Biometrika (2016), **xx**, x, *pp.* 1–7

Printed in Great Britain

Supplementary material for Partial least squares for dependent data

BY MARCO SINGER, TATYANA KRIVOBOKOVA, AXEL MUNK
Institute for Mathematical Stochastics, Goalgorithm2e.styldschmidtstr. 7, 37077 Göttingen, Germany
msinger@gwdg.de tkrivob@uni-goettingen.de munk@math.uni-goettingen.de

AND BERT DE GROOT
Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
bgroot@gwdg.de

S1. PROOFS

S1·1. *Derivation of the population partial least squares components*

Let denote $K_i \in \mathbb{R}^{k\times i}$ the matrix representation of a base for $\mathcal{K}_i(\Sigma^2, Pq)$. Then

$$\sum_{t=1}^{n} E\left(y_t - X_t^{\mathrm{T}} K_i\alpha\right)^2 = \sum_{t=1}^{n}[V^2]_{t,t}\left(\|q\|^2 + \eta_2^2 - 2\alpha^{\mathrm{T}} K_i^{\mathrm{T}} Pq + \alpha^{\mathrm{T}} K_i^{\mathrm{T}}\Sigma^2 K_i\alpha\right).$$

Minimizing this expression with respect to $\alpha \in \mathbb{R}^i$ gives $K_i^{\mathrm{T}}\Sigma^2 K_i\alpha = K_i Pq$. Since the matrix $K_i^{\mathrm{T}}\Sigma^2 K_i$ is invertible, we get the least squares fit β_i in Section 2.

Assume now that the first $i < a$ partial least squares base vectors $w_1, \dots, w_i$ have been calculated and consider for $\lambda \in \mathbb{R}$ the Lagrange function

$$\sum_{t,s=1}^{n} \operatorname{cov}\left(y_t - X_t^{\mathrm{T}}\beta_i, X_s^{\mathrm{T}} w\right) - \lambda(\|w\|^2 - 1) = w^{\mathrm{T}}\left(Pq - \Sigma^2\beta_i\right)\sum_{t,s=1}^{n}[V^2]_{t,s} - \lambda(\|w\|^2 - 1).$$

Maximizing with respect to w yields

$$w_{i+1} = (2\lambda)^{-1}\left(Pq - \Sigma^2\beta_i\right)\sum_{t,s=1}^{n}[V^2]_{t,s} \propto Pq - \Sigma^2\beta_i.$$

Since $\beta_i \in \mathcal{K}_i(\Sigma^2, Pq)$, we get $w_{i+1} \in \mathcal{K}_{i+1}(\Sigma^2, Pq)$ and w_{i+1} is orthogonal to $w_1, \dots, w_i$.

S1·2. *Proof of Theorem* 1

First consider

$$\begin{aligned} E\left(\|b - Pq\|^2\right) =& E\left[\left\|\frac{1}{\|V\|^2}\left\{(PN^{\mathrm{T}} + \eta_1 F^{\mathrm{T}})V^2 Nq + \eta_2(PN^{\mathrm{T}} + \eta_1 F^{\mathrm{T}})V^2 f\right\} - Pq\right\|^2\right] \\ =& \left\{E\left(\left\|\frac{1}{\|V\|^2} PN^{\mathrm{T}} V^2 Nq - Pq\right\|^2\right) + \frac{\eta_2^2}{\|V\|^4} E\left(\left\|PN^{\mathrm{T}} V^2 f\right\|^2\right)\right\} \\ &+ \frac{\eta_1^2}{\|V\|^4}\left\{E\left(\left\|F^{\mathrm{T}} V^2 Nq\right\|^2\right) + \eta_2^2 E\left(\left\|F^{\mathrm{T}} V^2 f\right\|^2\right)\right\} = S_1 + S_2, \end{aligned}$$

due to the independence of N, F and f. It is easy to see that

$$S_2 = \frac{\|V^2\|^2}{\|V\|^4}\eta_1^2 k\left(\|q\|^2+\eta_2^2\right).$$

Furthermore, with notation $A_0 = N^{\mathrm{T}}V^2N$ we get

$$S_1 = \frac{1}{\|V\|^4}E\left(q^{\mathrm{T}}A_0P^{\mathrm{T}}PA_0q\right) - \|Pq\|^2 + \frac{\eta_2^2}{\|V\|^4}E\left(\left\|PN^{\mathrm{T}}V^2f\right\|^2\right).$$

Consider now $E\left(q^{\mathrm{T}}A_0P^{\mathrm{T}}PA_0q\right)$ as a quadratic form with respect to the matrix $P^{\mathrm{T}}P$. Denote $\kappa = E\left(N_{1,1}^4\right) - 3$. First, $E\left(A_0q\right) = E\left(N^{\mathrm{T}}V^2Nq\right) = \|V\|^2 q$ and

$$\begin{aligned}
\mathrm{var}(A_0q) &= \left[\sum_{a,b=1}^{l} q_aq_b \sum_{t,s,u,v=1}^{n} V_u^{\mathrm{T}}V_sV_t^{\mathrm{T}}V_vE(N_{s,i}N_{u,a}N_{t,j}N_{v,b})\right]_{i,j=1}^{l} - \|V\|^4qq^{\mathrm{T}} \\
&= \left[q_iq_j\|V\|^4 + \left(q_iq_j+\delta_{i,j}\|q\|^2\right)\|V^2\|^2 + \kappa\sum_{t=1}^{n}\|V_t\|^4\delta_{i,j}q_i^2\right]_{i,j=1}^{l} - \|V\|^4qq^{\mathrm{T}} \\
&= \|V^2\|^2\left(qq^{\mathrm{T}} + \|q\|^2I_l\right) + \kappa\sum_{t=1}^{n}\|V_t\|^4\mathrm{diag}\left(q_1^2,\ldots,q_l^2\right),
\end{aligned}$$

where $\mathrm{diag}(v_1,\ldots,v_l)$ denotes the diagonal matrix with entries $v_1,\ldots,v_l\in\mathbb{R}$ on its diagonal and δ is the Kronecker delta. In the second equation we made use of $E\left(N_{s,i}N_{u,a}N_{t,j}N_{v,b}\right) = \delta_{i,a}\delta_{j,b}\delta_{s,u}\delta_{t,v} + \delta_{i,b}\delta_{j,a}\delta_{s,v}\delta_{t,u} + \delta_{i,j}\delta_{a,b}\delta_{t,s}\delta_{u,v} + \kappa\,\delta_{t,s}\delta_{s,u}\delta_{u,v}\delta_{i,j}\delta_{j,a}\delta_{a,b}$. Hence,

$$\begin{aligned}
\frac{1}{\|V\|^4}E\left(q^{\mathrm{T}}A_0P^{\mathrm{T}}PA_0q\right) =& \frac{1}{\|V\|^4}\mathrm{tr}\left\{P^{\mathrm{T}}P\mathrm{var}\left(A_0q\right)\right\} - \frac{1}{\|V\|^4}E\left(q^{\mathrm{T}}A_0\right)P^{\mathrm{T}}PE\left(A_0q\right) \\
=& \frac{\|V^2\|^2}{\|V\|^4}\left(q^{\mathrm{T}}P^{\mathrm{T}}Pq + \|P\|^2\|q\|^2\right) + q^{\mathrm{T}}P^{\mathrm{T}}Pq + \kappa\sum_{t=1}^{n}\frac{\|V_t\|^4}{\|V\|^4}\sum_{i=1}^{l}\|P_i\|^2q_i^2.
\end{aligned}$$

The remaining term in S_1 follows trivially, proving the result. $E\|\Sigma^2 - A\|^2$ is obtained using similar calculations. □

S1·3. *Proof of Theorem* 2

LEMMA S1. *Assume that for* $\nu\in(0,1]$ *and some constants* $\delta,\epsilon>0$ *it holds that* $\mathrm{pr}\left(\|A-\Sigma^2\|_{\mathcal{L}}\le\delta\right)\ge 1-\nu/2$ *and* $\mathrm{pr}\left(\|b-Pq\|\le\epsilon\right)\ge 1-\nu/2$. *Then each of the inequalities*

$$\begin{aligned}
\|A^{1/2}-\Sigma\| &\le 2^{-1}\delta\|\Sigma^{-1}\|\{1+o(1)\}, \\
\|A^{-1/2}b - \Sigma^{-1}Pq\| &\le \epsilon\|\Sigma^{-1}\|_{\mathcal{L}} + 2^{-1}\delta(\|Pq\|+\epsilon)\|\Sigma^{-2}\|\|\Sigma^{-1}\|\{1+o(1)\}
\end{aligned}$$

hold with probability at least $1-\nu/2$.

Proof: We show the result by using the Fréchet-derivative for functions $F:\mathbb{R}^{k\times k}\to\mathbb{R}^{k\times k}$. Due to the fact that $\eta_1>0$ it holds that Σ^2 is positive definite and thus invertible.

It holds due to Higham (2008), Problem 7.4, that $F'(\Sigma^2)B$ for an arbitrary $B\in\mathbb{R}^{k\times k}$ is given as the solution in $X\in\mathbb{R}^{k\times k}$ of $B=\Sigma X + X\Sigma$, i.e. due to the symmetry and positive definitiness of Σ we have $F'(\Sigma^2)B = 2^{-1}\Sigma^{-1}B$. We take the orthonormal base $\{E_{i,j}, i,j=1,\ldots,k\}$

for the space $(\mathbb{R}^{k\times k}, \|\cdot\|)$ with $E_{i,j}$ corresponding to the matrix that has zeros everywhere except at the position (i,j), where it is one. The Hilbert-Schmidt norm $\|F'(\Sigma^2)\|_{HS}$ is

$$\|F'(\Sigma^2)\|_{HS}^2 = 4^{-1}\sum_{i,j=1}^{k}\|\Sigma^{-1}E_{i,j}\|^2 = 4^{-1}\sum_{i,j=1}^{k}[\Sigma^{-1}]_{i,j}^2 = 4^{-1}\|\Sigma^{-1}\|^2.$$

This yields with the Taylor expansion for Fréchet-differentiable maps

$$\|A^{1/2}-\Sigma\|_{\mathcal{L}} \le \|F'(\Sigma)(A-\Sigma^2)\| + o(\|A-\Sigma^2\|) \le 2^{-1}\|\Sigma^{-1}\|\delta\{1+o(1)\}.$$

For the second inequality we see first that

$$\|A^{-1/2}b - \Sigma^{-1}Pq\| \le \epsilon\|\Sigma^{-1}\|_{\mathcal{L}} + \left\|(A^{-1/2}-\Sigma^{-1})b\right\|. \tag{S1}$$

The Fréchet-derivative of the map $F:\mathbb{R}^{k\times k}\to\mathbb{R}^{k\times k}, A\mapsto A^{-1/2}$ is $F'(\Sigma^2)B = -2^{-1}\Sigma^{-2}B\Sigma^{-1}$ and

$$\|F'(\Sigma^2)\|_{HS}^2 = 4^{-1}\sum_{i,j=1}^{k}\|\Sigma^{-2}E_{i,j}\Sigma^{-1}\|^2 \le 4^{-1}\|\Sigma^{-2}\|^2\|\Sigma^{-1}\|^2.$$

Here we used the submultiplicativity of the Frobenius norm with the Hadamard product of matrices. Thus we get via Taylor's theorem

$$\|A^{-1/2}-\Sigma^{-1}\| \le 2^{-1}\|\Sigma^{-2}\|\|\Sigma^{-1}\|\|A-\Sigma^2\| + o(\delta).$$

Plugging this into (S1) yields

$$\|A^{-1/2}b-\Sigma^{-1}Pq\| \le \epsilon\|\Sigma^{-1}\|_{\mathcal{L}} + 2^{-1}\delta(\|Pq\|+\epsilon)\|\Sigma^{-2}\|\|\Sigma^{-1}\|\left\{1+o(1)\right\},$$

where we used that $\|b\| \le \|Pq\| + \epsilon$. □

Equivalence of conjugate gradient and partial least squares: We denote $\tilde{A} = A^{1/2}$ and $\tilde{b} = A^{-1/2}b$. The partial least squares optimization problem is

$$\min_{v\in\mathcal{K}_i(A,b)} \|y - Xv\|^2,$$

whereas the conjugate gradient problem studied in Nemirovskii (1986) is

$$\min_{v\in\mathcal{K}_i(\tilde{A}^2,\tilde{A}\tilde{b})} \|\tilde{b} - \tilde{A}v\|^2. \tag{S2}$$

It is easy to see that the Krylov space $\mathcal{K}_i(\tilde{A}^2,\tilde{A}\tilde{b}) = \mathcal{K}_i(A,b)$ $(i=1,\ldots,k)$. We have

$$\arg\min_{v\in\mathcal{K}_i(\tilde{A}^2,\tilde{A}\tilde{b})} \|\tilde{b}-\tilde{A}v\|^2 = \arg\min_{\mathcal{K}_i(A,b)} \|y-Xv\|^2, i=1,\ldots,k.$$

Thus it holds

$$\widehat{\beta}_i = \arg\min_{v\in\mathcal{K}_i(\tilde{A}^2,\tilde{A}\tilde{b})} \|\tilde{b}-\tilde{A}v\|^2,$$

Furthermore we have $\Sigma\beta(\eta_1) = \Sigma^{-1}Pq$, i.e. the correct problem in the population is solved by $\beta(\eta_1)$ as well. Now we will restate the main result in Nemirovskii (1986) in our context:

THEOREM S1. *Nemirovskii*
Assume that there are $\tilde{\delta} = \tilde{\delta}(\nu,n) > 0$, $\tilde{\epsilon} = \tilde{\epsilon}(\nu,n) > 0$ such that for $\nu\in(0,1]$ it holds that

$\mathrm{pr}\left(\|\Sigma - A^{1/2}\|_{\mathcal{L}} \leq \tilde{\delta}\right) \geq 1-\nu/2$, $\mathrm{pr}\left(\|\Sigma^{-1}Pq - A^{-1/2}b\| \leq \tilde{\epsilon}\right) \geq 1-\nu/2$ *and the conditions*

1. *there is an* $L = L(\nu, n)$ *such that with probability at least* $1-\nu/2$ *it holds that* $\max\left\{\|A^{1/2}\|_{\mathcal{L}}, \|\Sigma\|_{\mathcal{L}}\right\} \leq L$,
2. *there is a vector* $u \in \mathbb{R}^k$ *and constants* $R, \mu > 0$ *such that* $\beta(\eta_1) = \Sigma^{\mu}u$, $\|u\| \leq R$

are satisfied. If we stop according to the stopping rule a^* *as defined in* (4) *with* $\tau \geq 1$ *and* $\zeta < \tau^{-1}$ *then we have for any* $\theta \in [0,1]$ *with probability at least* $1-\nu$

$$\left\|\Sigma^{\theta}\{\widehat{\beta}_{a^*} - \beta(\eta_1)\}\right\|^2 \leq C^2(\mu,\tau,\zeta) R^{2(1-\theta)/(1+\mu)} \left(\tilde{\epsilon} + \tilde{\delta} R L^{\mu}\right)^{2(\theta+\mu)/(1+\mu)}.$$

Proof: Note first that on the set where $\|\Sigma - A^{1/2}\|_{\mathcal{L}} \leq \tilde{\delta}$ holds with probability at least $1-\nu/2$ condition 1 also holds with $L = \|\Sigma\|_{\mathcal{L}} + \tilde{\delta}$. Constrained on the set where all the conditions of the theorem hold with probability at least $1-\nu$ we consider Nemirovskii's $(\Sigma, A^{1/2}, \Sigma^{-1}Pq, A^{-1/2}b)$ problem with errors $\tilde{\delta}$ and $\tilde{\epsilon}$. Furthermore by assumption Nemirovskii's $(2\theta, R, L, 1)$ conditions hold and thus the theorem follows by a simple application of the main theorem in Nemirovskii (1986). □

We will now apply Theorem S1 to our problem. Due to the fact that $\eta_1 > 0$ it holds that Σ^2 is positive definite and thus invertible. We note that the spectral norm is dominated by the Frobenius norm. From Markov's inequality we get

$$\mathrm{pr}\left(\|A - \Sigma^2\| \geq \delta\right) \leq \delta^{-2} E\left(\left\|A - \Sigma^2\right\|^2\right).$$

Using Theorem 1, $\sum_{t=1}^{n} \|V_i\|^4 \leq \|V^2\|^2$ and setting the right hand side to $\nu/2$ for $\nu \in (0,1]$ gives $\delta = \nu^{-1/2}\|V\|^{-2}\|V^2\|C_{\delta}$. In the same way $\epsilon = \nu^{-1/2}\|V\|^{-2}\|V^2\|C_{\epsilon}$. Lemma S1 gives with probability at least $1-\nu/2$ the concentration results required by Theorem S1 with

$$\tilde{\delta} = \nu^{-1/2}\frac{\|V^2\|}{\|V\|^2}C_{\delta}\{1+o(1)\}$$
$$\tilde{\epsilon} = \left(\nu^{-1/2}\frac{\|V^2\|}{\|V\|^2}C_{\epsilon} + \nu^{-1}\frac{\|V^2\|^2}{\|V\|^4}C_{\epsilon}C_{\delta}\right)\{1+o(1)\}$$

Conditions 1 and 2 of Theorem S1 hold with a probability of at least $1-\nu/2$ by choosing $L = \tilde{\delta} + \|\Sigma\|_{\mathcal{L}}$, $\mu = 1$ and $R = \|\Sigma^{-3}Pq\|$. Here we used that $\beta(\eta_1) = \Sigma^{-2}Pq$. Thus the theorem yields for $\theta = 1$

$$\left\|\Sigma\{\beta(\eta_1) - \widehat{\beta}_{a^*}\}\right\| \leq C(1,\tau,\zeta)\left(\tilde{\epsilon} + \tilde{\delta}RL\right).$$

Denote $c(\tau,\zeta) = C(1,\tau,\zeta)\{1+o(1)\}$. Finally we have $\|\Sigma^{-1}\|_{\mathcal{L}}^{-1}\|v\| \leq \|\Sigma v\|$ for any $v \in \mathbb{R}^k$ and thus the theorem is proven with

$$c_1(\nu) = \nu^{-1/2}c(\tau,\zeta)\|\Sigma^{-1}\|_{\mathcal{L}}\left(C_{\epsilon} + \|\Sigma\|_{\mathcal{L}}\|\Sigma^{-3}Pq\|C_{\delta}\right)$$
$$c_2(\nu) = \nu^{-1}c(\tau,\zeta)\|\Sigma^{-1}\|_{\mathcal{L}}\left(C_{\epsilon}C_{\delta} + \|\Sigma^{-3}Pq\|C_{\delta}^2\right).$$

□

S1·4. *Proof of Theorem* 3

The theorem is proved by contradiction. Assume that $\widehat{\beta}_1 \longrightarrow \beta_1$ in probability. Choosing $v \in \mathbb{R}^k$, $v \neq 0$, orthogonal to β_1 implies that $v^{\mathrm{T}}\widehat{\beta}_1$ converges in probability to zero. Next we show that the second moment vanishes as well.

Let $M_d(z) = \max_{i\in\{1,\ldots,n\}^d} E(\prod_{\nu=1}^d z_{i_\nu}^2)$ for a random vector $z = (z_1, \ldots, z_n)^{\mathrm{T}}$ with existing mixed $(2d)$th moments. Using $(a+b)^2 \leq 2(a^2+b^2)$ for $a, b \in \mathbb{R}$ we obtain

$$\begin{aligned} E\,(v^{\mathrm{T}}b)^4 \leq & \frac{8^2\|v\|^4}{\|V\|^8} E\left(\left\|PN^{\mathrm{T}}V^2Nq\right\|^4 + \eta_1^4\left\|F^{\mathrm{T}}V^2Nq\right\|^4 + \eta_2^4\left\|PN^{\mathrm{T}}V^2f\right\|^4 + \eta_1^4\eta_2^4\left\|F^{\mathrm{T}}V^2f\right\|^4\right) \\ \leq & 8^2\|v\|^4\left\{M_4(N_1)\|q\|^4 l^4\|P\|^4 + M_2(N_1)M_2(F_1)\eta_1^4\|q\|^4 l^2k^2\right. \\ & \left. + M_2(N_1)M_2(f_1)\eta_2^4 l^2\|P\|^4 + M_2(F_1)M_2(f_1)\eta_1^4\eta_2^4k^2\right\} < \infty, \quad n \in \mathbb{N}. \end{aligned}$$

Thus, $(v^{\mathrm{T}}b)^2$ is uniformally integrable by the theorem of de la Vallée-Poussin and it follows that the directional variance $\mathrm{var}(v^{\mathrm{T}}b)$ has to vanish in the limit as well. Now, calculations similar to Theorem 1 yield

$$\begin{aligned} \mathrm{var}(v^{\mathrm{T}}b) = & \frac{\|V^2\|^2}{\|V\|^4}\left\{\eta_1^2\|v\|^2\left(\|q\|^2+\eta_2^2\right) + \|P^{\mathrm{T}}v\|^2\left(\|q\|^2+\eta_2^2\right) + (v^{\mathrm{T}}Pq)^2\right\} \\ & + \sum_{t=1}^n \frac{\|V_t\|^4}{\|V\|^4}\sum_{i=1}^l q_i^2\left(v^{\mathrm{T}}P_i\right)^2\{E(N_{1,1}^4)-3\}, \quad v \in \mathbb{R}^k. \end{aligned}$$

We assumed that $\|V\|^{-2}\|V^2\|$ does not converge to zero. It remains to check under which conditions $\mathrm{var}(v^{\mathrm{T}}b)$ is larger than zero. This will always be the case if $v \neq 0$ and $\eta_1 > 0$, $l = 1$. For $\eta_1 = 0$ and $l > 1$ a vector v that lies in the range of P and is orthogonal to $\beta_1 \propto Pq$ exists, thus contradicting $\widehat{\beta}_1 \longrightarrow \beta_1$ in probability. □

S1·5. *Proof of Theorem* 4

It is easy to verify that $\|V\|^2 = \mathrm{tr}(T^2) = n\gamma(0)$ and $\left\|V^2\right\|^2 = n\gamma^2(0) + 2\sum_{t=1}^{n-1}\gamma^2(t)(n-t)$. If (6) is fulfilled, then

$$n\gamma(0) \leq \left\|V^2\right\|^2 \leq n\gamma^2(0)\left\{1 + 2c^2\frac{1-\exp(-2\rho(n-1))}{\exp(2\rho)-1}\right\} \leq n\gamma^2(0)\left\{1 + \frac{2c}{\exp(2\rho)-1}\right\}.$$

It follows that $\|V^2\| \sim n^{1/2}$. □

S1·6. *Proof of Theorem* 5

Let $\gamma : \mathbb{N} \to \mathbb{R}$ be the autocovariance function of a stationary time series that has zero mean. For the autocovariance matrix V^2 of the corresponding integrated process of order one we get $\left[V^2\right]_{t,s} = \sum_{i,j=1}^{t,s}\gamma(|i-j|)$, $(t, s = 1, \ldots, n)$. Let $t \geq s$. By splitting the sum into parts with $i < j$ and $i > j$ we get $\left[V^2\right]_{t,s} = s\gamma(0) + \sum_{j=1}^s\sum_{i=1}^{t-j}\gamma(i) + \sum_{j=2}^s\sum_{i=1}^{j-1}\gamma(i)$. Due to symmetry, $\left[V^2\right]_{t,s} = \left[V^2\right]_{s,t}$ for $s > t$.

First, consider the case that all $\gamma(j)$, $j > 0$ are negative. Using (6) we obtain

$$\gamma(0)s \geq \left[V^2\right]_{t,s} \geq \gamma(0)\left\{s - c\sum_{j=1}^s\sum_{i=1}^{t-j}\exp(-\rho j) - c\sum_{j=2}^s\sum_{i=1}^{j-1}\exp(-\rho j)\right\}, \; t \geq s.$$

Evaluation of the geometric sums gives

$$\left[V^2\right]_{t,s} \geq \gamma(0)\left(s\left\{1-\frac{2c}{\exp(\rho)-1}\right\}+c\frac{\exp(\rho)}{\{\exp(\rho)-1\}^2}\left\{1-\exp(-\rho s)\right\}\left[1+\exp\{\rho(s-t)\}\right]\right).$$

The second term on the right is always positive and the positivity of the first term is ensured by the condition $\rho > \log(2c+1)$. Hence, $\gamma(0)\left[1-2c\left\{\exp(\rho)-1\right\}^{-1}\right] s \leq \left[V^2\right]_{t,s} \leq \gamma(0)s,\ s \geq 1.$ If $\gamma(t),\, t \geq 1$ is not purely negative, it can be bound by

$$\gamma(0)\left[1-2c\left\{\exp(\rho)-1\right\}^{-1}\right] s \leq \left[V^2\right]_{t,s} \leq \gamma(0)\left[1+2c\left\{\exp(\rho)-1\right\}^{-1}\right] s.$$

We write δ_1 and δ_2 for the constants in the lower and upper bound, respectively, so that $\delta_1 \min\{s,t\} \leq \left[V^2\right]_{t,s} \leq \delta_2 \min\{s,t\}$ $(t, s = 1,\dots,n)$. This yields upper and lower bounds on the trace of V^2 and shows that $\|V\|^2 \sim n^2$. Additionally,

$$\left[V^4\right]_{t,t} = \sum_{l=1}^{n}\left[V^2\right]_{t,l}\left[V^2\right]_{l,t} = \sum_{l=1}^{t}\left[V^2\right]_{t,l}^2 + \sum_{l=t+1}^{n}\left[V^2\right]_{l,t}^2 \leq \frac{\delta_2^2}{6}t\left(6nt-4t^2+3t+1\right)$$

$$\left[V^4\right]_{t,t} \geq \frac{\delta_1^2}{6}t\left(6nt-4t^2+3t+1\right).$$

This implies upper and lower bounds on the trace of V^4 in the form $c\,n(n+1)(n^2+n+1)$ for $c \in \{\delta_1^2/6, \delta_2^2/6\}$ and thus $\|V^2\| \sim n^2$. □

S1·7. *Proof of Theorem* 6

First consider $n^{-1}X^{\mathrm{T}}\widehat{V}^{-2}y$. Define $X_u = (X_{u,1},\dots,X_{u,n})^{\mathrm{T}} = NP^{\mathrm{T}} + \eta_1 F$ and $y_u = (y_{u,1},\dots,y_{u,n})^{\mathrm{T}} = Nq + \eta_2 f$ such that $X = VX_u$ and $y = Vy_u$. By the triangle inequality

$$\left\|n^{-1}X^{\mathrm{T}}\widehat{V}^{-2}y - Pq\right\| \leq \left\|n^{-1}X^{\mathrm{T}}V^{-2}y - Pq\right\| + \left\|n^{-1}X^{\mathrm{T}}\left(\widehat{V}^{-2}-V^{-2}\right)y\right\|.$$

The first term on the right hand side is convergent to zero in probability due to Theorem 1. The second term can be bounded as

$$n^{-2}\left\|X^{\mathrm{T}}\left(\widehat{V}^{-2}-V^{-2}\right)y\right\|^2 \leq \|V\widehat{V}^{-2}V - I_n\|_{\mathcal{L}}^2\, n^{-1}\|X_u^{\mathrm{T}}\|_{\mathcal{L}}^2\, n^{-1}\|y_u\|^2.$$

Since both $X_{u,1},\dots,X_{u,n}$ and $y_{u,1},\dots,y_{u,n}$ are independent and identically distributed, it follows that $n^{-1}\|y_u\|^2$ is a strongly consistent estimator for $E(y_{u,1}^2)$, as well as that $n^{-1}\|X_u^{\mathrm{T}}\|_{\mathcal{L}}^2$ is bounded from above by $n^{-1}\|X_u^{\mathrm{T}}\|^2$, which is a strongly consistent estimator of $E(\|X_{u,1}\|^2)$. Convergence in probability of $\left\|V\widehat{V}^{-2}V - I_n\right\|_{\mathcal{L}}^2$ to zero implies the convergence of $b(\widehat{V})$ to Pq in probability. To obtain the convergence rate $\|n^{-1}X^{\mathrm{T}}V^{-2}y - Pq\| = O_p(r_n)$, use Theorem 1 and $\|V\widehat{V}^{-2}V - I_n\|_{\mathcal{L}} = O_p(r_n)$. The convergence of $\|n^{-1}X^{\mathrm{T}}\widehat{V}^{-2}X - \Sigma^2\|$ is proven in a similar way.

To show the consistency and the rate of the corrected partial least squares estimator, we follow the same lines as in the proof of Theorem 2. First, $\delta = r_n c_A(\nu)$ and $\epsilon = r_n c_b(\nu)$ for $\nu \in (0,1]$ with constants $c_A(\nu)$, $c_b(\nu)$ are taken, such that

$$\mathrm{pr}\{\|A(\widehat{V})^{1/2} - \Sigma\|_{\mathcal{L}} \leq r_n c_A(\nu)\} \geq 1-\nu/2,$$
$$\mathrm{pr}\{\|A(\widehat{V})^{-1/2}b(\widehat{V}) - \Sigma^{-1}Pq\| \leq r_n c_b(\nu)\} \geq 1-\nu/2.$$

Moreover, $L = \|\Sigma\|_{\mathcal{L}} + \delta$ and $R = \|\Sigma^{-3}Pq\|$, $\mu = 1$, satisfies conditions 1 and 2 in Theorem S1 with probability at least $1-\nu/2$. Thus, with probability at least $1-\nu$ we get by setting $\theta = 1$

$$\left\|\widehat{\beta}_{a^*}(\widehat{V}) - \beta(\eta_1)\right\| \leq r_n\, C(1,\tau,\zeta)\{1+o(1)\}\|\Sigma^{-1}\|_{\mathcal{L}}\left[c_b(\nu) + c_A(\nu)\|\Sigma^{-3}Pq\|\left\{\|\Sigma\|_{\mathcal{L}} + r_n c_A(\nu)\right\}\right],$$

where the constants ζ, τ are taken from the definition of a^*. □

S2. ADDENDUM TO SECTION 5, SIMULATIONS

Figure S1 shows the differences in empirical mean squared error of $\hat{\beta}_1$ for various dependence structures considered in Section 5 in the setting with $l = i = 1$. We calculated

$$n\text{MSE}(\hat{\beta}_1) = n\,500^{-1}\sum_{i=1}^{500}(\hat{\beta}_{1,i} - \beta_1)^2,$$

where $\hat{\beta}_{1,i}$ denotes a partial least squares estimator in the ith Monte Carlo simulation based on n observations. If an autoregressive dependence is present in the data and is ignored in the partial least squares algorithm, $n\text{MSE}(\hat{\beta}_1)$ is proportional to a constant, which is larger than in the corrected partial least squares case. Ignoring the integrated dependence in the data leads to $n\text{MSE}(\hat{\beta}_1)$ growing linearly with n, which confirms our theoretical findings in Section 3.

REFERENCES

HIGHAM, N. (2008). *Functions of Matrices: Theory and Computation*. Phildadelphia: SIAM, 1st ed.

NEMIROVSKII, A. (1986). The regularizing properties of the adjoint gradient method in ill-posed problems. *Comput. Math. Math. Phys.* **26**, 7–16.

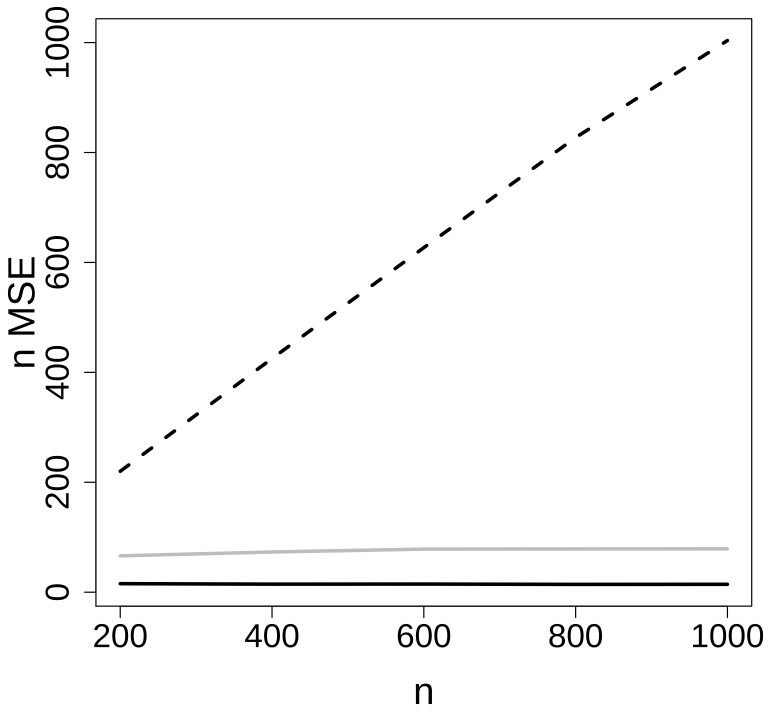

Fig. S1: Empirical mean squared eror of $\widehat{\beta}_1$ multiplied by n. The dependence structures are: autoregressive (grey), autoregressive integrated moving average (black, dashed) and corrected partial least squares on integrated data (black, solid).