Supplementary material for
Partial least squares for dependent data

BY MARCO SINGER, TATYANA KRIVOBOKOVA, AXEL MUNK

Institute for Mathematical Stochastics, Goalgorithm2e.stylschnidstr. 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. 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 $K_i(\Sigma^2, Pq)$ . Then

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

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

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

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

Maximizing with respect to $w$ yields

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

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

S1-2. Proof of Theorem 1

First consider

$$
E \left( \|b - Pq\|^2 \right) = E \left( \left\| \frac{1}{\|V\|^2} \left( (PN^T + \eta_1 F^T) V^2 Nq + \eta_2 (PN^T + \eta_1 F^T) V^2 f \right) - Pq \right\|^2 \right)
$$

$$
= \left\{ E \left( \left\| \frac{1}{\|V\|^2} PN^T V^2 Nq - Pq \right\|^2 \right) + \frac{\eta_2^2}{\|V\|^4} E \left( \|PN^T V^2 f\|^2 \right) \right\} + \frac{\eta_1^2}{\|V\|^4} \left\{ E \left( \|F^T V^2 Nq\|^2 \right) + \eta_2^2 E \left( \|F^T V^2 f\|^2 \right) \right\} = S_1 + S_2,
$$
due to the independence of \( N, F \) and \( f \). It is easy to see that

\[
S_2 = \frac{\|V^2\|^2}{\|V\|^4} \eta_2^2 k (\|q\|^2 + \eta_2^2).
\]

Furthermore, with notation \( A_0 = N^T V^2 N \) we get

\[
S_1 = \frac{1}{\|V\|^4} E(q^T A_0 P^T P A_0 q) - \|Pq\|^2 + \frac{\eta_2^2}{\|V\|^4} E\left( \|P N^T V^2 f\|^2 \right).
\]

Consider now \( E(q^T A_0 P^T P A_0 q) \) as a quadratic form with respect to the matrix \( P^T P \). Denote \( \kappa = E(N^4_{1,1}) - 3 \). First, \( E(A_0 q) = E(N^T V^2 N q) = \|V\|^2 q \) and

\[
\text{var}(A_0 q) = \left[ \sum_{a,b=1}^l q_a q_b \sum_{t,s,u,v=1}^n V^a_t V^s_t V^v_t E(n_{a,i} n_{a,i} n_{t,j} n_{v,b}) \right]_{i,j=1}^l - \|V\|^4 q^T q
\]

\[
= \left[ q_i q_j \|V\|^4 + (q_i q_j + \delta_{i,j}) \|q\|^2 \right] \|V^2\|^2 + \kappa \sum_{t=1}^n \|V_t\|^4 \delta_{i,j} \|q\|^2 - \|V\|^4 q^T q
\]

\[
= \|V^2\|^2 (q^T q + \|q\|^2 I_l) + \kappa \sum_{t=1}^n \|V_t\|^4 \text{diag} (q^2_1, \ldots, q^2_l),
\]

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

\[
\frac{1}{\|V\|^4} E(q^T A_0 P^T P A_0 q) = \frac{1}{\|V\|^4} \text{tr} \{ P^T P \text{var}(A_0 q) \} - \frac{1}{\|V\|^4} E(q^T A_0) P^T P E(A_0 q)
\]

\[
= \|V^2\|^2 q^T P q + \|P\|^2 \|q\|^2 + \kappa \sum_{t=1}^n \|V_t\|^4 \|P_t\|^2 \|q_t\|^2.
\]

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

\[ \square \]

**Lemma S1.** Assume that for \( \nu \in (0, 1) \) and some constants \( \delta, \epsilon > 0 \) it holds that \( \Pr \left( \|A - \Sigma^2\|_\mathcal{L} \leq \delta \right) \geq 1 - \nu/2 \) and \( \Pr \left( \|b - Pq\| \leq \epsilon \right) \geq 1 - \nu/2 \). Then each of the inequalities

\[
\|A^{1/2} - \Sigma\|_{\mathcal{L}} \leq 2^{-1} \delta \|\Sigma^{-1}\| \|1 + o(1)\|,
\]

\[
\|A^{-1/2} b - \Sigma^{-1} Pq\| \leq \epsilon \|\Sigma^{-1}\|_{\mathcal{L}} + 2^{-1} \delta (\|Pq\| + \epsilon) \|\Sigma^{-2}\| \|\Sigma^{-1}\| \|1 + o(1)\|
\]

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

**Proof:** We show the result by using the Frechét-derivative for functions \( F : \mathbb{R}^{k \times k} \rightarrow \mathbb{R}^{k \times k} \). Due to the fact that \( \eta_1 > 0 \) it holds that \( \Sigma^2 \) 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 \( B = X \Sigma + \Sigma X \), i.e. due to the symmetry and positive definiteness of \( \Sigma \) 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
\[
\|F'(\Sigma^2)\|^2_{HS} = 4^{-1} \sum_{i,j=1}^{k} \|\Sigma^{-1}E_{i,j}\|^2 = 4^{-1} \sum_{i,j=1}^{k} (\Sigma^{-1})_{i,j}^2 = 4^{-1}\|\Sigma^{-1}\|^2.
\]
This yields with the Taylor expansion for Fréchet-differentiable maps
\[
\|A^{1/2} - \Sigma\|_C \leq \|F'(\Sigma)(A - \Sigma^2)\| + o(\|A - \Sigma^2\|) \leq 2^{-1}\|\Sigma^{-1}\|\delta\{1 + o(1)\}.
\]
For the second inequality we see first that
\[
\|A^{-1/2}b - \Sigma^{-1}Pq\| \leq \epsilon\|\Sigma^{-1}\|C + \epsilon\|A^{-1/2} - \Sigma^{-1}\|\delta\|Pq\|.
\]
The Fréchet-derivative of the map \(F : \mathbb{R}^{k \times k} \rightarrow \mathbb{R}^{k \times k}, A \mapsto A^{-1/2}\) is \(F'(\Sigma^2)B = -2^{-1}\Sigma^{-2}B\Sigma^{-1}\) and
\[
\|F'(\Sigma^2)\|^2_{HS} = 4^{-1} \sum_{i,j=1}^{k} \|\Sigma^{-2}E_{i,j}\|_2^2 \leq 4^{-1}\|\Sigma^{-2}\|\|\Sigma^{-1}\|^2.
\]
Here we used the submultiplicativity of the Frobenius norm with the Hadamard product of matrices. Thus we get via Taylor’s theorem
\[
\|A^{-1/2} - \Sigma^{-1}\| \leq 2^{-1}\|\Sigma^{-2}\|\|\Sigma^{-1}\|\|A - \Sigma^2\| + o(\delta).
\]
Plugging this into (S1) yields
\[
\|A^{-1/2}b - \Sigma^{-1}Pq\| \leq \epsilon\|\Sigma^{-1}\|C + 2^{-1}\delta(\|Pq\| + \epsilon)\|\Sigma^{-2}\|\|\Sigma^{-1}\|^2\{1 + o(1)\},
\]
where we used that \(\|b\| \leq \|Pq\| + \epsilon.\) \(\square\)

**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
\[
\min_{v \in \mathcal{K}_{i}(\tilde{A}, \tilde{b})} \|y - Xv\|^2,
\]
whereas the conjugate gradient problem studied in Nemirovskii (1986) is
\[
\min_{v \in \mathcal{K}_{i}(\tilde{A}^2, \tilde{b})} \|\tilde{b} - \tilde{A}v\|^2. \tag{S2}
\]
It is easy to see that the Krylov space \(\mathcal{K}_{i}(\tilde{A}^2, \tilde{b}) = \mathcal{K}_{i}(A, b)\) \((i = 1, \ldots, k)\). We have
\[
\arg \min_{v \in \mathcal{K}_{i}(\tilde{A}^2, \tilde{b})} \|\tilde{b} - \tilde{A}v\|^2 = \arg \min_{v \in \mathcal{K}_{i}(A, b)} \|y - Xv\|^2, \quad i = 1, \ldots, k.
\]
Thus it holds
\[
\hat{\beta}_i = \arg \min_{v \in \mathcal{K}_{i}(\tilde{A}^2, \tilde{b})} \|\tilde{b} - \tilde{A}v\|^2,
\]
Furthermore we have \(\Sigma\hat{\beta}(\eta_i) = \Sigma^{-1}Pq\), i.e. the correct problem in the population is solved by \(\beta(\eta_i)\) as well. Now we will restate the main result in Nemirovskii (1986) in our context:  

**Theorem S1. Nemirovskii**

*Assume that there are \(\delta = \delta(\nu, n) > 0, \bar{\epsilon} = \bar{\epsilon}(\nu, n) > 0\) such that for \(\nu \in (0, 1]\) it holds that*
We will now apply Theorem S1 to our problem. Due to the fact that with probability at least \(1 - \nu/2\) it holds that

\[
\max \{ \| A^{1/2} \|_{L^2}, \| \Sigma \|_{L^2} \} \leq L,
\]

1. there is an \(L = L(\nu, n)\) such that with probability at least \(1 - \nu/2\) it holds that
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 \(\alpha^*\) as defined in (4) with \(\tau \geq 1\) and \(\varsigma < \tau^{-1}\) then we have for any \(\theta \in [0, 1]\) with probability at least \(1 - \nu\)

\[
\| \Sigma \theta \{ \beta_{\alpha^*} - \beta(\eta_1) \} \|^2 \leq C^2(\mu, \tau, \zeta) R^{2(1 - \theta)/(1 + \mu)} (\bar{\epsilon} + \delta R L)^{(2(\theta + \mu))/(1 + \mu)}.
\]

Proof: Note first that on the set where \(\| \Sigma - A^{1/2} \|_{L^2} \leq \bar{\delta}\) holds with probability at least \(1 - \nu/2\) condition 1 also holds with \(L = \| \Sigma \|_{L^2} + \bar{\delta}\). Constrained on the set where all the conditions of the theorem hold with probability at least \(1 - \nu\) we consider Nemirovskii’s \((\Sigma, A^{1/2}, \Sigma - 1) P q, A^{-1/2} b\) problem with errors \(\bar{\delta}\) and \(\bar{\epsilon}\). 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). \(\square\)

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

\[
\text{pr}\left( \|A - \Sigma^2\| \leq \bar{\delta}\right) \leq \bar{\delta}^{-2} E \left( \|A - \Sigma^2\|^2 \right).
\]

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

\[
\bar{\delta} = \nu^{-1/2} \|V^2\| C_\delta \{1 + o(1)\}
\]

\[
\bar{\epsilon} = \left( \nu^{-1/2} \|V^2\| C_\epsilon + \nu^{-1/2} \|V^2\|^2 \|C\| \|C_\delta\| \right) \{1 + o(1)\}
\]

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

\[
\|\Sigma \theta \{ \beta_{\alpha^*} - \beta(\eta_1) \} \| \leq C(1, \tau, \zeta) \left( \bar{\epsilon} + \delta R L \right).
\]

Denote \(c(\tau, \zeta) = C(1, \tau, \zeta) \{1 + o(1)\}\). Finally we have \(\|\Sigma^{-1}\|_{L^2}^{-1} \|v\| \leq \|\Sigma v\|\) for any \(v \in \mathbb{R}^k\) and thus the theorem is proven with

\[
c_1(\nu) = \nu^{-1/2} \|c(\tau, \zeta)\| \|\Sigma^{-1}\|_{L^2} \left( C_\epsilon + \|\Sigma\|_{L^2} \|\Sigma^{-3} P q\| C_\delta \right)
\]

\[
c_2(\nu) = \nu^{-1} \|c(\tau, \zeta)\| \|\Sigma^{-1}\|_{L^2} \left( C_\epsilon C_\delta + \|\Sigma^{-3} P q\| C_\delta^2 \right).
\]

\(\square\)
S1.4. Proof of Theorem 3

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

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

\[
E(v^2 b)^4 \leq \frac{8^2 ||v||^4}{||v||^4} E \left( \left| \prod_{v=1}^{d} T N V^2 N q \right| \right)^4 + \eta_1^4 \left| \prod_{v=1}^{d} T N V^2 f \right|^4 + \eta_2^4 \left| \prod_{v=1}^{d} T N V^2 f \right|^4
\]

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

\[
\text{var}(v^T b) = \frac{||V^2||^2}{||V||^4} \left\{ \eta_2^2 ||v||^2 (||q||^2 + \eta_2^2) + ||P^T v||^2 (||q||^2 + \eta_2^2) + (v^T P q)^2 \right\}
\]

\[
+ \sum_{i=1}^{n} \frac{||V_i||^4}{||v||^4} \sum_{i=1}^{l} q_i^2 (v^T P_i)^2 \left\{ E(N_{i,1}^4) - 3 \right\}, \quad v \in \mathbb{R}^k.
\]

We assumed that \( ||V||^2 ||V^2|| \) does not converge to zero. It remains to check under which conditions \( \text{var}(v^T b) \) 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 \( \beta_1 \propto P q \) exists, thus contradicting \( \hat{\beta}_1 \rightarrow \beta_1 \) in probability.

S1.5. Proof of Theorem 4

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

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

It follows that \( ||V||^2 \sim n^{1/2} \).

S1.6. Proof of Theorem 5

Let \( \gamma : \mathbb{N} \rightarrow \mathbb{R} \) 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

\[
[V^2]_{t,s} = \sum_{i,j=1}^{t,s} \gamma(|i - j|), \quad (t, s = 1, \ldots, n). \]

Let \( t \geq s \). By splitting the sum into parts with \( i < j \) and \( i > j \) we get \( [V^2]_{t,s} = s \gamma(0) + \sum_{j=1}^{s} \sum_{i=1}^{t-j} \gamma(i) + \sum_{j=2}^{s} \sum_{i=1}^{j-1} \gamma(i) \). Due to symmetry, \( [V^2]_{t,s} = [V^2]_{s,t} \) for \( s > t \).

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

\[
\gamma(0)s \geq [V^2]_{t,s} \geq \gamma(0) \left\{ s - c \sum_{j=1}^{s} \sum_{i=1}^{t-j} \exp(-\rho j) - c \sum_{j=2}^{s} \sum_{i=1}^{j-1} \exp(-\rho j) \right\}, \quad t \geq s.
\]
Evaluation of the geometric sums gives

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

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

If \(\gamma(t), t \geq 1\) is not purely negative, it can be bound by

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

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

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

\[
[V^2]_{t,t} \geq \frac{\delta_2^2}{6} t (6nt - 4t^2 + 3t + 1).
\]

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

---

S1.7. **Proof of Theorem 6**

First consider \(n^{-1}X^\top \hat{V}^{-2}y\). Define \(X_u = (X_{u,1}, \ldots, X_{u,n})^\top = N P^\top + \eta_1 F\) and \(y_u = (y_{u,1}, \ldots, y_{u,n})^\top = N q + \eta_2 F\) such that \(X = VX_u\) and \(y = VX_u\). By the triangle inequality

\[
\left\| n^{-1}X^\top \hat{V}^{-2}y - Pq \right\| \leq \left\| n^{-1}X^\top V^{-2}y - Pq \right\| + \left\| n^{-1}X^\top \left( \hat{V}^{-2} - V^{-2} \right) y \right\|.
\]

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

\[
n^{-2} \left\| X^\top \left( \hat{V}^{-2} - V^{-2} \right) y \right\|^2 \leq \left\| V \hat{V}^{-2} V - I_n \right\|_F^2 n^{-1} \| X_u \|_F^2 n^{-1} \| y_u \|^2.
\]

Since both \(X_{u,1}, \ldots, X_{u,n}\) and \(y_{u,1}, \ldots, y_{u,n}\) 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 \(n^{-1} \| X_u \|_F^2\) is bounded from above by \(n^{-1} \| X_u \|_F^2\), which is a strongly consistent estimator of \(E(\|X_{u,1}\|_F^2)\).

Convergence in probability of \(\left\| V \hat{V}^{-2} V - I_n \right\|_F\) to zero implies the convergence of \(b(\hat{V})\) to \(Pq\) in probability. To obtain the convergence rate \(\|n^{-1}X^\top V^{-2}y - Pq\| = O_p(r_n)\), use Theorem 1 and \(\| V \hat{V}^{-2} V - I_n \|_F = O_p(r_n)\). The convergence of \(\|n^{-1}X^\top V^{-2}X - \Sigma^2\|\) 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

\[
\Pr\{\|A(\hat{V})^{1/2} - \Sigma\|_F \leq r_n c_A(\nu)\} \geq 1 - \nu/2,
\]

\[
\Pr\{\|A(\hat{V})^{1/2} b(\hat{V}) - \Sigma^{-1} Pq\| \leq r_n c_b(\nu)\} \geq 1 - \nu/2.
\]
Moreover, \( L = \| \Sigma \|_L + \delta \) and \( R = \| \Sigma^{-3} P q \| \), \( \mu = 1 \), 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 \( \theta = 1 \)

\[
\| \hat{\beta}_{a^*}(\hat{V}) - \beta(\eta_1) \| \leq r_n C(1, \tau, \zeta) \{ 1 + o(1) \} \| \Sigma^{-1} \|_L \left[ c_b(\nu) + c_A(\nu) \| \Sigma^{-3} P q \| \right]^{\{ \| \Sigma \|_L + r_n c_A(\nu) \}},
\]

where the constants \( \zeta, \tau \) are taken from the definition of \( a^* \).

\[ \square \]

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

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

where \( \hat{\beta}_{1,i} \) denotes a partial least squares estimator in the \( i \)th 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, \( n_{MSE}(\hat{\beta}_1) \) 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 \( n_{MSE}(\hat{\beta}_1) \) growing linearly with \( n \), which confirms our theoretical findings in Section 3.

**References**


Fig. S1: Empirical mean squared error of $\hat{\beta}_1$ 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).