# Prediction of Ordered Random Effects in a Simple Small Area Model

Yaakov Malinovsky* and Yosef Rinott*,**

*The Hebrew University of Jerusalem and **LUISS, Rome

## Supplementary Material

In this Supplement we provide some of the simulations and technical proofs. Equations in this Supplement are indicated by $S$, e.g., ($S3.1$), and similarly, lemmas that appear only in the Supplement are numbered with $S$, e.g., Lemma $S3.1$. Equations, lemmas, and Theorems without $S$, refer to the article itself. Most of the notation is defined in the article, and this Supplement cannot be read independently.

## S1. Simulations for Conjecture 1

**Conjecture 1.** *The optimal $\gamma$ in the sense of Theorem 4, $\gamma^o$, satisfies*

$$\lim_{m\to\infty} \gamma^o = \sqrt{\gamma^*}.$$

We justify Conjecture 1 by simulations. First we consider the case that both the area random effect $u_i$ and the sampling error $e_i$ have a normal distribution, and take $m = 5, 10, 20, 100$, and then repeat the simulation with $e_i$ having a translated exponential distribution. The red lines in Figure 1S are the lower and upper bounds of (3.6) to the optimal $\gamma$ of Theorem 4 and the blue line is the optimal $\gamma$, both as functions of $\gamma^*$. The simulations were done as follows: we set $\sigma_u^2 = 1$. Different values of $\sigma_e^2$ define the different values of $\gamma^*$. Setting without loss of generality $\mu = 0$, we generated $y_i = 0 + u_i + e_i$, $i = 1, \ldots, m$. For each value of $\gamma^*$ we ran 1,000 simulations. By suitably averaging over these simulations, we then approximated $E\{L(\widehat{\boldsymbol{\theta}}_{()}^{[2]}(\gamma), \boldsymbol{\theta}_{()})\}$ for each $\gamma \in [0, 1]$ using an exhaustive search with step-size of 0.001 and found $\gamma^o$, the value of $\gamma$ that minimizes $E\{L(\widehat{\boldsymbol{\theta}}_{()}^{[2]}(\gamma), \boldsymbol{\theta}_{()})\}$.

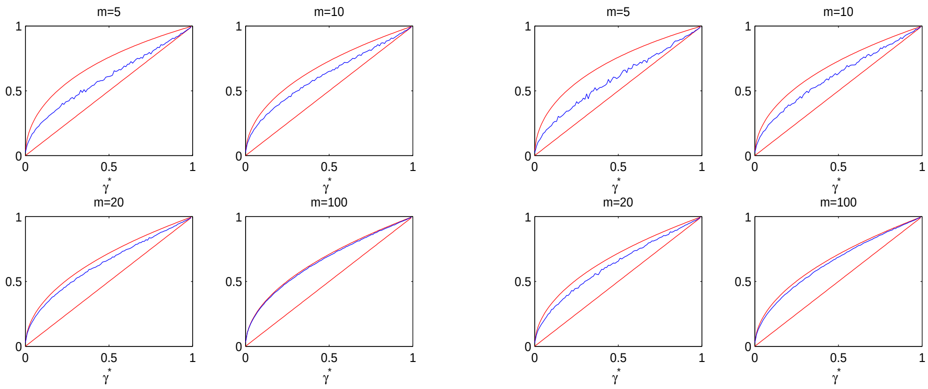


**Figure 1S:** $\gamma^o$ (the optimal $\gamma$) as a function of $\gamma^*$ (blue line) and the range of optimal $\gamma$ from Theorem 4 as a function of $\gamma^*$ (red lines) when:

1. Both the area random effect $u_i$ and the sampling error $e_i$ are normal (left four graphs).
2. The area random effects $u_i$ are normal, but the sampling errors $e_i$ are from a location exponential distribution (an exponential distribution translated by a constant) (right four graphs).

## S2. Simulations for Conjecture 2, and comparison of predictors

### S2.1. Known variances

For normal F and G, Conjecture 2 says that the predictor $\widehat{\theta}^{[3]}_{(i)}$ is better than $\widehat{\theta}^{[2]}_{(i)}(\gamma)$ for all values of $\gamma$ (including the optimal) in the sense that $E\{L(\widehat{\boldsymbol{\theta}}^{[3]}_{()}, \boldsymbol{\theta}_{()})\} \leq E\{L(\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma), \boldsymbol{\theta}_{()})\}$. Recall that $E\{L(\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma^o), \boldsymbol{\theta}_{()})\} \leq E\{L(\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma), \boldsymbol{\theta}_{()})\}$ for all $\gamma$. The simulations below support Conjecture 2. Figure 2S shows a sample of simulation results for $m = 30$ and 100. We compare the expected loss in predicting $\boldsymbol{\theta}_{()}$ by $\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma^o)$ to that of $\widehat{\boldsymbol{\theta}}^{[3]}_{()}$. While doing these simulations, we also compared the expected loss in predicting $\theta_{(m)}$ by $\widehat{\theta}^{[2]}_{(m)}(\gamma^o)$ to that of $\widehat{\theta}^{[3]}_{(m)}$.

The simulations show that the expected losses of the predictors $\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma^o)$ and $\widehat{\boldsymbol{\theta}}^{[3]}_{()}$ are rather close, while the predictor $\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma^*)$ is far worse. This suggests that the linear predictor $\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma^o)$ can be used without much loss. It is important to note that given $\gamma^o$, this estimator is easy to calculate. For large $m$ one may take $\gamma^o = \sqrt{\gamma^*}$, whereas for small $m$, the approximation of Section 3.2 can be used.

The simulation was done as follows: we set $\sigma_u^2 = 1$. Different values of $\sigma_e^2$ define the different values of $\gamma^*$. Setting $\mu = 0$, we generated $y_i = 0 + u_i + e_i$, $i = 1, .., m$. For each value of $\gamma^*$ we ran 1,000 simulations and approximated $E\{L(\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma), \boldsymbol{\theta}_{()})\}$ for each $\gamma$ in the range (3.6). Using an exhaustive search with step-size of 0.001 we found $\gamma^o$, the minimizer of $E\{L(\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma), \boldsymbol{\theta}_{()})\}$. We approximated $\widehat{\theta}^{[3]}_{(i)}$ in the following way: when both F and G are normal, $\theta_i | y_i \sim N\left(\gamma^* y_i + (1-\gamma^*)\mu, \gamma^* \sigma_e^2\right)$. Hence, for each $y_i$, $i = 1, ..., m$, we generated 1,000 random variables from $N\left(\gamma^* y_i + (1-\gamma^*)\overline{y}, \gamma^* \sigma_e^2\right)$, sorted them, and approximated $\widehat{\theta}^{[3]}_{(i)}$. We approximated $E\{L(\widehat{\boldsymbol{\theta}}^{[3]}_{()}, \boldsymbol{\theta}_{()})\}$ in the same way as we approximated $E\{L(\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma^*), \boldsymbol{\theta}_{()})\}$.

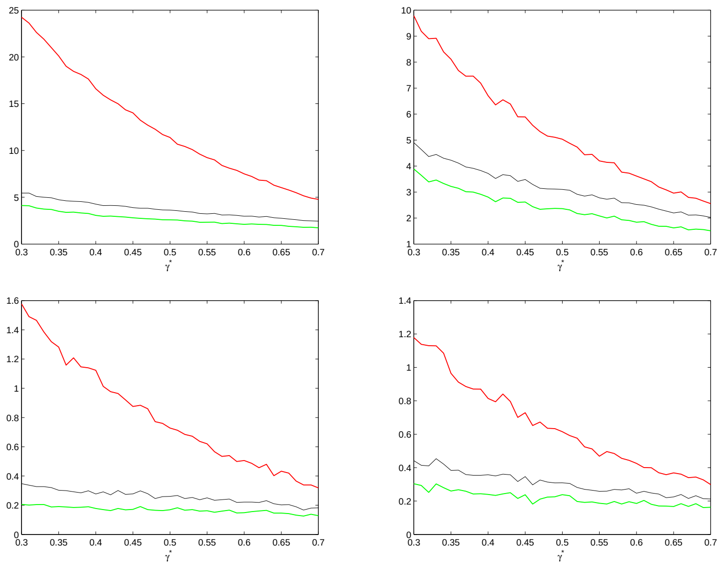

**Figure 2S:**

- Comparison of $E\{L(\widehat{\boldsymbol{\theta}}_{()}, \boldsymbol{\theta}_{()})\}$ as a function of $\gamma^*$, for the predictors $\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma^*)$, $\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma^o)$, $\widehat{\boldsymbol{\theta}}^{[3]}_{()}$ (red, black, green lines), where F and G are normal and $m = 100$ (upper left ), $m = 30$ (upper right)

- Comparison of the MSE of $\widehat{\theta}^{[2]}_{(m)}(\gamma^*)$, $\widehat{\theta}^{[2]}_{(m)}(\gamma^o)$, $\widehat{\theta}^{[3]}_{(m)}$ (red, black, green lines) for predicting $\theta_{(m)}$, as a function of $\gamma^*$, where F and G are normal and $m = 100$ (bottom left), $m = 30$ (bottom right)

### S2.2. Unknown variances

Figure 3$S$ compares the risks when only $\sigma_u^2$ is unknown and its estimator (4.1) is plugged-in. Otherwise, the simulations are similar to those of the previous section. The case that both variances, $\sigma_u^2$ and $\sigma_e^2$ are unknown is considered in Section 4 in the article.

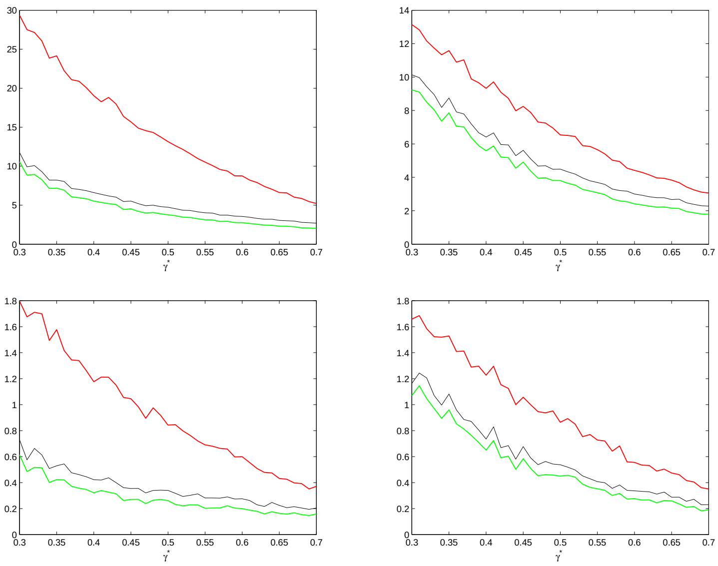

**Figure 3S:**

- Comparison of $E\{L(\widehat{\boldsymbol{\theta}}_{(\,)}, \boldsymbol{\theta}_{(\,)})\}$ as a function of $\gamma^*$, for the predictors $\widehat{\boldsymbol{\theta}}^{[2]}_{(\,)}(\gamma^*)$, $\widehat{\boldsymbol{\theta}}^{[2]}_{(\,)}(\sqrt{\gamma^*})$, $\widehat{\boldsymbol{\theta}}^{[3]}_{(\,)}$ (red, black, green lines), where F and G are normal and $m = 100$ (upper left ), $m = 30$ (upper right)
- Comparison of the MSE of $\widehat{\theta}^{[2]}_{(m)}(\gamma^*)$, $\widehat{\theta}^{[2]}_{(m)}(\sqrt{\gamma^*})$, $\widehat{\theta}^{[3]}_{(m)}$ (red, black, green lines) for predicting $\theta_{(m)}$, as a function of $\gamma^*$, where F and G are normal and $m = 100$ (bottom left), $m = 30$ (bottom right)

## S3. Proof of Theorem 5

For the proof of Theorem 5 we need some further lemmas. In the sequel, $\mathbb{I}$ denotes an indicator function, and $\varphi$ and $\Phi$ denote the standard normal density and cdf.

**Lemma S3.1.** *Set* $\psi(a) := \int_0^\infty t^2\Phi\left(at\right)\varphi(t)dt$, $\varrho_1(a) = \frac{1}{4} + \left(\frac{1}{4\pi} + \frac{1}{8}\right)\mathbb{I}\left(a \geq 1\right)$, *and* $\varrho_2(a) = \frac{1}{4}\mathbb{I}\left(a = 0\right) + \left(\frac{3}{8} + \frac{a}{4\pi}\right)\mathbb{I}\left(0 < a < \frac{\pi}{2}\right) + \frac{1}{2}\mathbb{I}\left(a \geq \frac{\pi}{2}\right)$. *Then*

$$\varrho_1(a) \leq \psi(a) \leq \varrho_2(a) \textit{ for all } a \geq 0, \textit{with equalities for } a = 0, a = 1.$$

**Proof of Lemma S3.1.** Note that $\psi(a) = \int_0^\infty t^2\Phi\left(at\right)\varphi(t)dt$ is increasing in $a$, and thus for $0 \leq a < \infty$, we have $1/4 = \psi(0) \leq \psi(a) \leq \psi(\infty) = 1/2$. A simple calculation shows that $\psi(1) = \frac{1}{4} + \left(\frac{1}{4\pi} + \frac{1}{8}\right)$, and the lower bound follows.

The upper bound follows readily once we show that for $a > 0$, $\psi(a) \leq \left(\frac{3}{8} + \frac{a}{4\pi}\right)$. We use the latter inequality only for $0 < a \leq \frac{\pi}{2}$ since for $a \geq \pi/2, 1/2$ is a better upper bound. (In fact $1/2$ is a good bound since for $a > 1$, that $\psi(a) > \psi(1) = \frac{1}{4} + \left(\frac{1}{4\pi} + \frac{1}{8}\right) \approx 0.4546$.)

To show $\psi(a) \leq \left(\frac{3}{8} + \frac{a}{4\pi}\right)$ for $a > 0$ we compute Taylor's expansion around $a = 1$,

$$\Phi(at) = \Phi(t) + t\varphi(t)(a-1) - \frac{a^*t^3}{2}\varphi(a^*t)(a-1)^2,$$

with $a^*$ between 1 and $a$. It follows that

$$\Phi(at) \leq \Phi(t) + t\varphi(t)(a-1), \quad for \quad t \geq 0 \quad and \quad a \geq 0.$$

Therefore,

$$\psi(a) = \int_0^\infty t^2\Phi\left(at\right)\varphi(t)dt \leq \int_0^\infty t^2\Phi\left(t\right)\varphi(t)dt + (a-1)\int_0^\infty t^3\varphi^2(t)dt$$
$$= \left(\frac{1}{4\pi} + \frac{3}{8}\right) + \frac{a-1}{4\pi} = \left(\frac{3}{8} + \frac{a}{4\pi}\right) \quad for \quad all \quad a \geq 0. \qquad \square$$

**Lemma S3.2.** *Let $Z \sim N(0,1)$. Then $2\varrho_1(a) - \frac{1}{2} \le E\left(|Z|Z\Phi\left(aZ\right)\right) \le 2\varrho_2(a) - \frac{1}{2}$. Equalities hold when $a = 0$ or $a = 1$.*

**Proof of Lemma S3.2.**

$$E\left(|Z|\Phi\left(aZ\right)Z\right) = \int_{-\infty}^{\infty} |t|t\Phi\left(at\right)\varphi(t)dt = \int_{0}^{\infty} t^2\Phi\left(at\right)\varphi(t)dt - \int_{-\infty}^{0} t^2\Phi\left(at\right)\varphi(t)dt$$
$$= 2\int_{0}^{\infty} t^2\Phi\left(at\right)\varphi(t)dt - \frac{1}{2} = 2\psi(a) - \frac{1}{2}. \qquad (S3.1)$$

The result now follows from Lemma S3.1. □

**Lemma S3.3.** *For Model (1.1) with $F$ and $G$ normal, $m = 2$, and $\mu = 0$,*

$$E\left(\theta_{(2)}y_{(2)}\right) \le 2\sigma_u^2\varrho_2(a) + \frac{\sigma_e^2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}$$

*and*

$$E\left(\theta_{(2)}y_{(2)}\right) \ge 2\sigma_u^2\varrho_1(a) + \frac{\sigma_e^2}{\pi}\sqrt{\gamma^*(1-\gamma^*)},$$

*where $a = \sqrt{\frac{\gamma^*}{1-\gamma^*}}$.*

**Proof of Lemma S3.3.** Kella (1986) (see also David and Nagaraja (2003)) shows that

$$E_\mu\left(\theta_{(i)}|\mathbf{y}\right) = \Phi\left(\triangle\right)\mu_1 + \Phi\left(-\triangle\right)\mu_2 + (-1)^i\sigma\sqrt{2}\varphi\left(\triangle\right), \qquad (S3.2)$$

where $\triangle = \gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}$, $\sigma^2 = \gamma^*\sigma_e^2$, $\mu_i = \gamma^* y_i$, $i = 1,2$. Therefore,

$$E(\theta_{(2)}y_{(2)}) = E(y_{(2)}E(\theta_{(2)}|\mathbf{y})) = E\left(y_{(2)}\left(\Phi\left(\triangle\right)\mu_1 + \Phi\left(-\triangle\right)\mu_2 + \sigma\sqrt{2}\varphi\left(\triangle\right)\right)\right)$$
$$= \gamma^* E\left(y_{(2)}\Phi\left(\gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}\right)(y_1-y_2)\right) + \gamma^* E\left(y_{(2)}y_2\right) + \sigma\sqrt{2}E\left(y_{(2)}\varphi\left(\gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}\right)\right). \qquad (S3.3)$$

We now calculate the latter three terms. For the first we use the relation $y_{(2)} = \frac{y_1+y_2}{2} + \frac{|y_1-y_2|}{2}$. We have

$$E\left(y_{(2)}\Phi\left(\gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}\right)(y_1-y_2)\right) = E\left(\frac{y_1+y_2}{2}\Phi\left(\gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}\right)(y_1-y_2)\right)$$
$$+ E\left(\frac{|y_1-y_2|}{2}\Phi\left(\gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}\right)(y_1-y_2)\right) = E\frac{y_1+y_2}{2}E\left(\Phi\left(\gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}\right)(y_1-y_2)\right)$$
$$+ E\left(\frac{|y_1-y_2|}{2}\Phi\left(\gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}\right)(y_1-y_2)\right) = E\left(\frac{|y_1-y_2|}{2}\Phi\left(\gamma^*\frac{y_1-y_2}{\sigma\sqrt{2}}\right)(y_1-y_2)\right).$$

The penultimate equality follows from the fact that for iid normal variables $y_i$, $y_1 - y_2$ and $y_1 + y_2$ are independent, and the last equality holds because for $\mu = 0$ we have $E(y_i) = 0$. The substitution $Z = \frac{y_1 - y_2}{[2(\sigma_u^2 + \sigma_e^2)]^{1/2}}$ and standard calculations show that the last term above equals

$$\frac{\sigma_u^2}{\gamma^*} E\left(|Z|\Phi\left(\sqrt{\frac{\gamma^*}{1-\gamma^*}} Z\right) Z\right),$$

where $Z$ is a standard normal random variable.

Let $a = \sqrt{\frac{\gamma^*}{1-\gamma^*}}$. Using (S3.1) we obtain

$$E\left(y_{(2)}\Phi\left(\gamma^* \frac{y_1 - y_2}{\sigma\sqrt{2}}\right)(y_1 - y_2)\right) = \frac{\sigma_u^2}{\gamma^*} E\left(|Z|\Phi\left(a\, Z\right) Z\right) = \frac{\sigma_u^2}{\gamma^*}(2\psi(a) - 1/2).$$

To calculate the second term of (S3.3) we use a result from Siegel (1993), see also Rinott and Samuel-Cahn (1994). It yields the second equality below, while the others are straightforward:

$$E\left(y_{(2)} y_2\right) = Cov(y_2, y_{(2)}) = Cov(y_2, y_2)P(y_2 = y_{(2)}) + Cov(y_2, y_1)P(y_2 = y_{(1)}) = \frac{\sigma_u^2}{2\gamma^*}.$$

The third part of (S3.3) is computed like the second part above to give

$$E\left(y_{(2)}\varphi\left(\gamma^* \frac{y_1 - y_2}{\sqrt{2}}\right)\right) = \sqrt{\frac{\sigma_u^2}{2\gamma^*}} E\left(|Z|\varphi\left(a\, Z\right)\right).$$

The latter expectation becomes

$$\int_{-\infty}^{\infty} |t|\varphi\left(a\, t\right)\varphi(t)dt = 2\int_0^{\infty} t\,\varphi\left(a\, t\right)\varphi(t)dt = \frac{1-\gamma^*}{\pi}.$$

Combining these results, we get

$$E\left(\theta_{(2)} y_{(2)}\right) = 2\sigma_u^2\psi(a) + \frac{\sigma_e^2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}. \qquad (S3.4)$$

From (S3.4) and Lemma S3.1, Lemma S3.3 follows readily. □

**Proof of Theorem 5**. It is easy to see that we can assume $\mu = 0$ without loss of generality. We use the calculations of Theorem 3. Lemma S3.3 is used instead of Lemma 1 for a better upper bound of $E\left(\theta_{(1)} y_{(1)} + \theta_{(2)} y_{(2)}\right)$ for the normal case and $m = 2$.

Below we use the notation of Lemma 1. By symmetry $E(\theta_{(1)}y_{(1)}) = E(\theta_{(2)}y_{(2)})$. Therefore, by Lemma S3.3 with $a = \sqrt{\frac{\gamma^*}{1-\gamma^*}}$, we have $E(\theta_{(1)}y_{(1)} + \theta_{(2)}y_{(2)}) \le 4\sigma_u^2\varrho_2(a) + 2\frac{\sigma_e^2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}$. By (A.2) and the above inequality we obtain

$$E\sum_{i=1}^{2}\big(y_{(i)} - \theta_{(i)}\big)\big(y_{(i)} - \overline{y}\big) = 2(\sigma_u^2 + \sigma_e^2) - \sigma_e^2 - E\sum_{i=1}^{2}\theta_{(i)}y_{(i)}$$
$$\ge 2(\sigma_u^2 + \sigma_e^2) - \sigma_e^2 - 4\sigma_u^2\varrho_2(a) - 2\frac{\sigma_e^2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}$$
$$= 2\sigma_u^2 - 4\sigma_u^2\varrho_2(a) + \sigma_e^2\left(1 - \frac{2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}\right) := \kappa(\gamma^*).$$

Recall from the proof of Theorem 3 the notation

$$D(\gamma) := E\{L(\widehat{\boldsymbol{\theta}}\,^{[2]}_{(\,)}(\gamma), \boldsymbol{\theta}_{(\,)})\} - E\{L(\widehat{\boldsymbol{\theta}}\,^{[1]}_{(\,)}, \boldsymbol{\theta}_{(\,)})\}.$$

In order to prove part 1 of Theorem 5, we have to show that its conditions imply $D(\gamma) \le 0$.

By (A.3) for $m = 2$,

$$D(\gamma) = (1-\gamma)^2(\sigma_u^2 + \sigma_e^2) - 2(1-\gamma)E\sum_{i=1}^{2}(y_{(i)} - \theta_{(i)})(y_{(i)} - \overline{y})$$
$$\le (1-\gamma)^2(\sigma_u^2 + \sigma_e^2) - 2(1-\gamma)\kappa(\gamma^*) = (1-\gamma)[(1-\gamma)(\sigma_u^2 + \sigma_e^2) - 2\kappa(\gamma^*)].$$

We assume $0 \le \gamma \le 1$ and therefore $D(\gamma) \le 0$ provided $\gamma \ge 1 - 2\frac{\kappa(\gamma^*)}{\sigma_u^2 + \sigma_e^2} =: \omega(\gamma^*)$.

For $\gamma^* = 0$ $(a = 0)$, $\omega(\gamma^*) = -1$ and clearly $D(\gamma) \le 0$ for all $\gamma$.

Next we show that in the range $0 < \gamma^* < \dfrac{\pi^2}{\pi^2 + 4} \approx 0.71$ $\left(0 < a < \dfrac{\pi}{2}\right)$ the function $\omega(\gamma^*)$ has a single zero at $c \approx 0.4119$, and $\omega(\gamma^*) < 0$ for $\gamma^* < c$. This implies that $\gamma > \omega(\gamma^*)$ and therefore $D(\gamma) < 0$.

In this range of $\gamma^*$, $\omega(\gamma^*) = 1 + 4\gamma^*\left(\frac{a}{2\pi} - \frac{1}{4}\right) - 2(1-\gamma^*)\left(1 - \frac{2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}\right)$. Substituting $\gamma^* = \dfrac{a^2}{1+a^2}$ we get $\omega(\gamma^*) = 1 + \frac{1}{1+a^2}\left(\frac{2}{\pi}a^3 - a^2 - 2 + \frac{4}{\pi}\frac{a}{1+a^2}\right)$. The function $\omega(\gamma^*)$ has the same zeros as the function $P(a) := \frac{\pi}{2}(1+a^2)^2\omega(\gamma^*)$ and straightforward calculations show that $P(a) = a^5 + a^3 - \frac{\pi}{2}a^2 + 2a - \frac{\pi}{2}$, and that this function is increasing in $a$ and therefore in $\gamma^*$. By numerical calculation we obtain that it vanishes at $c \approx 0.4119$.

The second part of Theorem 5 is proved by showing that $\gamma^* \ge \omega(\gamma^*)$ and therefore $1 \ge \gamma \ge \gamma^*$ implies $\gamma \ge \omega(\gamma^*)$.

In the range $0 < \gamma^* < \dfrac{\pi^2}{\pi^2+4} \approx 0.71$ $\left(0 < a < \dfrac{\pi}{2}\right)$, $\omega(\gamma^*) = 1+4\gamma^*\left(\frac{a}{2\pi} - \frac{1}{4}\right) - 2(1-\gamma^*)\left(1 - \frac{2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}\right) \le 1 - 2(1-\gamma^*)\frac{\pi-1}{\pi}$. Therefore, $\omega(\gamma^*) - \gamma^* \le \frac{2-\pi}{\pi}(1-\gamma^*) < 0$.

In the range $\gamma^* \ge \dfrac{\pi^2}{\pi^2+4}$, $\omega(\gamma^*) = 1 - 2(1-\gamma^*)\left(1 - \frac{2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}\right) \le 1 - 2(1-\gamma^*)\frac{\pi-1}{\pi}$. Therefore, $\omega(\gamma^*) - \gamma^* \le \frac{2-\pi}{\pi}(1-\gamma^*) < 0$.

For the proof the last part we use the same calculation as in Theorem 4 with m=2 to obtain

$$\partial E\{L(\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma), \boldsymbol{\theta}_{()})\}/\partial\gamma = 0 \quad \text{if and only if} \quad \gamma = 1 - \frac{E\sum_{i=1}^{2}\left(y_{(i)} - \theta_{(i)}\right)\left(y_{(i)} - \overline{y}\right)}{(\sigma_u^2 + \sigma_e^2)}.$$

By (A.2) we have

$$E\sum_{i=1}^{2}\left(y_{(i)} - \theta_{(i)}\right)\left(y_{(i)} - \overline{y}\right) = 2(\sigma_u^2 + \sigma_e^2) - \sigma_e^2 - E\sum_{i=1}^{2}\theta_{(i)}y_{(i)}.$$

By (S3.4) we have $E\sum_{i=1}^{2}\theta_{(i)}y_{(i)} = 4\sigma_u^2\psi(a) + 2\frac{\sigma_e^2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}$. Hence,

$$E\sum_{i=1}^{2}\left(y_{(i)} - \theta_{(i)}\right)\left(y_{(i)} - \overline{y}\right) = 2\sigma_u^2\left(1 - 2\psi(a)\right) + \sigma_e^2\left(1 - \frac{2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}\right).$$

Finally, using the convexity of $E\{L(\widehat{\theta}^{[2]}(\gamma), \theta)\}$, the optimal $\gamma$ is

$$\gamma^o = \gamma^*\left(4\psi(a) - 1\right) + (1-\gamma^*)\frac{2}{\pi}\sqrt{\gamma^*(1-\gamma^*)}. \qquad \square$$

## S4. Proof of Theorem 6

Note that $\widehat{\theta}^{[2]}_{(i)}(\gamma) = (1-\gamma)\overline{y} + \gamma g_i(y)$, and from (S3.2) $\widehat{\theta}^{[3]}_{(i)} = E_{\widehat{\mu}}\left(\theta_{(i)}|y\right) = (1-\gamma^*)\overline{y} + \gamma^* f_i(y)$, where $f_i(y)$ and $g_i(y)$ are functions of $y = (y_1, y_2)$ defined for $i = 1, 2$ by

$$f_i \equiv f_i(y) = (-1)^i\left(\Phi\left(\triangle\right)(y_1 - y_2) + \frac{\sigma}{\gamma^*}\sqrt{2}\varphi\left(\triangle\right)\right) + y_i, \quad g_i \equiv g_i(y) = y_{(i)},$$

$$\triangle = \gamma^*\frac{y_1 - y_2}{\sigma\sqrt{2}}, \quad \sigma^2 = \gamma^*\sigma_e^2.$$

We have

$$E\left((\widehat{\theta}^{[3]}_{(i)} - \theta_{(i)})^2|y\right) = Var\left(\theta_{(i)}|y\right) + \left(E\left((\theta_{(i)} - \widehat{\theta}^{[3]}_{(i)})|y\right)\right)^2$$
$$= Var\left(\theta_{(i)}|y\right) + \left((1-\gamma^*)\left(\mu - \overline{y}\right)\right)^2 = Var\left(\theta_{(i)}|y\right) + \left((1-\gamma^*)\overline{y}\right)^2,$$

where the last equality holds because for m=2 we can assume that $\mu = 0$ without loss of generality. In the same way,

$$E\left((\widehat{\theta}^{[2]}_{(i)}(\gamma) - \theta_{(i)})^2|y\right) = Var\left(\theta_{(i)}|y\right) + \left(E\left((\theta_{(i)} - \widehat{\theta}^{[2]}_{(i)}(\gamma))|y\right)\right)^2$$
$$= Var\left(\theta_{(i)}|y\right) + ((1-\gamma^*)\mu + \gamma^* f_i - (1-\gamma)\overline{y} - \gamma g_i)^2 = Var\left(\theta_{(i)}|y\right) + (\gamma^* f_i - (1-\gamma)\overline{y} - \gamma g_i)^2 .$$

Therefore,

$$\begin{aligned}
&d(\gamma) := E\{L(\widehat{\boldsymbol{\theta}}^{[3]}_{()}, \boldsymbol{\theta}_{()})\} - E\{L(\widehat{\boldsymbol{\theta}}^{[2]}_{()}(\gamma), \boldsymbol{\theta}_{()})\} \\
&= \sum_{i=1}^{2} E\left\{E\left((\widehat{\theta}^{[3]}_{(i)} - \theta_{(i)})^2|y\right)\right\} - \sum_{i=1}^{2} E\left\{E\left((\widehat{\theta}^{[2]}_{(i)}(\gamma) - \theta_{(i)})^2|y\right)\right\} \\
&= 2E\left((1-\gamma^*)\overline{y}\right)^2 - E\left(\gamma^* f_1 - \gamma g_1 - (1-\gamma)\overline{y}\right)^2 - E\left(\gamma^* f_2 - \gamma g_2 - (1-\gamma)\overline{y}\right)^2 \\
&= 2\left((1-\gamma^*)^2 - (1-\gamma)^2\right) E\left(\overline{y}^{\,2}\right) - E\left(\gamma^* f_1 - \gamma g_1\right)^2 - E\left(\gamma^* f_2 - \gamma g_2\right)^2 \\
&+ 2(1-\gamma)E\big[\left((\gamma^*(f_1+f_2) - \gamma(g_1+g_2)\right)(\overline{y})\big].
\end{aligned}$$

From the definitions of $f_i$ and $g_i$ it follows that $f_1 + f_2 - g_1 - g_2 \equiv 0$, and the last term vanishes. It is now easy to see that $d(\gamma^*) \leq 0$. □