Supplementary Materials for "Tensor Envelope Partial Least Squares Regression"

Xin Zhang and Lexin Li

Florida State University and University of California, Berkeley

1 Proofs and Technical Details

Proof of Lemma 1

Proof. From the vectorized linear model (3), we can see that $\boldsymbol{B}_{(m+1)}$ is the regression coefficient matrix for the multivariate linear regression of $\boldsymbol{Y}$ on $\text{vec}(\boldsymbol{X})$. Therefore $\boldsymbol{B}_{(m+1)} = \text{cov}^{-1}\{\text{vec}(\boldsymbol{X})\}\text{cov}\{\text{vec}(\boldsymbol{X}), \boldsymbol{Y}\} = \left(\boldsymbol{\Sigma}_m^{-1} \otimes \cdots \otimes \boldsymbol{\Sigma}_1^{-1}\right) \boldsymbol{C}_{(m+1)}$. It follows from the basic property of Tucker operator that $\boldsymbol{B} = [\![\boldsymbol{C}; \boldsymbol{\Sigma}_1^{-1}, \ldots, \boldsymbol{\Sigma}_m^{-1}, \boldsymbol{I}_r]\!]$. □

Proof of Lemma 2

Proof. Recall that in the tensor PLS regression, $\text{E}(\boldsymbol{Y} \mid \boldsymbol{X}) = \boldsymbol{B}_{(m+1)}\text{vec}(\boldsymbol{X}) = \text{E}(\boldsymbol{Y} \mid \boldsymbol{T}) = \boldsymbol{\Psi}_{(m+1)}\text{vec}(\boldsymbol{T})$, where $\boldsymbol{T} = [\![\boldsymbol{X}; \boldsymbol{W}_1^\intercal, \ldots, \boldsymbol{W}_m^\intercal]\!]$. Therefore,

$$\begin{aligned}
\boldsymbol{B}_{(m+1)}\text{vec}(\boldsymbol{X}) &= \boldsymbol{\Psi}_{(m+1)}\text{vec}(\boldsymbol{T}) \\
&= \boldsymbol{\Psi}_{(m+1)}\text{vec}([\![\boldsymbol{X}; \boldsymbol{W}_1^\intercal, \ldots, \boldsymbol{W}_m^\intercal]\!]) \\
&= \boldsymbol{\Psi}_{(m+1)} \left(\boldsymbol{W}_m^\intercal \otimes \cdots \otimes \boldsymbol{W}_1^\intercal\right) \cdot \text{vec}(\boldsymbol{X}) \\
&= [\![\boldsymbol{\Psi}; \boldsymbol{W}_1, \ldots, \boldsymbol{W}_m, \boldsymbol{I}_r]\!]_{(m+1)} \cdot \text{vec}(\boldsymbol{X}).
\end{aligned}$$

This implies that $\boldsymbol{B} = [\![\boldsymbol{\Psi}; \boldsymbol{W}_1, \ldots, \boldsymbol{W}_m, \boldsymbol{I}_r]\!]$ under the PLS regression assumption, and hence $\widehat{\boldsymbol{B}}_{\text{PLS}} = [\![\widehat{\boldsymbol{\Psi}}; \widehat{\boldsymbol{W}}_1, \ldots, \widehat{\boldsymbol{W}}_m, \boldsymbol{I}_r]\!]$. The rest of the proof follows from the fact that $\text{cov}\{\text{vec}(\boldsymbol{T})\} = (\boldsymbol{W}_m^\intercal \boldsymbol{\Sigma}_m \boldsymbol{W}_m) \otimes \cdots \otimes (\boldsymbol{W}_1^\intercal \boldsymbol{\Sigma}_1 \boldsymbol{W}_1)$ and then applying Lemma 1 on the regression of $\boldsymbol{Y}$ on $\boldsymbol{T}$ and □

Proof of Proposition 1

Proof. From Lemma 2, we have

$$\begin{aligned}
\widehat{\boldsymbol{B}}_{\text{PLS}} &= [\![\widehat{\boldsymbol{\Psi}}; \widehat{\boldsymbol{W}}_1, \ldots, \widehat{\boldsymbol{W}}_m, \boldsymbol{I}_r]\!] \\
&= [\![[\![\widehat{\boldsymbol{C}}_{\boldsymbol{T}}; (\widehat{\boldsymbol{W}}_1^\intercal \widehat{\boldsymbol{\Sigma}}_1 \widehat{\boldsymbol{W}}_1)^{-1}, \ldots, (\widehat{\boldsymbol{W}}_m^\intercal \widehat{\boldsymbol{\Sigma}}_m \widehat{\boldsymbol{W}}_m)^{-1}, \boldsymbol{I}_r]\!]; \widehat{\boldsymbol{W}}_1, \ldots, \widehat{\boldsymbol{W}}_m, \boldsymbol{I}_r]\!] \\
&= [\![\widehat{\boldsymbol{C}}_{\boldsymbol{T}}; \widehat{\boldsymbol{W}}_1 (\widehat{\boldsymbol{W}}_1^\intercal \widehat{\boldsymbol{\Sigma}}_1 \widehat{\boldsymbol{W}}_1)^{-1}, \ldots, \widehat{\boldsymbol{W}}_m (\widehat{\boldsymbol{W}}_m^\intercal \widehat{\boldsymbol{\Sigma}}_m \widehat{\boldsymbol{W}}_m)^{-1}, \boldsymbol{I}_r]\!] \\
&= [\![\widehat{\boldsymbol{C}}; \widehat{\boldsymbol{W}}_1 (\widehat{\boldsymbol{W}}_1^\intercal \widehat{\boldsymbol{\Sigma}}_1 \widehat{\boldsymbol{W}}_1)^{-1} \widehat{\boldsymbol{W}}_1^\intercal, \ldots, \widehat{\boldsymbol{W}}_m (\widehat{\boldsymbol{W}}_m^\intercal \widehat{\boldsymbol{\Sigma}}_m \widehat{\boldsymbol{W}}_m)^{-1} \widehat{\boldsymbol{W}}_m^\intercal, \boldsymbol{I}_r]\!],
\end{aligned}$$

where the last equality follows from $\widehat{\boldsymbol{C}}_{\boldsymbol{T}} = [\![\widehat{\boldsymbol{C}}; \widehat{\boldsymbol{W}}_1^\intercal, \ldots, \widehat{\boldsymbol{W}}_m^\intercal, \boldsymbol{I}_r]\!]$. The conclusion then follows from $\widehat{\boldsymbol{W}}_1 (\widehat{\boldsymbol{W}}_1^\intercal \widehat{\boldsymbol{\Sigma}}_1 \widehat{\boldsymbol{W}}_1)^{-1} \widehat{\boldsymbol{W}}_1^\intercal = \boldsymbol{P}_{\widehat{\boldsymbol{W}}_k(\widehat{\boldsymbol{\Sigma}}_k)} \widehat{\boldsymbol{\Sigma}}_k^{-1}$ and $\widehat{\boldsymbol{B}}_{\text{OLS}} = [\![\widehat{\boldsymbol{C}}; \widehat{\boldsymbol{\Sigma}}_1^{-1}, \ldots, \widehat{\boldsymbol{\Sigma}}_m^{-1}, \boldsymbol{I}_r]\!]$. □

Proof of Proposition 2

Proof. First, $\boldsymbol{X} \times_k \boldsymbol{Q}_k \perp\!\!\!\perp \boldsymbol{X} \times_k \boldsymbol{P}_k$ implies $0 = \text{cov}\{\text{vec}(\boldsymbol{X} \times_k \boldsymbol{Q}_k), \text{vec}(\boldsymbol{X} \times_k \boldsymbol{P}_k)\} = \boldsymbol{\Sigma}_m \otimes \cdots \otimes \boldsymbol{Q}_k \boldsymbol{\Sigma}_k \boldsymbol{P}_k \otimes \cdots \otimes \boldsymbol{\Sigma}_1$. It holds if and only if $\boldsymbol{Q}_k \boldsymbol{\Sigma}_k \boldsymbol{P}_k = 0$. By the definition of a reducing subspace (Cook et al., 2010), $\boldsymbol{Q}_k \boldsymbol{\Sigma}_k \boldsymbol{P}_k = 0$ if and only if $\text{span}(\boldsymbol{P}_k) = \text{span}(\boldsymbol{\Gamma}_k)$ is a reducing subspace of $\boldsymbol{\Sigma}_k$, i.e., $\boldsymbol{\Sigma}_k = \boldsymbol{\Gamma}_k \boldsymbol{\Omega}_k \boldsymbol{\Gamma}_k^\intercal + \boldsymbol{\Gamma}_{0k} \boldsymbol{\Omega}_{0k} \boldsymbol{\Gamma}_{0k}^\intercal$.

Second, we substitute $\boldsymbol{X} = \boldsymbol{X} \times_k \boldsymbol{P}_k + \boldsymbol{X} \times_k \boldsymbol{Q}_k$ into (3) and get

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{B}_{(m+1)} \text{vec}(\boldsymbol{X} \times_k \boldsymbol{P}_k) + \boldsymbol{B}_{(m+1)} \text{vec}(\boldsymbol{X} \times_k \boldsymbol{Q}_k) + \boldsymbol{\varepsilon} \\
&= (\boldsymbol{B} \times_k \boldsymbol{P}_k)_{(m+1)} \text{vec}(\boldsymbol{X} \times_k \boldsymbol{P}_k) + (\boldsymbol{B} \times_k \boldsymbol{Q}_k)_{(m+1)} \text{vec}(\boldsymbol{X} \times_k \boldsymbol{Q}_k) + \boldsymbol{\varepsilon}.
\end{aligned}$$

Therefore, $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{X} \times_k \boldsymbol{Q}_k \mid \boldsymbol{X} \times_k \boldsymbol{P}_k$ implies $\boldsymbol{B} \times_k \boldsymbol{Q}_k = 0$, which is equivalent to $\boldsymbol{B} = \boldsymbol{B} \times_k \boldsymbol{P}_k = \boldsymbol{B} \times_k \boldsymbol{\Gamma}_k \boldsymbol{\Gamma}_k^\intercal$, which gives the parametrization $\boldsymbol{B} = [\![\boldsymbol{\Theta}; \boldsymbol{\Gamma}_1, \cdots, \boldsymbol{\Gamma}_m, \boldsymbol{I}_r]\!]$. □

Proof of Proposition 3

Proof. We first show $\mathcal{T}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{B}) \subseteq \mathcal{E}_{\boldsymbol{\Sigma}_m}\left(\boldsymbol{B}_{(m)}\right) \otimes \cdots \otimes \mathcal{E}_{\boldsymbol{\Sigma}_1}\left(\boldsymbol{B}_{(1)}\right)$. From Definition 2, this means we need to show that (a) $\mathcal{E}_{\boldsymbol{\Sigma}_m}\left(\boldsymbol{B}_{(m)}\right) \otimes \cdots \otimes \mathcal{E}_{\boldsymbol{\Sigma}_1}\left(\boldsymbol{B}_{(1)}\right)$ is a reducing subspace of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$, and (b) it contains $\text{span}(\boldsymbol{B}_{(m+1)})$. It is straightforward to see (a) from $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Sigma}_m \otimes \cdots \otimes \boldsymbol{\Sigma}_1$ and that $\mathcal{E}_{\boldsymbol{\Sigma}_k}\left(\boldsymbol{B}_{(k)}\right)$ is a reducing subspace of $\boldsymbol{\Sigma}_k$ for each k. To show (b), we can write $\boldsymbol{B}_{(k)} = \boldsymbol{P}_k \boldsymbol{B}_{(k)}$, where $\boldsymbol{P}_k$ is the projection on to $\mathcal{E}_{\boldsymbol{\Sigma}_k}\left(\boldsymbol{B}_{(k)}\right)$, because

$\text{span}(\boldsymbol{B}_{(k)}) \subseteq \mathcal{E}_{\boldsymbol{\Sigma}_k}\left(\boldsymbol{B}_{(k)}\right)$ by definition. This further implies that $\boldsymbol{B} = \boldsymbol{B} \times_k \boldsymbol{P}_k$, $k = 1, \ldots, m$, and that $\boldsymbol{B} = \boldsymbol{B} \times_1 \boldsymbol{P}_1 \times_2 \cdots \times_m \boldsymbol{P}_m$. Taking mode-$(m+1)$ matricization on both sides of the last equation, we have (b).

We next show $\mathcal{T}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{B}) \supseteq \mathcal{E}_{\boldsymbol{\Sigma}_m}\left(\boldsymbol{B}_{(m)}\right) \otimes \cdots \otimes \mathcal{E}_{\boldsymbol{\Sigma}_1}\left(\boldsymbol{B}_{(1)}\right)$. By definition, we can write $\mathcal{T}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{B}) = \mathcal{E}_m \otimes \cdots \otimes \mathcal{E}_1$ for some $\mathcal{E}_k \subseteq \mathbb{R}^{p_k}$, $k = 1, \ldots, m$. It remains to show that $\mathcal{E}_k \supseteq \mathcal{E}_{\boldsymbol{\Sigma}_k}\left(\boldsymbol{B}_{(k)}\right)$. We achieve this by showing (c) $\mathcal{E}_k$ is a reducing subspace of $\boldsymbol{\Sigma}_k$, (d) $\mathcal{E}_k$ contains $\text{span}(\boldsymbol{B}_{(k)})$, and then by noticing $\mathcal{E}_{\boldsymbol{\Sigma}_k}\left(\boldsymbol{B}_{(k)}\right)$ is the smallest subspace satisfying (c) and (d). Note that (d) can be directly obtained from Proposition 2. To get (c), we recall that $\mathcal{T}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\boldsymbol{B})$ is a reducing subspace of $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Sigma}_m \otimes \cdots \otimes \boldsymbol{\Sigma}_1$ and thus

$$\boldsymbol{P}_{\mathcal{E}_m \otimes \cdots \otimes \mathcal{E}_1}(\boldsymbol{\Sigma}_m \otimes \cdots \otimes \boldsymbol{\Sigma}_1)\boldsymbol{Q}_{\mathcal{E}_m \otimes \cdots \otimes \mathcal{E}_1} = 0,$$

where $\boldsymbol{P}_{\mathcal{E}_m \otimes \cdots \otimes \mathcal{E}_1} = \boldsymbol{P}_{\mathcal{E}_m} \otimes \cdots \otimes \boldsymbol{P}_{\mathcal{E}_1}$ and $\boldsymbol{Q}_{\mathcal{E}_m \otimes \cdots \otimes \mathcal{E}_1} = \boldsymbol{I} - \boldsymbol{P}_{\mathcal{E}_m \otimes \cdots \otimes \mathcal{E}_1}$. Expand the above equation we get,

$$\boldsymbol{P}_{\mathcal{E}_m}\boldsymbol{\Sigma}_m \otimes \cdots \otimes \boldsymbol{P}_{\mathcal{E}_1}\boldsymbol{\Sigma}_1 - \boldsymbol{P}_{\mathcal{E}_m}\boldsymbol{\Sigma}_m\boldsymbol{P}_{\mathcal{E}_m} \otimes \cdots \otimes \boldsymbol{P}_{\mathcal{E}_1}\boldsymbol{\Sigma}_1\boldsymbol{P}_{\mathcal{E}_1} = 0,$$

which implies the following equality by right-multiplying $\boldsymbol{P}_{\mathcal{E}_m} \otimes \boldsymbol{P}_{\mathcal{E}_2} \otimes \boldsymbol{I}_{p_1}$:

$$\boldsymbol{P}_{\mathcal{E}_m}\boldsymbol{\Sigma}_m\boldsymbol{P}_{\mathcal{E}_m} \otimes \cdots \otimes \boldsymbol{P}_{\mathcal{E}_2}\boldsymbol{\Sigma}_2\boldsymbol{Q}_{\mathcal{E}_2} \otimes \boldsymbol{P}_{\mathcal{E}_1}\boldsymbol{\Sigma}_1\boldsymbol{Q}_{\mathcal{E}_1} = 0,$$

which implies $\mathcal{E}_1$ reduces $\boldsymbol{\Sigma}_1$. Similarly we get $\mathcal{E}_k$ reduces $\boldsymbol{\Sigma}_k$ for all $k = 1, \ldots, m$. This completes the proof. □

Proof of Lemma 3

Proof. From Algorithm 4, we know that $\boldsymbol{w}_{sk}$ is the dominant eigenvector of $\tilde{\boldsymbol{C}}_{(s-1)k}\tilde{\boldsymbol{C}}_{(s-1)k}^{\intercal}$, which equals to $\boldsymbol{Q}_{(s-1)k}\tilde{\boldsymbol{C}}_{0k}\tilde{\boldsymbol{C}}_{0k}^{\intercal}\boldsymbol{Q}_{(s-1)k}$. The conclusion follows from noticing $\tilde{\boldsymbol{C}}_k = \tilde{\boldsymbol{C}}_{0k}\tilde{\boldsymbol{C}}_{0k}^{\intercal}$. □

Proof of Theorem 1

Proof. From Lemma 3, we see that $\boldsymbol{w}_{1k}$ is the first eigenvector of $\tilde{\boldsymbol{C}}_k$. Then for $s > 1$, we have $\boldsymbol{w}_{sk}$ is the first eigenvector of $\boldsymbol{Q}_{(s-1)k}\tilde{\boldsymbol{C}}_k\boldsymbol{Q}_{(s-1)k}$ where $\boldsymbol{Q}_{(s-1)k}$ is the projection

onto the orthogonal complement of $\text{span}\big(\boldsymbol{\Sigma}_k(\boldsymbol{w}_{1k}, \ldots, \boldsymbol{w}_{(s-1)k})\big)$. Following the proof of Proposition 4.1 in Cook et al. (2013), we have

$$\mathcal{W}_k^{(0)} \subset \ldots \subset \mathcal{W}_k^{(u_k)} = \mathcal{E}_{\boldsymbol{\Sigma}_k}(\tilde{\boldsymbol{C}}_k) = \mathcal{W}_k^{(u_k+1)} = \ldots = \mathcal{W}_k^{(p_k)},$$

where $\mathcal{E}_{\boldsymbol{\Sigma}_k}(\tilde{\boldsymbol{C}}_k)$ is the $\boldsymbol{\Sigma}_k$-envelope of $\text{span}(\tilde{\boldsymbol{C}}_k)$. We next need to show $\mathcal{E}_{\boldsymbol{\Sigma}_k}\big(\boldsymbol{B}_{(k)}\big) = \mathcal{E}_{\boldsymbol{\Sigma}_k}(\tilde{\boldsymbol{C}}_k)$. From Lemma 3, we see that $\text{span}(\tilde{\boldsymbol{C}}_k) = \text{span}(\boldsymbol{C}_{(k)})$. From Lemma 1, we see that $\text{span}(\boldsymbol{B}_{(k)}) = \text{span}(\boldsymbol{\Sigma}_k^{-1}\boldsymbol{C}_{(k)})$. Finally, from Proposition 2.4 of Cook et al. (2010), we have $\mathcal{E}_{\boldsymbol{\Sigma}_k}\big(\boldsymbol{B}_{(k)}\big) = \mathcal{E}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\Sigma}_k^{-1}\boldsymbol{B}_{(k)})$ and thus $\mathcal{E}_{\boldsymbol{\Sigma}_k}\big(\boldsymbol{B}_{(k)}\big) = \mathcal{E}_{\boldsymbol{\Sigma}_k}(\boldsymbol{C}_{(k)}) = \mathcal{E}_{\boldsymbol{\Sigma}_k}(\tilde{\boldsymbol{C}}_k)$. □

Proof of Theorem 2

Proof. From Theorem 1, if d_k is chosen as u_k, the population value $\text{span}(\boldsymbol{W}_k) = \mathcal{W}_k^{(u_k)} = \mathcal{E}_{\boldsymbol{\Sigma}_k}\big(\boldsymbol{B}_{(k)}\big)$. Since the sample version of Algorithm 4 is based on eigen-decomposition and $\sqrt{n}$-consistent sample covariance matrices $\widehat{\boldsymbol{C}}$, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}$, and $\widehat{\boldsymbol{\Sigma}}_k$, $k = 1, \ldots, m$, it is clear that $\widehat{\boldsymbol{W}}_k$ is $\sqrt{n}$-consistent for $\boldsymbol{W}_k$. Hence $\widehat{\boldsymbol{W}}_k$ is $\sqrt{n}$-consistent for the envelope $\mathcal{E}_{\boldsymbol{\Sigma}_k}\big(\boldsymbol{B}_{(k)}\big)$.

For $d_k \geq u_k$, we have $\mathcal{W}_k^{(d_k)} = \mathcal{W}_k^{(u_k)} = \mathcal{E}_{\boldsymbol{\Sigma}_k}\big(\boldsymbol{B}_{(k)}\big)$, therefore the projection $\boldsymbol{P}_{\widehat{\boldsymbol{W}}_k(\widehat{\boldsymbol{\Sigma}}_k)}$ is $\sqrt{n}$-consistent for $\boldsymbol{P}_{\boldsymbol{W}_k(\boldsymbol{\Sigma}_k)}$. Since $\boldsymbol{W}_k$ is a semi-orthogonal basis for the envelope $\mathcal{E}_{\boldsymbol{\Sigma}_k}\big(\boldsymbol{B}_{(k)}\big)$, it is a reducing subspace of $\boldsymbol{\Sigma}_k$ and hence $\boldsymbol{P}_{\boldsymbol{W}_k(\boldsymbol{\Sigma}_k)} = \boldsymbol{P}_{\boldsymbol{W}_k}$ by definition. Therefore, $\boldsymbol{P}_{\widehat{\boldsymbol{W}}_k(\widehat{\boldsymbol{\Sigma}}_k)}$ is $\sqrt{n}$-consistent estimator for the projection onto the envelope $\mathcal{E}_{\boldsymbol{\Sigma}_k}\big(\boldsymbol{B}_{(k)}\big)$. Then from Proposition 1, we recall that $\widehat{\boldsymbol{B}}_{\text{PLS}} = \widehat{\boldsymbol{B}}_{\text{OLS}} \times_1 \boldsymbol{P}_{\widehat{\boldsymbol{W}}_1(\widehat{\boldsymbol{\Sigma}}_1)} \times_2 \cdots \times_m \boldsymbol{P}_{\widehat{\boldsymbol{W}}_m(\widehat{\boldsymbol{\Sigma}}_m)}$. The $\sqrt{n}$-consistency of $\widehat{\boldsymbol{B}}_{\text{PLS}}$ then follows from the $\sqrt{n}$-consistency of $\widehat{\boldsymbol{B}}_{\text{OLS}}$ and $\boldsymbol{P}_{\widehat{\boldsymbol{W}}_k(\widehat{\boldsymbol{\Sigma}}_k)}$. □

2 Additional simulations

We consider an additional simulation example, with a univariate response and a 3-way tensor predictor with higher dimension and rank. Specifically, the simulation setup is similar to that in Section 5.2, with two main changes. The first is that the response is now a scalar, which allows a more direct comparison with the CP method of Zhou et al. (2013) that was designed for a univariate response. The second is that we now consider both the original setup of a predictor tensor $\boldsymbol{X} \in \mathbb{R}^{20\times 20\times 20}$ with a core tensor $\boldsymbol{\Theta} \in \mathbb{R}^{2\times 2\times 2}$ $(p = 20, u = 2)$, and a new setup of a predictor tensor $\boldsymbol{X} \in \mathbb{R}^{40\times 40\times 40}$ with

(p, u)	Model	Prediction			
		OLS	CP	TEPLS	TEPLS-CV
$(20, 2)$	I	84.4 (1.3)	7.2 (0.2)	1.4 (0.1)	1.4 (0.1)
	II	42.0 (0.6)	5.6 (0.1)	1.0 (0.1)	1.0 (0.1)
	III	41.0 (0.5)	5.4 (0.1)	1.0 (0.1)	1.0 (0.1)
$(40, 5)$	I	1504.9 (21.4)	17.7 (0.3)	4.0 (0.1)	4.4 (0.1)
	II	367.5 (4.9)	4.8 (0.1)	1.2 (0.1)	1.2 (0.1)
	III	361.9 (5.3)	4.5 (0.1)	1.0 (0.1)	1.0 (0.1)
(p, u)	Model	Estimation			
		OLS	CP	TEPLS	TEPLS-CV
$(20, 2)$	I	$> 10^4$	$> 10^3$	1.1 (0.1)	1.1 (0.1)
	II	60.9 (0.3)	20.4 (0.2)	1.7 (0.1)	1.7 (0.1)
	III	$> 10^6$	$> 10^5$	1.7 (0.1)	1.7 (0.1)
$(40, 5)$	I	$> 10^5$	$> 10^4$	5.8 (0.1)	6.2 (0.1)
	II	305.9 (1.5)	30.9 (0.3)	6.4 (0.1)	6.4 (0.1)
	III	$> 10^7$	$> 10^5$	6.3 (0.1)	6.3 (0.1)

Table S1: Univariate response and 3-way predictor. Performance under various scenarios and comparison of estimators. OLS, CP, tensor envelope PLS with true and estimated envelope dimensions. Reported are the average and standard error (in parenthesis) of the prediction mean squared error evaluated on an independent testing data, and the estimation error, all based on 100 data replications.

a core tensor $\mathbf{\Theta} \in \mathbb{R}^{5\times5\times5}$ $(p = 40, u = 5)$. The latter has a higher predictor dimension and rank, and is comparable to the dimension of the ADHD real data example in Section 6.2. The sample size for training and testing data is still fixed at $n = 200$. Table S1 summarizes the prediction and estimation results based on 100 data replications. It is again clearly seen that the proposed tensor envelope PLS method is more competitive than the alternative solutions in terms of both prediction and estimation accuracy across all model scenarios.

References

Cook, R. D., Helland, I. S., and Su, Z. (2013). Envelopes and partial least squares regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(5):851–877.

Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statist. Sinica*, 20(3):927–960.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552.