v+2)^2} > 0.$$

□

**Theorem 5.**

*Proof.* Let $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_0 + \boldsymbol{\lambda}/\sqrt{n}$. For convenience, we denote the mean with respect to a design $\xi$ by $\mathbb{E}_\xi$ and let

$$Z_{in} = Z_n(\boldsymbol{x}_i, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0) := \nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} + \frac{1}{2n} \boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}. \quad (10)$$

With this notation, we can rewrite equation (5) as

$$I_{12}(\xi; \boldsymbol{\gamma}_1) = 1 + \log\left(1 + \mathbb{E}_\xi Z_n\right) - \mathbb{E}_\xi \log\left(1 + Z_n\right). \quad (11)$$

We clearly have that,

$$\sup_{\boldsymbol{x}\in\mathcal{X}} |Z_n(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0)| \leq \sup_{\boldsymbol{x}\in\mathcal{X}} \left|\nabla h(\boldsymbol{x}; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}}\right| + \sup_{\boldsymbol{x}\in\mathcal{X}} \left|\frac{1}{2n} \boldsymbol{\lambda}^T h''(\boldsymbol{x}; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right|$$

Now, from the Cauchy-Shwartz inequality and from the assumptions of the Theorem:

$$|\nabla h(\boldsymbol{x}; \boldsymbol{\gamma}_0)^T \boldsymbol{\lambda}| \leq \|\nabla h(\boldsymbol{x}; \boldsymbol{\gamma}_0)^T\| \cdot \|\boldsymbol{\lambda}\| \leq M_1 L \quad (12)$$

and, by the definition of the spectral norm:

$$\begin{aligned} |\boldsymbol{\lambda}^T h''(\boldsymbol{x}; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}| &\leq \|\boldsymbol{\lambda}\| \|h''(\boldsymbol{x}; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\| \leq \|\boldsymbol{\lambda}\|^2 \sup_{\lambda \neq 0} \frac{\|h''(\boldsymbol{x}; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\|}{\|\boldsymbol{\lambda}\|} \\ &= \|\boldsymbol{\lambda}\|^2 \|h''(\boldsymbol{x}; \bar{\boldsymbol{\gamma}})\|_2 \leq M_2 L^2. \quad (13) \end{aligned}$$

Thus, we get the upper bound:

$$\sup_{\boldsymbol{x}\in\mathcal{X}} |Z_n(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0)| \leq \frac{M_1 L}{\sqrt{n}} + \frac{M_2 L^2}{2n}$$

where the constants $L, M_1, M_2$ do not depend on $\xi$ nor on $n$.

It is therefore possible to find an integer $n^* := n^*(L, M_1, M_2)$ such that, for all $n \geq n^*$:

$$\max\left\{\frac{M_1 L}{\sqrt{n}}, \frac{M_2 L^2}{2n}\right\} < \frac{1}{4} \tag{14}$$

Thus, for $n \geq n^*$, we have that $\sup_{\boldsymbol{x}\in\mathcal{X}} |Z_n(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0)| < \frac{1}{2}$. As a consequence, $Z_n < \frac{1}{2}$ and, replacing each term of the expected value $\mathbb{E}_\xi|Z_n|$ with its maximum value,

$$\mathbb{E}_\xi Z_n \leq \mathbb{E}_\xi |Z_n| \leq \sup_{\boldsymbol{x}\in\mathcal{X}} |Z_n(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0)| < \frac{1}{2}.$$

This enables us to expand in infinite series both the terms $\log(1+\mathbb{E}_\xi Z_n)$ and $\log(1+Z_n)$ in (11), obtaining:

$$\begin{aligned}
I_{12}(\xi; \boldsymbol{\gamma}_1) \;=\; & 1 + \left[\sum_{i=1}^{k}\left(\nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} + \frac{1}{2n}\boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right)\xi(\boldsymbol{x}_i)\right] \\
& -\frac{1}{2}\left[\sum_{i=1}^{k}\left(\nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} + \frac{1}{2n}\boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right)\xi(\boldsymbol{x}_i)\right]^2 \\
& -\sum_{i=1}^{k}\left[\left(\nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} + \frac{1}{2n}\boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right)\right]\xi(\boldsymbol{x}_i) \\
& +\frac{1}{2}\sum_{i=1}^{k}\left[\left(\nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} + \frac{1}{2n}\boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right)\right]^2\xi(\boldsymbol{x}_i) \\
& +\sum_{m=3}^{\infty}\frac{(-1)^{m+1}}{m}[(\mathbb{E}_\xi Z_n)^m - \mathbb{E}_\xi Z_n^m]
\end{aligned}$$

The first terms of the two expansions simplify, and after some calculations, we obtain, as in Theorem 2:

$$I_{12}(\xi; \boldsymbol{\gamma}_1) = 1 + \frac{\zeta(\xi, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0)}{n} + O(n^{-3/2}),$$

where the $O(n^{-3/2})$ term is now computed exactly:

$$\begin{aligned}
O(n^{-3/2}) \;=\; & \frac{1}{2}\left[\sum_i \frac{1}{4n^2}\left(\boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right)^2 \xi(\boldsymbol{x}_i) - \frac{1}{4n^2}\left(\sum_i \boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\xi(\boldsymbol{x}_i)\right)^2\right] \\
& -\frac{1}{2n}\left[\left(\sum_i \nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}}\xi(\boldsymbol{x}_i)\right)\left(\sum_i \boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\xi(\boldsymbol{x}_i)\right) + \right. \\
& \left. -\left(\sum_i \nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}}\boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\xi(\boldsymbol{x}_i)\right)\right] \\
& +\sum_{m=3}^{\infty}\frac{(-1)^{m+1}}{m}\left[(\mathbb{E}_\xi Z_n)^m - \mathbb{E}_\xi Z_n^m\right]
\end{aligned}$$

Let $\text{Var}_\xi f(\boldsymbol{X})$ and $\text{Cov}_\xi(f(\boldsymbol{X}), g(\boldsymbol{X}))$ denote variances and covariances with respect to the design $\xi$, respectively. We can then write:

$$\begin{aligned} n[I_{12}(\xi;\boldsymbol{\gamma}_1) - 1] - \zeta(\xi, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0) = \\ = \frac{1}{8n} \text{Var}_\xi \left(\boldsymbol{\lambda}^T h^{''}(\boldsymbol{X};\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right) + \frac{1}{2\sqrt{n}} \text{Cov}_\xi \left(\boldsymbol{\lambda}^T h^{''}(\boldsymbol{X};\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}, \nabla h(\boldsymbol{X};\boldsymbol{\gamma}_0)^T \boldsymbol{\lambda}\right) \\ +n \sum_{m=3}^{\infty} \frac{(-1)^{m+1}}{m} \left[(\mathbb{E}_\xi Z_n)^m - \mathbb{E}_\xi Z_n^m\right] \end{aligned}$$

Thus, using the Cauchy-Shwartz bound $\text{Cov}_\xi(f,g) \leq \sqrt{\text{Var}_\xi(f)\,\text{Var}_\xi(g)}$,

$$\begin{aligned} |n[I_{12}(\xi;\boldsymbol{\gamma}_1) - 1] - \zeta(\xi, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0)| \leq \frac{1}{8n} \text{Var}_\xi \left(\boldsymbol{\lambda}^T h^{''}(\boldsymbol{X};\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right) \\ + \frac{1}{2\sqrt{n}} \sqrt{\text{Var}_\xi \left(\boldsymbol{\lambda}^T h^{''}(\boldsymbol{X};\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right) \text{Var}_\xi \left(\nabla h(\boldsymbol{X};\boldsymbol{\gamma}_0)^T \boldsymbol{\lambda}\right)} \\ + \sum_{m=3}^{\infty} \frac{n}{m} \left|(\mathbb{E}_\xi Z_n)^m - \mathbb{E}_\xi Z_n^m\right|. \end{aligned} \tag{15}$$

We need to prove that the right-hand side of (15) is bounded from above by a quantity that is inversely proportional to $n$ and independent of $\xi$.

Let us start with

$$\sum_{m=3}^{\infty} \frac{n}{m} \left|(\mathbb{E}_\xi Z_n)^m - \mathbb{E}_\xi Z_n^m\right| \leq \sum_{m=3}^{\infty} \frac{n}{m} \left[(\mathbb{E}_\xi |Z_n|)^m + \mathbb{E}_\xi |Z_n|^m\right] \leq \sum_{m=3}^{\infty} \frac{2n}{m} \mathbb{E}_\xi |Z_n|^m, \tag{16}$$

and taking into account the expression of $Z_n$ given in (10), we have that

$$\begin{aligned} \sum_{m=3}^{\infty} \frac{2n}{m} \mathbb{E}_\xi |Z_n|^m &\leq \sum_{m=3}^{\infty} \frac{2n}{m} \mathbb{E}_\xi \left(\left|\nabla h(\boldsymbol{X};\boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}}\right| + \left|\frac{1}{2n} \boldsymbol{\lambda}^T h^{''}(\boldsymbol{X};\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right|\right)^m \\ &\leq \sum_{m=3}^{\infty} \frac{2n}{m} \left[2^m\, \mathbb{E}_\xi \left|\nabla h(\boldsymbol{x}_i;\boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}}\right|^m + 2^m\, \mathbb{E}_\xi \left|\frac{1}{2n} \boldsymbol{\lambda}^T h^{''}(\boldsymbol{x}_i;\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}\right|^m\right], \end{aligned} \tag{17}$$

where the last inequality follows from the fact that, given $a, b \in \mathbb{R}$ and $m \in \mathbb{N}$,

$$(|a| + |b|)^m = \sum_{j=0}^{m} \binom{m}{j} |a|^j |b|^{m-j} \leq 2^m \max\{|a|^m, |b|^m\} \leq 2^m (|a|^m + |b|^m),$$

with $a = \nabla h(\boldsymbol{X};\boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}}$ and $b = \frac{1}{2n} \boldsymbol{\lambda}^T h^{''}(\boldsymbol{X};\bar{\boldsymbol{\gamma}})\boldsymbol{\lambda}$.

In addition, in the right-hand side of (17) $2/m < 1$ as in the summation $m > 2$, and thus:

$$\begin{aligned}
\sum_{m=3}^{\infty} \frac{2n}{m} &\left[ 2^m \, \mathbb{E}_\xi \left| \nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} \right|^m + 2^m \, \mathbb{E}_\xi \left| \frac{1}{2n} \boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}}) \boldsymbol{\lambda} \right|^m \right] \\
&\leq \sum_{m=3}^{\infty} n \left[ \mathbb{E}_\xi \left| 2\nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} \right|^m + \mathbb{E}_\xi \left| \frac{1}{n} \boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}}) \boldsymbol{\lambda} \right|^m \right] \\
&\leq \sum_{m=3}^{\infty} n \left[ \left( \frac{2M_1 L}{\sqrt{n}} \right)^m + \left( \frac{M_2 L^2}{n} \right)^m \right], \qquad (18)
\end{aligned}$$

where the last inequality follows from the assumptions of the Theorem:

$$\begin{aligned}
\mathbb{E}_\xi \left| 2\nabla h(\boldsymbol{x}_i; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} \right|^m &\leq \sup_{\boldsymbol{x} \in \mathcal{X}} \left| 2\nabla h(\boldsymbol{x}; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} \right|^m \\
&= \left[ \sup_{\boldsymbol{x} \in \mathcal{X}} \left| 2\nabla h(\boldsymbol{x}; \boldsymbol{\gamma}_0)^T \frac{\boldsymbol{\lambda}}{\sqrt{n}} \right| \right]^m \leq \left( \frac{2M_1 L}{\sqrt{n}} \right)^m
\end{aligned}$$

and

$$\mathbb{E}_\xi \left| \frac{1}{n} \boldsymbol{\lambda}^T h''(\boldsymbol{x}_i; \bar{\boldsymbol{\gamma}}) \boldsymbol{\lambda} \right|^m \leq \sup_{\boldsymbol{x} \in \mathcal{X}} \left| \frac{1}{n} \boldsymbol{\lambda}^T h''(\boldsymbol{x}; \bar{\boldsymbol{\gamma}}) \boldsymbol{\lambda} \right|^m = \left( \sup_{\boldsymbol{x} \in \mathcal{X}} \left| \frac{1}{n} \boldsymbol{\lambda}^T h''(\boldsymbol{x}; \bar{\boldsymbol{\gamma}}) \boldsymbol{\lambda} \right| \right)^m \leq \left( \frac{M_2 L^2}{n} \right)^m.$$

In addition, if $n > n^*$ then (14) is fulfilled and thus $2M_1 L/\sqrt{n} < 1$ and $M_2 L^2/n < 1$. As a consequence, the two power series in (18) converge, and from (16)-(18) we obtain:

$$\begin{aligned}
\sum_{m=3}^{\infty} \frac{n}{m} \left| (\mathbb{E}_\xi Z)^m - \mathbb{E}_\xi Z^m \right| &\leq \sum_{m=3}^{\infty} n \left[ \left( \frac{2M_1 L}{\sqrt{n}} \right)^m + \left( \frac{M_2 L^2}{n} \right)^m \right] \\
&= n \left[ \left( 1 - \frac{2M_1 L}{\sqrt{n}} \right)^{-1} - 1 - \frac{2M_1 L}{\sqrt{n}} - \left( \frac{2M_1 L}{\sqrt{n}} \right)^2 + \right. \\
&\quad \left. + \left( 1 - \frac{M_2 L^2}{n} \right)^{-1} - 1 - \frac{M_2 L^2}{n} - \left( \frac{M_2 L^2}{n} \right)^2 \right] \qquad (19)
\end{aligned}$$

In particular, from (14) we have that $1 - 2M_1 L/\sqrt{n} > 1/2$, and thus:

$$\left( 1 - \frac{2M_1 L}{\sqrt{n}} \right)^{-1} - 1 - \frac{2M_1 L}{\sqrt{n}} - \left( \frac{2M_1 L}{\sqrt{n}} \right)^2 = \frac{(2M_1 L/\sqrt{n})^3}{1 - 2M_1 L/\sqrt{n}} \leq \frac{16 M_1^3 L^3}{n^{3/2}}.$$

Similarly, the upper bound in (14) guarantees that:

$$\left( 1 - \frac{M_2 L^2}{n} \right)^{-1} - 1 - \frac{M_2 L^2}{n} - \left( \frac{M_2 L^2}{n} \right)^2 \leq \frac{(M_2 L^2/n)^3}{1 - M_2 L^2/n} \leq \frac{2 M_2^3 L^6}{n^3}$$

Using these two upper-bounds in (19), we have:

$$\sum_{m=3}^{\infty} \frac{n}{m} \left| (\mathbb{E}_\xi Z)^m - \mathbb{E}_\xi Z^m \right| \leq \frac{16 M_1^3 L^3}{n^{1/2}} + \frac{2 M_2^3 L^6}{n^2}. \qquad (20)$$

It then remains to consider the other terms in the right-hand side of (15). Using the trivial upper bound $\mathrm{Var}(X) \leq \mathbb{E}X^2$ and arguments similar to those used above,

$$\mathrm{Var}_\xi \left( \boldsymbol{\lambda}^T h^{''}(\boldsymbol{X}; \bar{\boldsymbol{\gamma}}) \boldsymbol{\lambda} \right) \leq \mathbb{E}_\xi \left| \boldsymbol{\lambda}^T h^{''}(\boldsymbol{X}; \bar{\boldsymbol{\gamma}}) \boldsymbol{\lambda} \right|^2 \leq |\boldsymbol{\lambda}^T \boldsymbol{\lambda}|^2 \mathbb{E}_\xi \| h^{''}(\boldsymbol{X}; \bar{\boldsymbol{\gamma}}) \|_2^2 \leq L^4 M_2^2$$

and

$$\mathrm{Var}_\xi \left( \nabla h(\boldsymbol{X}; \boldsymbol{\gamma}_0)^T \boldsymbol{\lambda} \right) \leq \mathbb{E}_\xi \left( \nabla h(\boldsymbol{X}; \boldsymbol{\gamma}_0)^T \boldsymbol{\lambda} \right)^2 \leq \mathbb{E}_\xi \| \nabla h(\boldsymbol{X}; \boldsymbol{\gamma}_0) \|^2 \| \boldsymbol{\lambda} \|^2 \leq L^2 M_1^2.$$

Thus, putting everything together, we get:

$$|n(I_{12}(\xi; \boldsymbol{\gamma}_1) - 1) - \zeta(\xi, \boldsymbol{\lambda}, \boldsymbol{\gamma}_0)| \leq \frac{16 L^3 M_1^3}{\sqrt{n}} + \frac{2 L^6 M_2^3}{n^2} + \frac{L^4 M_2^2}{8n} + \frac{L^3 M_1 M_2}{2\sqrt{n}}$$

where the constants $K, M_1$ and $M_2$ are independent of $\xi$.

□

# References

A. C. Atkinson, A. N. Donev, and R. D. Tobias. *Optimum experimental designs, with SAS*. Oxford University Press, 2007. ISBN 978–0–19–929659–0.

S. H. Tung. On lower and upper bounds of the difference between the arithmetic and the geometric mean. *Mathematics of Computation*, 29(131):834–836, 1975.