Calibration and Partial Calibration on Principal Components when the Number of Auxiliary Variables is Large

Hervé CARDOT[1] *and Camelia GOGA*[1] *and Muhammad-Ahmed SHEHZAD*[2]

(1) IMB, Université de Bourgogne Franche Comté, DIJON, FRANCE

(2) Bahauddin Zakariya University, MULTAN, PAKISTAN

Supplementary Material

S1 Estimation of the total electricity consumption

A sample of 5 auxiliary information curves is drawn in Figure 1, corresponding to measurements over a period of 24 hours at a half an hour scale ($p = 336$).

The distribution of the proportion of positive sampling weights are given in Figure 2. It can be seen that the proportion of negative weights increases as the number of dimensions increases. Note also that this proportion is slightly smaller for the estimated principal components.

In Figure 3, the MSE is drawn for all the seven considered days and various dimensions starting from $r = 1$ to $r = 336$.

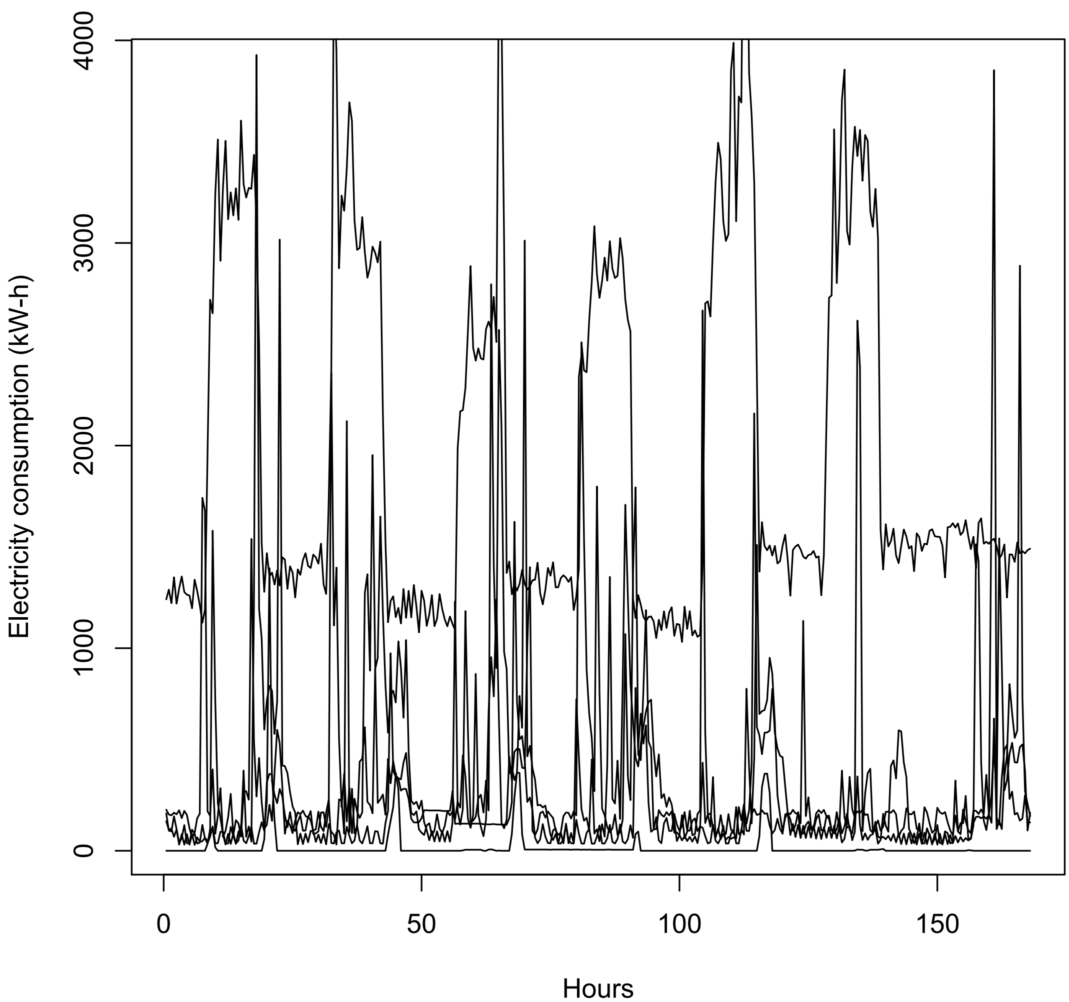

Figure 1: A sample of 5 electricity load curves observed every half an hour during the first week.

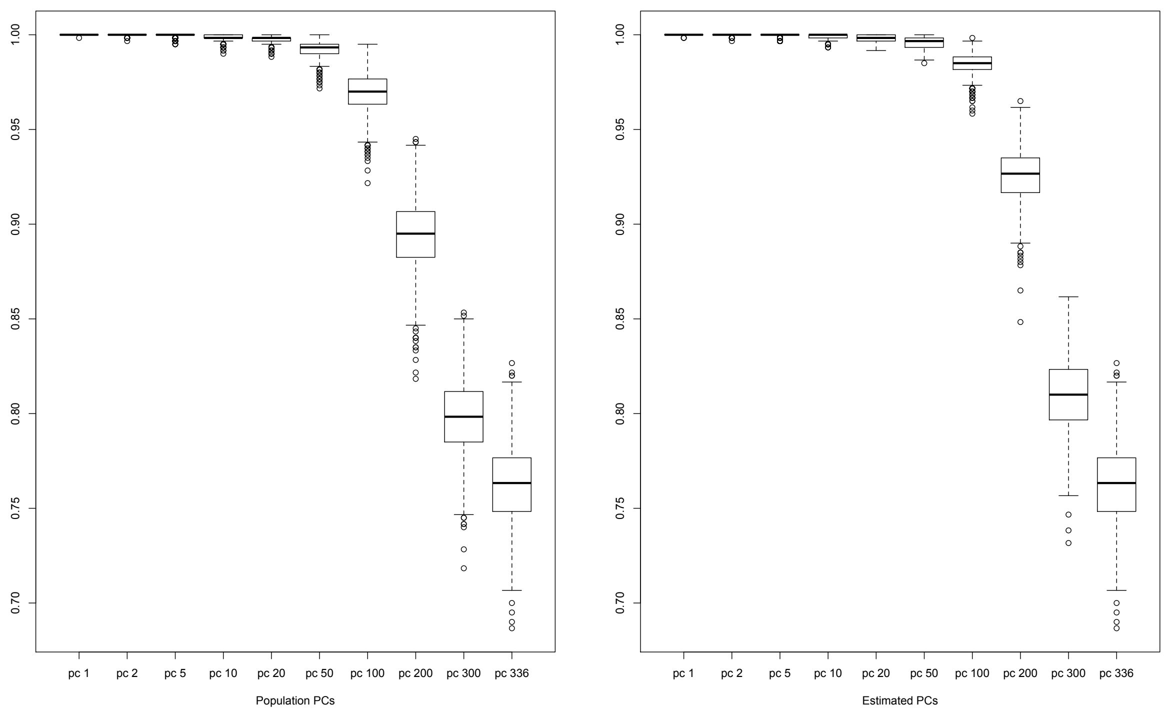

Figure 2: Proportion of positive calibrated weights for different values of the dimension r. On the right for calibration on the population principal components and on left for the sample principal components.

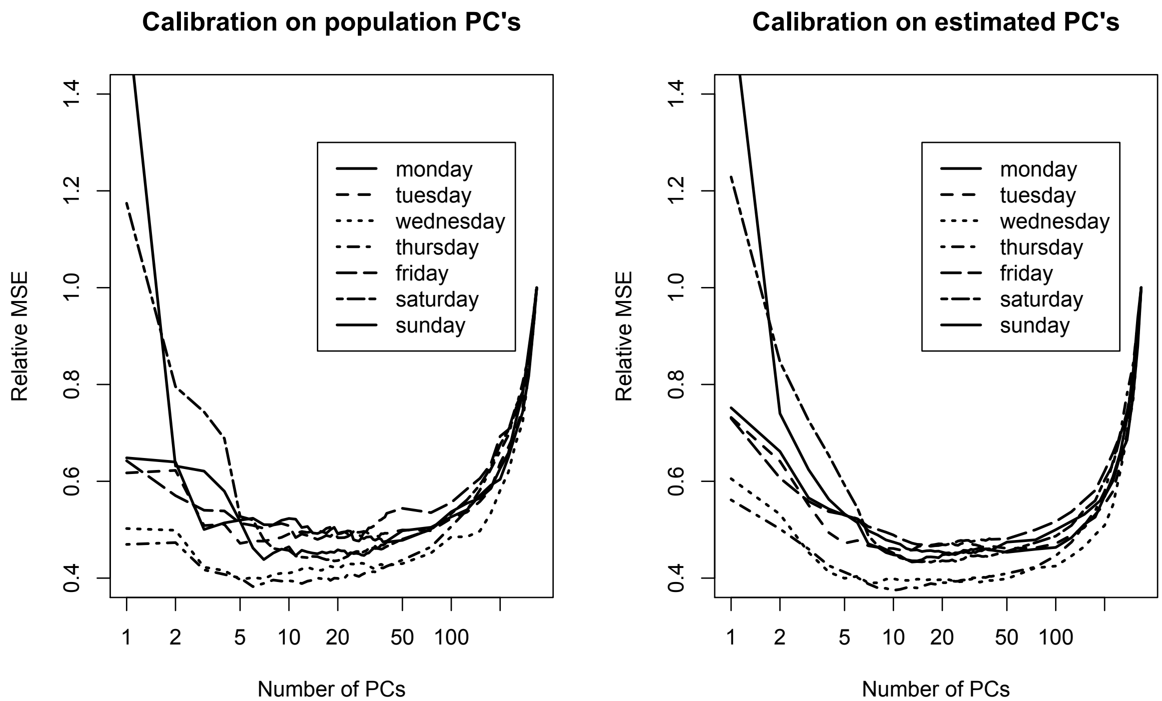

Figure 3: Relative MSE, for different values of the dimension r and the different days of the week and a sample size of $n = 600$ for calibration estimators based on the population (left) and the sample (right) principal components. The horizontal axis is at a log scale.

The calibration error has also been evaluated. It is the difference between the total of the original auxiliary variables and their "estimation" on the samples obtained with the weights $w_k^{\rm pc}(r)$ and the weights $w_k^{\rm epc}(r)$. The distribution of the estimation squared calibration error $\| \sum_{k \in s} w_k \mathbf{x}_k - t_{\mathbf{x}} \|^2$ is drawn in Figure 4 for calibration on the estimated principal components (the distribution for the population principal components calibration is very similar and not presented here). We have also plotted the distribution (in the first boxplot) of the squared error for the weights obtained with the data driven choice of the dimension r. For both approaches, the distributions of the errors are very similar. The errors are high and highly variable when the number of principal components is small (the mean value is about 1300 for $r = 1$) and then they decrease rapidly (the mean value is close to 720 for $r = 5$ and close to 600 for $r = 10$). For larger values of r, the decrease is slower. When the dimension r is not chosen in advance, the mean value of the error is roughly the same as the mean error squared corresponding to $r = 10$ but with a variability that is much larger.

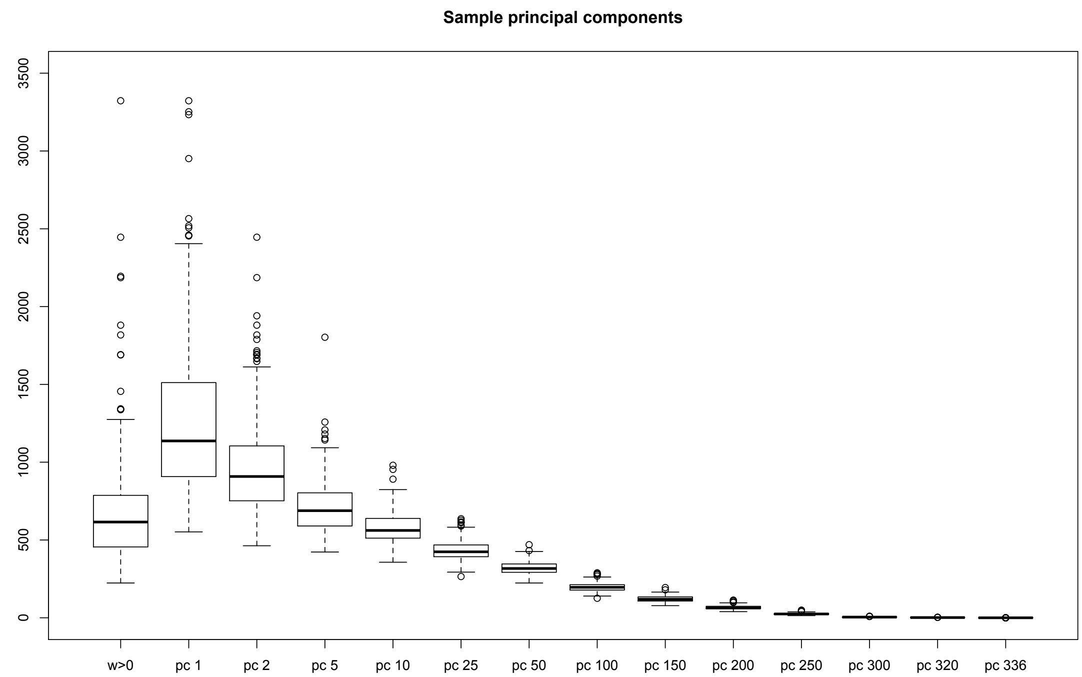

Figure 4: Calibration errors for the original variables, in terms of MSE, for different number of principal components estimated in the sample. The first boxplot ($w > 0$) corresponds to the data-driven choice of the number of components.

S2 Proofs

Throughout the proofs, we use the letter C to denote a generic constant whose value may vary from place to place. This constant does not depend on N. For sake of clarity, subscript $_N$ has been suppressed when there were no ambiguity.

For a vector $\mathbf{v}$, we denote by $\|\mathbf{v}\|$ its Euclidean norm. The spectral norm of a matrix $\mathbf{A}$ is denoted by $\|\mathbf{A}\| = \sup_{\mathbf{v}} \|\mathbf{A}\mathbf{v}\|/\|\mathbf{v}\|$. We often use the following well known inequality, $\|\mathbf{A}\|^2 \leq \mathrm{tr}\left(\mathbf{A}^T\mathbf{A}\right)$ as well as the equality $\|\mathbf{A}^T\mathbf{A}\| = \|\mathbf{A}\mathbf{A}^T\|$.

Proof of Proposition 1

We may write

$$\hat{t}_{yw}^{\mathrm{pc}}(r) - t_y = \left(\tilde{t}_{y,\mathbf{x}}^{\mathrm{diff}}(r) - t_y\right) + \left(\hat{t}_{\mathbf{z}_r d} - t_{\mathbf{z}_r}\right)^T \left(\tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r) - \hat{\boldsymbol{\gamma}}_{\mathbf{z}}(r)\right), \qquad \text{(S2.1)}$$

where

$$\tilde{t}_{y,\mathbf{x}}^{\mathrm{diff}}(r) = \hat{t}_{yd} - \left(\hat{t}_{\mathbf{z}_r d} - t_{\mathbf{z}_r}\right)^T \tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r). \qquad \text{(S2.2)}$$

We get by linearity of the Horvitz-Thompson estimators, that

$$\left(\hat{t}_{\mathbf{z}_r d} - t_{\mathbf{z}_r}\right)^T \tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r) = \widehat{t}_{\mathbf{z}_r^T \tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r)} - t_{\mathbf{z}_r^T \tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r)} \qquad \text{(S2.3)}$$

By construction, the new real variable $\mathbf{z}_r^T \tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r)$ is the projection of $(y_k)_{k \in U}$ onto the space generated by the first r principal components and we have that

$$\sum_{k \in U_N} \left(\mathbf{z}_{kr}^T \tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r)\right)^2 \leq \sum_{k \in U_N} y_k^2.$$

Consequently, we get with (S2.3) and with classical properties of Horvitz-Thompson estimators (and because assumptions (A1) and (A2) hold) that

$$\frac{1}{N}\left(\hat{t}_{\mathbf{z}_r d} - t_{\mathbf{z}_r}\right)^T \tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r) = O_p\left(\frac{1}{\sqrt{n}}\right). \qquad \text{(S2.4)}$$

Recall that $(\mathbf{Z}_1, \ldots, \mathbf{Z}_r) = \mathbf{X}\mathbf{G}_r$, and

$$N^{-1}\sum_{k\in U}\mathbf{z}_{kr}\mathbf{z}_{kr}^T = \mathbf{G}_r^T\left(N^{-1}\mathbf{X}^T\mathbf{X}\right)\mathbf{G}_r = \mathrm{diag}(\lambda_j)_{j=1}^r := \mathbf{\Lambda}_r.$$

By definition of the principal components and assumption (A5) we have that $\sum_{k\in U_N}\|\mathbf{z}_{kr}\|^2 = \left(\sum_{j=1}^r \lambda_j\right) N \leq CNr$. Following the same lines as in Breidt and Opsomer (2000), we obtain with assumptions (A1)-(A5) that

$$\frac{1}{N}\left(\hat{t}_{\mathbf{z}_r d} - t_{\mathbf{z}_r}\right) = O_p\left(\sqrt{\frac{r}{n}}\right). \tag{S2.5}$$

It remains to bound $\hat{\boldsymbol{\gamma}}_{\mathbf{z}_r} - \tilde{\boldsymbol{\gamma}}_{\mathbf{z}_r}$. For that, we first bound the estimation error of $\mathbf{\Lambda}_r$. Considering the spectral norm for square matrices, we have for some constant C, with classical algebra (see *e.g.* Cardot *et al.* 2010, Proposition 1),

$$E_p\left\|\mathbf{\Lambda}_r - \frac{1}{N}\sum_{k\in s} d_k\mathbf{z}_{kr}\mathbf{z}_{kr}^T\right\|^2 \leq \frac{C}{N^2}\sum_{k\in U_N}\|\mathbf{z}_{kr}\|^4. \tag{S2.6}$$

Expanding now each $\mathbf{z}_{kr}$ in the eigenbasis $\mathbf{v}_1, \ldots, \mathbf{v}_r$, we have $\|\mathbf{z}_{kr}\|^2 = \sum_{j=1}^r \langle x_k, v_j\rangle^2$ and thus

$$\begin{aligned}\frac{1}{N}\sum_{k\in U_N}\|\mathbf{z}_{kr}\|^4 &\leq \frac{r}{N}\sum_{k\in U_N}\sum_{j=1}^r\langle x_k, v_j\rangle^4 \\ &\leq r\sum_{j=1}^r\left[\frac{1}{N}\sum_{k\in U_N}\langle x_k, v_j\rangle^4\right] \\ &\leq C_4 r^2, \end{aligned} \tag{S2.7}$$

thanks to assumption (A6). We deduce from (S2.6) and (S2.7) that

$$\left\| \mathbf{\Lambda}_r - \frac{1}{N}\sum_{k\in s} d_k \mathbf{z}_{kr}\mathbf{z}_{kr}^T \right\| = O_p\left(\frac{r}{\sqrt{n}}\right). \tag{S2.8}$$

Note that as in Cardot *et al.* (2010), we can deduce from previous upper bounds that

$$\left\| \left(\frac{1}{N}\sum_{k\in s} d_k \mathbf{z}_{kr}\mathbf{z}_{kr}^T\right)^{-1} \right\| = \frac{1}{\lambda_r} + O_p\left(\frac{r}{\sqrt{n}}\right), \tag{S2.9}$$

and

$$\begin{aligned}\left\| \mathbf{\Lambda}_r^{-1} - \left(\frac{1}{N}\sum_{k\in s} d_k \mathbf{z}_{kr}\mathbf{z}_{kr}^T\right)^{-1} \right\| &\leq \frac{1}{\lambda_r}\left\| \left(\frac{1}{N}\sum_{k\in s} d_k \mathbf{z}_{kr}\mathbf{z}_{kr}^T\right)^{-1} \right\| \left\| \mathbf{\Lambda}_r - \left(\frac{1}{N}\sum_{k\in s} d_k \mathbf{z}_{kr}\mathbf{z}_{kr}^T\right) \right\| \\ &= O_p\left(\frac{r}{\sqrt{n}}\right). \end{aligned} \tag{S2.10}$$

An application of Cauchy-Schwarz inequality as well as the bound obtained in (S2.7), gives with assumptions (A1)-(A6) that there is some constant C such that

$$\begin{aligned}\frac{1}{N^2}E_p\left\| \sum_{k\in U_N} y_k \mathbf{z}_{kr} - \sum_{k\in s} d_k y_k \mathbf{z}_{kr}\right\|^2 &\leq \frac{C}{N^2}\sum_{k\in U_N} y_k^2 \left\|\mathbf{z}_{kr}\right\|^2 \\ &\leq \frac{C}{N^2}\left(\sum_{k\in U_N} y_k^4\right)^{1/2}\left(\sum_{k\in U_N}\left\|\mathbf{z}_{kr}\right\|^4\right)^{1/2} \\ &\leq C\frac{r}{N}. \end{aligned} \tag{S2.11}$$

Note finally, that we have for some constant C

$$\begin{aligned}\left\|\frac{1}{N}\sum_{k\in U_N}\mathbf{z}_{kr}y_k\right\|^2 &= \frac{1}{N^2}\sum_{k,\ell\in U_N} y_k y_\ell \mathbf{z}_{kr}^T\mathbf{z}_{\ell r}\\ &\leq \lambda_1 \frac{1}{N}\sum_{k\in U_N} y_k^2\\ &\leq C, \end{aligned} \tag{S2.12}$$

because the largest eigenvalue of the non negative $N\times N$ matrix $N^{-1}\mathbf{X}\mathbf{G}_r\left(\mathbf{X}\mathbf{G}_r\right)^T$ is equal to λ_1 and $N^{-1}\sum_{k\in U_N} y_k^2$ is supposed to be bounded.

Consequently, we get with previous upper bounds,

$$\begin{aligned}\|\hat{\boldsymbol{\gamma}}_{\mathbf{z}}(r)-\tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r)\| \leq& \left\|\boldsymbol{\Lambda}_r^{-1}-\left(\frac{1}{N}\sum_{k\in s} d_k\mathbf{z}_{kr}\mathbf{z}_{kr}^T\right)^{-1}\right\| \left\|\frac{1}{N}\sum_{k\in U_N}\mathbf{z}_{kr}y_k\right\|\\ &+\left\|\frac{1}{N}\sum_{k\in U_N}\mathbf{z}_{kr}y_k-\frac{1}{N}\sum_{k\in s} d_k\mathbf{z}_{kr}y_k\right\| \left\|\left(\frac{1}{N}\sum_{k\in s} d_k\mathbf{z}_{kr}\mathbf{z}_{kr}^T\right)^{-1}\right\|\\ &= O_p\left(\frac{r}{\sqrt{n}}\right)\end{aligned} \tag{S2.13}$$

and with (S2.5),

$$\frac{1}{N}\left(\hat{t}_{\mathbf{z}_r d}-t_{\mathbf{z}_r}\right)^T\left(\hat{\boldsymbol{\gamma}}_{\mathbf{z}}(r)-\tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r)\right) = O_p\left(\frac{r^{3/2}}{n}\right). \tag{S2.14}$$

Finally, using again decomposition (S2.1), we get with previous bounds

$$\frac{1}{N}\left(\hat{t}_{yw}^{\mathrm{pc}}(r)-t_y\right) = \frac{1}{N}\left(\tilde{t}_{y,\mathbf{x}}^{\mathrm{diff}}(r)-t_y\right)+O_p\left(\frac{r^{3/2}}{n}\right). \tag{S2.15}$$

□

Proof of Proposition 2

The proof follows the same lines as the proof of Proposition 1. We first write

$$\hat{t}_{yw}^{\text{epc}}(r) - t_y = \tilde{t}_{y,\mathbf{x}}^{\text{diff}}(r) - t_y + \left(\hat{t}_{\mathbf{x}d} - t_{\mathbf{x}}\right)^T \left(\tilde{\boldsymbol{\beta}}_{\mathbf{x}}^{\text{pc}}(r) - \hat{\boldsymbol{\beta}}_{\mathbf{x}}^{\text{epc}}(r)\right), \qquad \text{(S2.16)}$$

and note that, as in Proposition 1, we have that $N^{-1}\left(\tilde{t}_{y,\mathbf{x}}^{\text{diff}}(r) - t_y\right) = O_p(1/\sqrt{n})$.

We now look for an upper bound on the second term at the right-hand side of equality (S2.16). It can be shown easily that $N^{-1}\left(\hat{t}_{\mathbf{x}d} - t_{\mathbf{x}}\right) = O_p(\sqrt{p/n})$.

We can also write

$$\tilde{\boldsymbol{\beta}}_{\mathbf{x}}^{\text{pc}}(r) - \hat{\boldsymbol{\beta}}_{\mathbf{x}}^{\text{epc}}(r) = \mathbf{G}_r\left(\tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r) - \hat{\boldsymbol{\gamma}}_{\hat{\mathbf{z}}}(r)\right) + \left(\mathbf{G}_r - \hat{\mathbf{G}}_r\right)\hat{\boldsymbol{\gamma}}_{\hat{\mathbf{z}}}(r) \qquad \text{(S2.17)}$$

and bound each term at the right-hand side of the equality.

We denote by $\hat{\mathbf{G}}_r = (\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_r)$ the matrix whose columns are the orthonormal eigenvectors of $\hat{\boldsymbol{\Gamma}}$ associated to the r largest eigenvalues, $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_r \geq 0$. Note that these eigenvectors are unique up to sign change and, for $j = 1, \ldots, r$, we choose $\hat{\mathbf{v}}_j$ such that $\langle \hat{\mathbf{v}}_j, \mathbf{v}_j \rangle \geq 0$. Since $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are orthonormal vectors, the spectral norm of matrix $\mathbf{G}_r$ satisfies $\|\mathbf{G}_r\| = 1$. This is also true for $\hat{\mathbf{G}}_r$, and we have $\left\|\hat{\mathbf{G}}_r\right\| = 1$.

Now, using the fact that $N^{-1}\hat{N} = 1 + O_p(n^{-1/2})$ and $\|\widehat{\overline{\mathbf{X}}}\| = O_p(p/\sqrt{n})$, it can be shown that

$$\left\|\widehat{\boldsymbol{\Gamma}}-\frac{1}{N}\mathbf{X}^T\mathbf{X}\right\| = O_p\left(\frac{p}{\sqrt{n}}\right). \tag{S2.18}$$

We deduce with Lemma 4.3 in Bosq (2000) and equation (S2.18) that

$$\begin{aligned}\max_{j=1,\dots,p}|\lambda_j-\hat{\lambda}_j| &\leq \left\|\widehat{\boldsymbol{\Gamma}}-\frac{1}{N}\mathbf{X}^T\mathbf{X}\right\| \\ &= O_p\left(\frac{p}{\sqrt{n}}\right), \end{aligned} \tag{S2.19}$$

and

$$\|\mathbf{v}_j-\hat{\mathbf{v}}_j\| \leq \delta_j\left\|\widehat{\boldsymbol{\Gamma}}-\frac{1}{N}\mathbf{X}^T\mathbf{X}\right\|, \tag{S2.20}$$

with $\delta_1 = 2\sqrt{2}/(\lambda_1-\lambda_2)$ and $\delta_j = 2\sqrt{2}/(\min(\lambda_{j-1}-\lambda_j,\lambda_j-\lambda_{j+1}))$ for $j=2,\cdots,r$.

Consequently,

$$\begin{aligned}\left\|\mathbf{G}_r-\hat{\mathbf{G}}_r\right\|^2 &\leq \operatorname{tr}\left[\left(\mathbf{G}_r-\hat{\mathbf{G}}_r\right)^T\left(\mathbf{G}_r-\hat{\mathbf{G}}_r\right)\right] \\ &\leq \sum_{j=1}^r\|\mathbf{v}_j-\hat{\mathbf{v}}_j\|^2 \\ &\leq \sum_{j=1}^r\delta_j^2\left\|\widehat{\boldsymbol{\Gamma}}-\frac{1}{N}\mathbf{X}^T\mathbf{X}\right\|^2 \\ &= O_p\left(\frac{p^2r^3}{n}\right),\end{aligned}$$

with (S2.18), (S2.20) and the fact that $\max_{j=1,\cdots,r}\delta_j^2 = O(r^2)$ which comes from the fact that we have supposed that $\min_{j=1,\dots,r+1}(\lambda_j-\lambda_{j+1}) \geq c_\lambda r$ with $c_\lambda > 0$.

We also deduce from (S2.19) that

$$\left\| \frac{1}{N} \sum_{k \in U_N} \mathbf{z}_{kr} \mathbf{z}_{kr}^T - \frac{1}{N} \sum_{k \in s} d_k \hat{\mathbf{z}}_{kr} \hat{\mathbf{z}}_{kr}^T \right\| \leq \max_{j=1,\ldots,r} |\lambda_j - \hat{\lambda}_j| = O_p\left(\frac{p}{\sqrt{n}}\right). \qquad \text{(S2.21)}$$

Employing a similar technique as before (see the bound obtained in (S2.10)), we also get that

$$\left\| \left(\frac{1}{N} \sum_{k \in U_N} \mathbf{z}_{kr} \mathbf{z}_{kr}^T \right)^{-1} - \left(\frac{1}{N} \sum_{k \in s} d_k \hat{\mathbf{z}}_{kr} \hat{\mathbf{z}}_{kr}^T \right)^{-1} \right\| = O_p\left(\frac{p}{\sqrt{n}}\right), \qquad \text{(S2.22)}$$

and

$$\left\| \left(\frac{1}{N} \sum_{k \in s} d_k \hat{\mathbf{z}}_{kr} \hat{\mathbf{z}}_{kr}^T \right)^{-1} \right\|^2 = \frac{1}{\lambda_r^2} + O_p\left(\frac{p^2}{n}\right).$$

Using the inequality $\mathrm{tr}(\mathbf{AB}) \leq \|\mathbf{A}\| \ \mathrm{tr}(\mathbf{B})$ for any symmetric non negative matrices $\mathbf{A}$ and $\mathbf{B}$, we also have

$$\begin{aligned}
\left\| \frac{1}{N} \sum_{k \in s} d_k \hat{\mathbf{z}}_{kr} y_k \right\|^2 &= \frac{1}{N^2} \sum_{k,l \in s} d_k d_l y_k y_l \mathbf{x}_l^T \hat{\mathbf{G}}_r^T \hat{\mathbf{G}}_r \mathbf{x}_k \\
&= \frac{1}{N^2} \sum_{k,l \in s} d_k d_l y_k y_l \ \mathrm{tr}\left(\hat{\mathbf{G}}_r^T \hat{\mathbf{G}}_r \mathbf{x}_l \mathbf{x}_k^T \right) \\
&= \frac{1}{N^2} \mathrm{tr}\left[\hat{\mathbf{G}}_r^T \hat{\mathbf{G}}_r \left(\sum_{k,l \in s} d_k d_l y_k y_l \mathbf{x}_l \mathbf{x}_k^T \right) \right] \\
&\leq \left\| \hat{\mathbf{G}}_r^T \hat{\mathbf{G}}_r \right\| \frac{1}{N^2} \mathrm{tr}\left[\sum_{k,l \in s} d_k d_l y_k y_l \mathbf{x}_l \mathbf{x}_k^T \right] \\
&\leq \lambda_1 \frac{1}{N} \sum_{k \in s} (d_k y_k)^2
\end{aligned}$$

because $\left\|\hat{\mathbf{G}}_r^T\hat{\mathbf{G}}_r\right\| = 1$ and the largest eigenvalue of the non negative $N \times N$ matrix $N^{-1}\mathbf{X}\mathbf{X}^T$ is equal to λ_1. Note also that $N^{-1}\sum_{k\in s}(d_k y_k)^2 \leq C$, for some constant C, because $N^{-1}\sum_{k\in U_N} y_k^2$ is supposed to be bounded and $\max d_k^2 \leq \delta^{-2}$ (see Assumption A2).

We can now bound $\|\hat{\boldsymbol{\gamma}}_{\hat{\mathbf{z}}}(r)\|$. Combining previous inequalities, we have

$$\begin{aligned}\|\hat{\boldsymbol{\gamma}}_{\hat{\mathbf{z}}}(r)\|^2 &\leq \left\|\left(\frac{1}{N}\sum_{k\in s} d_k \hat{\mathbf{z}}_{kr}\hat{\mathbf{z}}_{kr}^T\right)^{-1}\right\|^2 \left\|\frac{1}{N}\sum_{k\in s} d_k \hat{\mathbf{z}}_{kr} y_k\right\|^2 \\ &= O_p(1). \end{aligned} \tag{S2.23}$$

Let us study now $N^{-1}\left(\sum_{k\in U_N}\mathbf{z}_{kr}y_k - \sum_{k\in s} d_k\hat{\mathbf{z}}_{kr}y_k\right)$. Writing $\mathbf{z}_{kr} - \hat{\mathbf{z}}_{kr} = \left(\mathbf{G}_r - \hat{\mathbf{G}}_r\right)^T \mathbf{x}_k$, we have

$$\begin{aligned}\frac{1}{N^2}\left\|\sum_{k\in U_N}(\mathbf{z}_{kr} - \hat{\mathbf{z}}_{kr})\, y_k\right\|^2 &= \frac{1}{N^2}\sum_{k,\ell\in U_N}\mathbf{x}_k^T\left(\mathbf{G}_r - \hat{\mathbf{G}}_r\right)\left(\mathbf{G}_r - \hat{\mathbf{G}}_r\right)^T\mathbf{x}_\ell y_k y_\ell \\ &= \frac{1}{N^2}\mathrm{tr}\left[\left(\mathbf{G}_r - \hat{\mathbf{G}}_r\right)\left(\mathbf{G}_r - \hat{\mathbf{G}}_r\right)^T\left(\sum_{k,\ell\in U_N}\mathbf{x}_\ell\mathbf{x}_k^T y_k y_\ell\right)\right] \\ &\leq \left\|\mathbf{G}_r - \hat{\mathbf{G}}_r\right\|^2 \frac{1}{N^2}\left\|\sum_{k\in U_N}\mathbf{x}_k y_k\right\|^2 \\ &\leq \left\|\mathbf{G}_r - \hat{\mathbf{G}}_r\right\|^2 \lambda_1\left(\frac{1}{N}\sum_{k\in U_N} y_k^2\right).\end{aligned}$$

because the largest eigenvalue of the non negative $N\times N$ matrix $N^{-1}\mathbf{X}\mathbf{G}_r\left(\mathbf{X}\mathbf{G}_r\right)^T$ is equal to λ_1. Since $N^{-1}\sum_{k\in U_N} y_k^2$ is supposed to be bounded, we obtain

$$\left\|\frac{1}{N}\left(\sum_{k\in U_N}(\mathbf{z}_{kr} - \hat{\mathbf{z}}_{kr})\, y_k\right)\right\| = O_p\left(\frac{pr^{3/2}}{\sqrt{n}}\right). \tag{S2.24}$$

Define $\alpha_k = 1 - \mathbb{1}_{\{k\in s\}} d_k$ and remember that $\hat{\mathbf{z}}_{kr} = \hat{\mathbf{G}}_r^T \mathbf{x}_k$. We have that

$$\left\| \frac{1}{N} \sum_{k\in U_N} \hat{\mathbf{z}}_{kr} y_k - \sum_{k\in s} d_k \hat{\mathbf{z}}_{kr} y_k \right\| = \left\| \hat{\mathbf{G}}_r^T \left(\frac{1}{N} \sum_{k\in U_N} \alpha_k \mathbf{x}_k y_k \right) \right\|$$
$$\leq \left\| \hat{\mathbf{G}}_r^T \right\| \left\| \frac{1}{N} \sum_{k\in U_N} \alpha_k \mathbf{x}_k y_k \right\|$$
$$= O_p\left(\sqrt{\frac{p}{n}} \right). \tag{S2.25}$$

Combining (S2.24) and (S2.25), we finally obtain that

$$\left\| N^{-1} \left(\sum_{k\in U_N} \mathbf{z}_{kr} y_k - \sum_{k\in s} d_k \hat{\mathbf{z}}_{kr} y_k \right) \right\| = O_p\left(\frac{p r^{3/2}}{\sqrt{n}} \right). \tag{S2.26}$$

Hence, using now a decomposition similar to (S2.13), we obtain

$$\|\hat{\boldsymbol{\gamma}}_{\hat{\mathbf{z}}}(r) - \tilde{\boldsymbol{\gamma}}_{\mathbf{z}}(r)\| \leq \left\| \boldsymbol{\Lambda}_r^{-1} - \left(\frac{1}{N} \sum_{k\in s} d_k \hat{\mathbf{z}}_{kr} \hat{\mathbf{z}}_{kr}^T \right)^{-1} \right\| \left\| \frac{1}{N} \sum_{k\in U_N} \mathbf{z}_{kr} y_k \right\|$$
$$+ \left\| \frac{1}{N} \sum_{k\in U_N} \mathbf{z}_{kr} y_k - \frac{1}{N} \sum_{k\in s} d_k \hat{\mathbf{z}}_{kr} y_k \right\| \left\| \left(\frac{1}{N} \sum_{k\in s} d_k \hat{\mathbf{z}}_{kr} \hat{\mathbf{z}}_{kr}^T \right)^{-1} \right\|$$
$$= O_p\left(\frac{p r^{3/2}}{\sqrt{n}} \right). \tag{S2.27}$$

Combining previous bounds we get

$$\left\| \frac{1}{N} \left(\hat{t}_{\mathbf{x}d} - t_{\mathbf{x}} \right)^T \left(\tilde{\boldsymbol{\beta}}_{\mathbf{x}}^{\mathrm{pc}}(r) - \hat{\boldsymbol{\beta}}_{\mathbf{x}}^{\mathrm{epc}}(r) \right) \right\| = O_p\left(\sqrt{\frac{p}{n}} \right) O_p\left(\frac{p r^{3/2}}{\sqrt{n}} \right)$$

and using again decomposition (S2.16), we finally get

$$\frac{1}{N} \left(\hat{t}_{yw}^{\mathrm{epc}}(r) - t_y \right) = \frac{1}{N} \left(\tilde{t}_{y,\mathbf{x}}^{\mathrm{diff}}(r) - t_y \right) + O_p\left(\frac{p^{3/2} r^{3/2}}{n} \right).$$

□

References

Bosq, D. (2000). *Linear processes in function spaces.* Lecture Notes in Statistics, Vol. 149. Springer-Verlag.

Breidt, J.F. and Opsomer, J.D. (2000). Local polynomial regression estimators in survey sampling. *Ann. Statist.* **28**, 1023-1053.

Cardot, H., Chaouch, M., Goga, C. and Labruère, C. (2010). Properties of design-based functional principal components analysis. *Journal of Statistical and Planning Inference* **140**, 75-91.