Supplement to "Envelope model for function-on-function linear regression"

Zhihua Su
Department of Statistics, University of Florida
and
Bing Li
Department of Statistics, Pennsylvania State University
and
Dennis Cook
School of Statistics, University of Minnesota

December 26, 2022

The supplementary materials contain proof of theorems, lemmas and propositions, details on estimation and asymptotics, as well as additional simulations.

1 Proof of Theorem 1

Proof. By the expansions in (7), (9) and (10), we have

$$\begin{aligned} Q_{\mathcal{E}(B;\Sigma_\epsilon)}(Y-\alpha) &= Q_{\mathcal{E}(B;\Sigma_\epsilon)}(BX+\epsilon) \\ &= \Big(\sum_{i\notin I} \chi_i \otimes \chi_i\Big)\Big[\sum_{i\in I}\sum_{j\in J} b_{ij}(\chi_i\otimes\psi_j)X + \sum_{i=1}^{\infty}\rho_i^{1/2}\,\nu_i\,\chi_i\Big] \\ &= \sum_{i\notin I}\rho_i^{1/2}\,\nu_i\,\chi_i, \end{aligned}$$

where, for the third equality, we have used the fact $\langle \chi_i, \chi_j\rangle_{\mathcal{H}_Y} = 0$ for any $i \in I$, $j \notin I$. Hence $Q_{\mathcal{E}(B;\Sigma_\epsilon)}(Y-\alpha) \perp\!\!\!\perp X$, which is equivalent to relation (11a). Next,

$$\begin{aligned} P_{\mathcal{E}(B;\Sigma_\epsilon)}(Y-\alpha) &= P_{\mathcal{E}(B;\Sigma_\epsilon)}(BX+\epsilon) \\ &= \Big(\sum_{i\in I} \chi_i \otimes \chi_i\Big)\Big[\sum_{i\in I}\sum_{j\in J} b_{ij}(\chi_i\otimes\psi_j)X + \sum_{i=1}^{\infty}\rho_i^{1/2}\,\nu_i\,\chi_i\Big] \\ &= \sum_{i\in I}\sum_{j\in J} b_{ij}(\chi_i\otimes\psi_j)X + \sum_{i\in I}\rho_i^{1/2}\,\nu_i\,\chi_i. \end{aligned}$$

Since elements in the two sets $\{\nu_i : i \in I\}$ and $\{\nu_i : i \notin I\}$ are independent of each other, we have $P_{\mathcal{E}(B;\Sigma_\epsilon)}(Y-\alpha) \perp\!\!\!\perp Q_{\mathcal{E}(B;\Sigma_\epsilon)}(Y-\alpha) \mid X$, which is equivalent to (b).

Similarly, since

$$\begin{aligned} P_{\mathcal{E}(B^*;\Sigma_X)}X &= \Big(\sum_{j\in J}\psi_j\otimes\psi_j\Big)\Big(\sum_{j=1}^{\infty}\langle X,\psi_i\rangle_{\mathcal{H}_X}\,\psi_i\Big) = \sum_{j\in J}\tau_i^{1/2}\,\xi_i\,\psi_i, \\ Q_{\mathcal{E}(B^*;\Sigma_X)}X &= \Big(\sum_{j\notin J}\psi_j\otimes\psi_j\Big)\Big(\sum_{j=1}^{\infty}\langle X,\psi_i\rangle_{\mathcal{H}_X}\,\psi_i\Big) = \sum_{j\notin J}\tau_i^{1/2}\,\xi_i\,\psi_i, \end{aligned}$$

we have $P_{\mathcal{E}(B^*;\Sigma_X)}\, X \perp\!\!\!\perp Q_{\mathcal{E}(B^*;\Sigma_X)}\, X$, proving (11c). Finally, because

$$\begin{aligned} Y &= \alpha + BX + \epsilon = \alpha + B\big(P_{\mathcal{E}(B^*;\Sigma_X)}\, X + Q_{\mathcal{E}(B^*;\Sigma_X)}\, X\big) + \epsilon \\ &= \alpha + B\, P_{\mathcal{E}(B^*;\Sigma_X)}\, X + \epsilon, \end{aligned}$$

and because $X \perp\!\!\!\perp \epsilon$, we have relation (d). $\square$

2 Proof of Theorem 2

PROOF. It suffices to show that $\mathrm{Lat}_B(\Sigma_\epsilon) = \mathrm{Lat}_B(\Sigma_Y)$. Since $Y = \alpha + BX + \epsilon$, we have $\Sigma_Y = B\Sigma_X B^* + \Sigma_\epsilon$. Let $\mathcal{S} \in \mathrm{Lat}_B(\Sigma_\epsilon)$ and $v \in \mathcal{S}$. Then $\Sigma_Y v = B\Sigma_X B^* v + \Sigma_\epsilon v$. Since $\mathcal{S} \in \mathrm{Lat}_B(\Sigma_\epsilon)$, we have $\Sigma_\epsilon v \in \mathcal{S}$. Since $B\Sigma_X B^* v \in \mathrm{ran}(B)$ and $\mathcal{S}$ contains $\mathrm{ran}(B)$, we have $B\Sigma_X B^* v \in \mathcal{S}$. Hence $\Sigma_Y \mathcal{S} \subseteq \mathcal{S}$, which implies $\mathcal{S} \in \mathrm{Lat}_B(\Sigma_Y)$.

Conversely, let $\mathcal{S} \in \mathrm{Lat}_B(\Sigma_Y)$ and $v \in \mathcal{S}$. Then $\Sigma_\epsilon v = \Sigma_Y v - B\Sigma_X B^* v$. Similar to the argument in the last paragraph, we have $\Sigma_Y v \in \mathcal{S}$, $B\Sigma_X B^* v \in \mathcal{S}$, which implies $\mathcal{S} \in \mathrm{Lat}_B(\Sigma_\epsilon)$. $\square$

3 Proof of Theorem 3

PROOF. Since $Y = \alpha + BX + \epsilon$, we have

$$\begin{aligned} Y - \alpha &= \sum_{i \in I} \sum_{j \in J} b_{ij} (\chi_i \otimes \psi_j) \sum_{k=1}^{\infty} \tau_k^{1/2} \xi_k \psi_k + \sum_{i=1}^{\infty} \rho_i^{1/2} \nu_i \chi_i \\ &= \sum_{i \in I} \left(\sum_{j \in J} b_{ij} \tau_j^{1/2} \xi_j + \rho_i^{1/2} \nu_i \right) \chi_i + \sum_{i \notin I} \rho_i^{1/2} \nu_i \chi_i \\ &= \sum_{i \in I} U_i \chi_i + \sum_{i \notin I} \rho_i^{1/2} \nu_i \chi_i. \end{aligned}$$

Since

$$\sum_{i\in I} U_i\chi_i = U^\intercal\chi(I) = U^\intercal Q D^{-1/2} D^{1/2} Q^\intercal \chi(I) = \zeta(I)^\intercal D^{1/2}\phi(I),$$

we have

$$Y = \alpha + \zeta(I)^\intercal \mathrm{diagvm}\big(\lambda(I)\big)^{1/2}\phi(I) + \sum_{i\notin I} \rho_i^{1/2}\nu_i\chi_i.$$

The claim follows by updating the notation as indicated in the theorem. For example, for $i \notin I$, $\chi_i = \phi_i$, $\nu_i = \zeta_i$ and $\rho_i = \lambda_i$. □

4 Proof of Theorem 4

We first introduce two lemmas that will be needed to prove Theorem 4. Lemma 1 was proved independently by Cook, Forzani and Liu (Cook et al., 2020, Prop. 3.3) for envelopes in model (1) with linearly constrained coefficients. The proof here, which uses our novel lattice context, is different from that of Cook et al.

Lemma 1. *Suppose $\Sigma \in \mathbb{R}^{r\times r}$ is a symmetric and positive definite matrix, and $M \in \mathbb{R}^{t\times r}$ is a matrix with $t > r$ and $M^\intercal M = I_r$. Then, for any matrix $A \in \mathbb{R}^{r\times s}$, we have*

$$\mathcal{E}(MA; M\Sigma M^\intercal) = M\mathcal{E}(A;\Sigma).$$

PROOF. We first show that the asserted equality is implied by

$$\mathcal{E}(MA; M\Sigma M^\intercal) \subseteq M\mathcal{E}(A;\Sigma) \subseteq MM^\intercal\mathcal{E}(MA, M\Sigma M^\intercal). \tag{1}$$

In fact, if this holds, then $\mathcal{E}(MA, M\Sigma M^\intercal) \subseteq MM^\intercal\mathcal{E}(MA, M\Sigma M^\intercal)$, but we know $\mathcal{E}(MA, M\Sigma M^\intercal)$ cannot be a proper subset of $MM^\intercal\mathcal{E}(MA, M\Sigma M^\intercal)$ because the dimension of the latter is no greater the dimension of the former. Hence the three spaces in (1) must be the same.

We now prove the first inclusion in (1), for which it suffices to show that $M\mathcal{E}(A;\Sigma) \in \mathrm{Lat}_{MA}(\Sigma)$. Because $\mathcal{E}(A;\Sigma)$ is an invariant subspace of Σ, we have $\Sigma\mathcal{E}(A;\Sigma) \subseteq \mathcal{E}(A;\Sigma)$.

Hence

$$M\Sigma M^{\intercal} M\mathcal{E}(A;\Sigma) = M\Sigma\mathcal{E}(A;\Sigma) \subseteq M\mathcal{E}(A;\Sigma),$$

which means $M\mathcal{E}(A;\Sigma)$ is a reducing subspace of $M\Sigma M^{\intercal}$. Since $\mathrm{span}(A) \subseteq \mathcal{E}(A;\Sigma)$, we have $\mathrm{span}(MA) \subseteq M\mathcal{E}(A;\Sigma)$. Therefore, $M\mathcal{E}(A;\Sigma) \in \mathrm{Lat}_{MA}(\Sigma)$.

To show the second inclusion in (2), it suffices to show

$$\mathcal{E}(A;\Sigma) \subseteq M^{\intercal}\mathcal{E}(MA, M\Sigma M^{\intercal}),$$

which is implied by $M^{\intercal}\mathcal{E}(MA, M\Sigma M^{\intercal}) \in \mathrm{Lat}_{A}(\Sigma)$. Since

$$\begin{aligned} \Sigma M^{\intercal}\mathcal{E}(MA, M\Sigma M^{\intercal}) &= M^{\intercal} M\Sigma M^{\intercal}\mathcal{E}(MA, M\Sigma M^{\intercal}) \\ &\subseteq M^{\intercal}\mathcal{E}(MA, M\Sigma M^{\intercal}), \end{aligned}$$

$M^{\intercal}\mathcal{E}(MA, M\Sigma M^{\intercal})$ is a reducing subspace of Σ. Since $\mathcal{E}(MA; M\Sigma M^{\intercal})$ contains $\mathrm{span}(MA)$, $M^{\intercal}\mathcal{E}(MA; M\Sigma M^{\intercal})$ must contain $\mathrm{span}(A)$. Therefore

$$M^{\intercal}\mathcal{E}(MA; M\Sigma M^{\intercal}) \in \mathrm{Lat}_{A}(\Sigma),$$

as desired. □

Lemma 2. *If the conditions in Lemma 1 are satisfied and $\Lambda \in \mathbb{R}^{r\times r}$ is a symmetric and positive semidefinite matrix such that $M\Lambda = \Lambda M = 0$, then*

$$\mathcal{E}(MA; M\Sigma M^{\intercal} + \Lambda) = \mathcal{E}(MA; M\Sigma M^{\intercal}).$$

PROOF. It suffices to show that

$$\begin{aligned} &\mathcal{E}(MA; M\Sigma M^{\intercal}) \in \mathrm{Lat}_{MA}(M\Sigma M^{\intercal} + \Lambda), \\ &\mathcal{E}(MA; M\Sigma M^{\intercal} + \Lambda) \in \mathrm{Lat}_{MA}(M\Sigma M^{\intercal}). \end{aligned} \tag{2}$$

To show the first relation, note that

$$\begin{aligned}&(M\Sigma M^\intercal + \Lambda)\mathcal{E}(MA; M\Sigma M^\intercal)\\ &= M\Sigma M^\intercal \mathcal{E}(MA; M\Sigma M^\intercal) + \Lambda\mathcal{E}(MA; M\Sigma M^\intercal).\end{aligned}$$

Let us show that the second term on the right is the space $\{0\}$. To do so, it suffices to show that $\mathcal{E}(MA; M\Sigma M^\intercal) \subseteq \mathrm{span}(M)$, which, in turn, is implied by $\mathrm{span}(M) \in \mathrm{Lat}_{MA}(M\Sigma M^\intercal)$. This is obviously true:

$$M\Sigma M^\intercal \mathrm{span}(M) \subseteq \mathrm{span}(M); \quad \mathrm{span}(MA) \subseteq \mathrm{span}(M).$$

Hence

$$\begin{aligned}(M\Sigma M^\intercal + \Lambda)\mathcal{E}(MA; M\Sigma M^\intercal) &= M\Sigma M^\intercal \mathcal{E}(MA; M\Sigma M^\intercal)\\ &\subseteq \mathcal{E}(MA; M\Sigma M^\intercal).\end{aligned}$$

Meanwhile, it is also true that $\mathrm{span}(MA) \subseteq \mathcal{E}(MA; M\Sigma M^\intercal)$. This proves the first relation (2).

To show the second relation in (2), note that

$$\begin{aligned}&M\Sigma M^\intercal \mathcal{E}(MA; M\Sigma M^\intercal + \Lambda)\\ &= (M\Sigma M^\intercal + \Lambda)\mathcal{E}(MA; M\Sigma M^\intercal + \Lambda) - \Lambda\mathcal{E}(MA; M\Sigma M^\intercal + \Lambda)\end{aligned}$$

Let us show the second term is the space $\{0\}$. Again, it suffices to show that $\mathrm{span}(M) \in \mathrm{Lat}_{MA}(M\Sigma M^\intercal + \Lambda)$. This is true because

$$\begin{aligned}&(M\Sigma M^\intercal + \Lambda)\mathrm{span}(M) = M\Sigma M^\intercal \mathrm{span}(M) \subseteq \mathrm{span}(M) \quad \text{and}\\ &\mathrm{span}(MA) \subseteq \mathrm{span}(M).\end{aligned}$$

It follows that

$$\begin{aligned} M\Sigma M^\intercal \mathcal{E}(MA; M\Sigma M^\intercal + \Lambda) &= (M\Sigma M^\intercal + \Lambda)\mathcal{E}(MA; M\Sigma M^\intercal + \Lambda) \\ &\subseteq \mathcal{E}(MA; M\Sigma M^\intercal + \Lambda). \end{aligned}$$

Again, it is obvious that $\mathcal{E}(MA; M\Sigma M^\intercal + \Lambda)$ contains $\mathrm{span}(MA)$. Hence the second relation in (2) holds. □

Now we prove Theorem 4. PROOF. Taking the inner product with c on both sides of (5), we have

$$\langle c, Y\rangle_{\mathcal{H}_Y} = \mu + \langle c, BX\rangle_{\mathcal{H}_Y} + \langle c, \epsilon\rangle_{\mathcal{H}_Y}. \tag{3}$$

Since

$$\begin{aligned} BX =&[BX]^\intercal_{\phi(I)}\phi(I) \\ =&[X]^\intercal_{\psi(J)}({}_{\phi(I)}[B]_{\psi(J)})^\intercal\phi(I) \\ =&[X]^\intercal_{\psi(J)}({}_{\phi(I)}[B]_{\psi(J)})^\intercal\langle\phi(I), c^\intercal\rangle_{\mathcal{H}_Y} c \\ =&\langle X, \psi(J)^\intercal\rangle_{\mathcal{H}_X}({}_{\phi(I)}[B]_{\psi(J)})^\intercal\langle\phi(I), c^\intercal\rangle_{\mathcal{H}_Y} c \\ =&c^\intercal\langle c, \phi(I)^\intercal\rangle_{\mathcal{H}_Y}({}_{\phi(I)}[B]_{\psi(J)})\langle\psi(J), X\rangle_{\mathcal{H}_X}, \end{aligned}$$

we have

$$\begin{aligned} \langle c, BX\rangle_{\mathcal{H}_Y} =&\langle c, c^\intercal\rangle_{\mathcal{H}_Y}\langle c, \phi(I)^\intercal\rangle_{\mathcal{H}_Y}({}_{\phi(I)}[B]_{\psi(J)})\langle\psi(J), X\rangle_{\mathcal{H}_X} \\ =&\langle c, \phi(I)^\intercal\rangle_{\mathcal{H}_Y}({}_{\phi(I)}[B]_{\psi(J)})\langle\psi(J), X\rangle_{\mathcal{H}_X} \\ =&\langle c, \phi(I)^\intercal\rangle_{\mathcal{H}_Y}({}_{\phi(I)}[B]_{\psi(J)})\langle\psi(J), b^\intercal\rangle_{\mathcal{H}_X}\langle b, X\rangle_{\mathcal{H}_X}. \end{aligned}$$

where, for the second equality, we have used the relation $\langle c, c^\intercal\rangle_{\mathcal{H}_Y} = I_l$. Substituting the above into (3), we have

$$\langle Y, c\rangle_{\mathcal{H}_Y} = \mu + \langle c, \phi(I)^\intercal\rangle_{\mathcal{H}_Y}({}_{\phi(I)}[B]_{\psi(J)})\langle\psi(J), b^\intercal\rangle_{\mathcal{H}_X}\langle b, X\rangle_{\mathcal{H}_X} + \langle\epsilon, c\rangle_{\mathcal{H}_Y}.$$

In other words,

$$\tilde{Y} = \mu + \Gamma({}_{\phi(I)}[B]_{\psi(J)})\Phi^\intercal\tilde{X} + \tilde{\epsilon}, \tag{4}$$

where $\Gamma = \langle c, \phi(I)^\intercal\rangle_{\mathcal{H}_Y}$ and $\Phi = \langle b, \psi(J)^\intercal\rangle$. It remains to show that span(Γ) and span(Φ) are the envelopes stated in the theorem.

Let

$$X^\triangle = \langle\psi(J), X\rangle_{\mathcal{H}_X}, \quad Y^\triangle = \langle\phi(I), Y\rangle_{\mathcal{H}_Y}, \quad \epsilon^\triangle = \langle\phi(I), \epsilon\rangle_{\mathcal{H}_Y}.$$

Taking $c = \phi(I)$ and $b = \psi(J)$, we get a special case of (4)

$$Y^\triangle = \mu + ({}_{\phi(I)}[B]_{\psi(J)})X^\triangle + \epsilon^\triangle. \tag{5}$$

Without assuming any further structure, (5) is just a multivariate linear model with response envelope $\mathcal{E}({}_{\phi(I)}[B]_{\psi(J)}; \Sigma_{\epsilon^\triangle}) = \mathbb{R}^s$, and predictor envelope $\mathcal{E}({}_{\phi(I)}[B]^\intercal_{\psi(J)}; \Sigma_{X^\triangle}) = \mathbb{R}^t$.

Now we show that the response envelope of (4) is span(Γ). We first derive the covariance matrix of $\tilde{\epsilon}$. Since

$$\epsilon = \sum_{i\in I}\langle\epsilon, \phi_i\rangle_{\mathcal{H}_Y}\phi_i + \sum_{i\notin I}\langle\epsilon, \phi_i\rangle_{\mathcal{H}_Y}\phi_i = \langle\epsilon, \phi(I)^\intercal\rangle_{\mathcal{H}_Y}\phi(I) + \sum_{i\notin I}\langle\epsilon, \phi_i\rangle_{\mathcal{H}_Y}\phi_i,$$

we have

$$\langle c, \epsilon\rangle_{\mathcal{H}_Y} = \sum_{i\in I}\langle\epsilon, \phi_i\rangle_{\mathcal{H}_Y}\langle c, \phi_i\rangle_{\mathcal{H}_Y} + \sum_{i\notin I}\langle\epsilon, \phi_i\rangle_{\mathcal{H}_Y}\langle c, \phi_i\rangle_{\mathcal{H}_Y}.$$

Hence

$$\begin{aligned}\Sigma_{\tilde{\epsilon}} =&\Gamma\Sigma_{\epsilon\triangle}\Gamma^\intercal + \sum_{i\notin I}\mathrm{var}\big(\langle\epsilon,\phi_i\rangle_{\mathcal{H}_Y}\big)\langle c,\phi_i\rangle_{\mathcal{H}_Y}\langle\phi_i,c^\intercal\rangle_{\mathcal{H}_Y}\\ \equiv&\Gamma\Sigma_{\epsilon\triangle}\Gamma^\intercal + \Lambda_\epsilon.\end{aligned}$$

Note that, if $i\in I$ and $j\notin I$, then $\langle c,\phi_i\rangle_{\mathcal{H}_Y}^\intercal\langle c,\phi_j\rangle_{\mathcal{H}_Y}=0$. This is because

$$0=\langle\phi_i,\phi_j\rangle_{\mathcal{H}_Y}=\Big\langle\langle\phi_i,c^\intercal\rangle_{\mathcal{H}_Y}c,\phi_j\Big\rangle_{\mathcal{H}_Y}=\langle\phi_i,c^\intercal\rangle_{\mathcal{H}_Y}\langle c,\phi_j\rangle_{\mathcal{H}_Y}.$$

This implies $\Gamma^\intercal\langle c,\phi_j\rangle_{\mathcal{H}_Y}=0$ whenever $j\notin I$, which further implies $\Lambda_\epsilon\Gamma=0$.

For the rest of the proof, we abbreviate ${}_{\phi(I)}[B]_{\psi(J)}$ by $[B]$, which will not cause any ambiguity. Since $\Sigma_{\tilde{\epsilon}}=\Gamma\Sigma_{\epsilon\triangle}\Gamma^\intercal+\Lambda_\epsilon$, we have

$$\mathcal{E}(\Gamma[B]\Phi^\intercal;\Sigma_{\tilde{\epsilon}})=\mathcal{E}(\Gamma[B];\Sigma_{\tilde{\epsilon}})=\mathcal{E}(\Gamma[B];\Gamma\Sigma_{\epsilon\triangle}\Gamma^\intercal+\Lambda_\epsilon).$$

By $\Lambda_\epsilon\Gamma=0$ and Lemma 2,

$$\mathcal{E}(\Gamma[B];\Gamma\Sigma_{\epsilon\triangle}\Gamma^\intercal+\Lambda_\epsilon)=\mathcal{E}(\Gamma[B];\Gamma\Sigma_{\epsilon\triangle}\Gamma^\intercal).$$

By Lemma 1,

$$\mathcal{E}(\Gamma[B];\Gamma\Sigma_{\epsilon\triangle}\Gamma^\intercal)=\Gamma\mathcal{E}([B],\Sigma_{\epsilon\triangle})=\Gamma\mathbb{R}^s=\mathrm{span}(\Gamma).$$

Therefore $\mathcal{E}(\Gamma[B]\Phi^\intercal;\Sigma_{\tilde{\epsilon}})=\mathrm{span}(\Gamma)$.

Next, we show that

$$\mathcal{E}(\Phi[B]^\intercal\Gamma^\intercal;\Sigma_{\tilde{X}})=\mathcal{E}(\Phi[B]^\intercal;\Sigma_{\tilde{X}})=\mathrm{span}(\Phi).$$

We first derive $\Sigma_{\tilde{X}}$. Since

$$X=\sum_{j\in J}\langle X,\psi_j\rangle_{\mathcal{H}_X}\psi_j+\sum_{j\notin J}\langle X,\psi_j\rangle_{\mathcal{H}_X}\psi_j,$$

we have

$$\begin{aligned}
\tilde{X} &= \langle b, X\rangle_{\mathcal{H}_X} \\
&= \sum_{j\in J}\langle X, \psi_j\rangle_{\mathcal{H}_X}\langle b, \psi_j\rangle_{\mathcal{H}_X} + \sum_{j\notin J}\langle X, \psi_j\rangle_{\mathcal{H}_X}\langle b, \psi_j\rangle_{\mathcal{H}_X} \\
&= \Phi X^{\triangle} + \sum_{j\notin J}\langle X, \psi_j\rangle_{\mathcal{H}_X}\langle b, \psi_j\rangle_{\mathcal{H}_X}.
\end{aligned}$$

Hence

$$\begin{aligned}
\Sigma_{\tilde{X}} =& \Phi\Sigma_{X^{\triangle}}\Phi^{\intercal} + \sum_{j\notin J}\mathrm{var}(\langle X, \psi_j\rangle_{\mathcal{H}_X})\langle b, \psi_j\rangle_{\mathcal{H}_X}\langle b^{\intercal}, \psi_j\rangle_{\mathcal{H}_X} \\
\equiv& \Phi\Sigma_{X^{\triangle}}\Phi^{\intercal} + \Lambda_X.
\end{aligned}$$

By the similar argument used in deriving the response envelope, we have that $\Lambda_X\Phi = 0$.

Note that

$$\mathcal{E}(\Phi[B]^{\intercal}\Phi; \Sigma_{\tilde{X}}) = \mathcal{E}(\Phi[B]^{\intercal}; \Sigma_{\tilde{X}}) = \mathcal{E}(\Phi[B]^{\intercal}; \Phi\Sigma_{X^{\triangle}}\Phi^{\intercal} + \Lambda_X).$$

By $\Lambda_X\Phi = 0$ and Lemma 2,

$$\mathcal{E}(\Phi[B]^{\intercal}; \Phi\Sigma_{X^{\triangle}}\Phi^{\intercal} + \Lambda_X) = \mathcal{E}(\Phi[B]^{\intercal}; \Phi\Sigma_{X^{\triangle}}\Phi^{\intercal}).$$

By Lemma 1,

$$\mathcal{E}(\Phi[B]^{\intercal}; \Phi\Sigma_{X^{\triangle}}\Phi^{\intercal}) = \Phi\mathcal{E}([B]^{\intercal}; \Sigma_{X^{\triangle}}) = \Phi\mathbb{R}^t = \mathrm{span}(\Phi).$$

Hence $\mathcal{E}(\Phi[B]^{\intercal}\Gamma; \Sigma_{\tilde{X}}) = \mathrm{span}(\Phi)$. □

5 Proof of Theorem 5

Proof. Since (X, Y) obeys the FELM in Definition 1, by (4), we have

$$\tilde{Y} = \mu + \Gamma({}_{\phi(I)}[B]_{\psi(J)})\Phi^{\intercal}\tilde{X} + \tilde{\epsilon},$$

thus we have

$$\Gamma({}_{\phi(I)}[B]_{\psi(J)})\Phi^{\intercal} = \Gamma\eta\Phi^{\intercal}.$$

Pre-multiply both sides by $\Gamma^{\intercal}$ and post-multiply both sides by Φ, we have ${}_{\phi(I)}[B]_{\psi(J)} = \eta$. $\square$

6 Details on estimation

The likelihood function of model (19) is

$$\begin{aligned}\ell(\mu, \Gamma, \Phi, \Omega, \Omega_0, \Delta, \Delta_0, \eta) &= -\frac{n(l+k)}{2}\log(2\pi) - \frac{n}{2}\log|\Phi\Delta\Phi^{\intercal} + \Phi_0\Delta_0\Phi_0^{\intercal}| \\ &- \frac{1}{2}\sum_{i=1}^{n}\tilde{X}_i^{\intercal}(\Phi\Delta\Phi^{\intercal} + \Phi_0\Delta_0\Phi_0^{\intercal})^{-1}\tilde{X}_i - \frac{n}{2}\log|\Gamma\Omega\Gamma^{\intercal} + \Gamma_0\Omega_0\Gamma_0^{\intercal}| \\ &- \frac{1}{2}\sum_{i=1}^{n}(\tilde{Y}_i - \mu - \Gamma\eta\Phi^{\intercal}\tilde{X}_i)^{\intercal}(\Gamma\Omega\Gamma^{\intercal} + \Gamma_0\Omega_0\Gamma_0^{\intercal})^{-1}(\tilde{Y}_i - \mu - \Gamma\eta\Phi^{\intercal}\tilde{X}_I).\end{aligned} \tag{6}$$

We maximize the likelihood function over $\Gamma^{\intercal}\Gamma = I_s$, $\Phi^{\intercal}\Phi = I_t$, $\Omega \succ 0$, $\Omega_0 \succ 0$, $\Delta \succ 0$, $\Delta_0 \succ 0$, $\eta \in \mathbb{R}^{s\times t}$, where $\succ$ denotes Loewner ordering, i.e., $A \succ B$ if and only if $A - B$ is a positive definite matrix. The likelihood involves the parameters μ, Γ, Φ, Ω, Ω_0, Δ, Δ_0, η. As shown in Cook (2018, page 114), given Γ and Φ, the MLE of parameters μ, Ω, Ω_0, Δ, Δ_0, η can all be expressed as explicit functions of Γ and Φ. Thus, the key is to find the MLE of Γ and Φ, or equivalently, $\mathcal{E}(\beta; \Sigma_{\tilde{\epsilon}})$ and $\mathcal{E}(\beta^{\intercal}; \Sigma_{\tilde{X}})$. To estimate $\mathcal{E}(\beta; \Sigma_{\tilde{\epsilon}})$ and $\mathcal{E}(\beta^{\intercal}; \Sigma_{\tilde{X}})$, we alternate between the following optimization problems

$$\hat{\mathcal{E}}(\beta; \Sigma_{\tilde{\epsilon}}) = \underset{\mathrm{span}(\Gamma)\in\mathcal{G}(l,s)}{\arg\min}\ \log|\Gamma^{\intercal}\hat{\Sigma}_{\tilde{Y}|\Phi^{\intercal}\tilde{X}}\Gamma| + \log|\Gamma^{\intercal}\hat{\Sigma}_{\tilde{Y}}^{-1}\Gamma|,$$

and

$$\hat{\mathcal{E}}(\beta^{\intercal};\Sigma_{\tilde{X}}) = \underset{\mathrm{span}(\Phi)\in\mathcal{G}(k,t)}{\arg\min} \log|\Phi^{\intercal}\hat{\Sigma}_{\tilde{X}|\Gamma^{\intercal}\tilde{Y}}\Phi| + \log|\Phi^{\intercal}\hat{\Sigma}_{\tilde{X}}^{-1}\Phi|,$$

where $\mathcal{G}(a,b), a \geq b$, denotes the $a\times b$ Grassmann manifold, $\hat{\Sigma}_{\tilde{X}}$ and $\hat{\Sigma}_{\tilde{Y}}$ denote the sample covariance matrices of $\tilde{X}$ and $\tilde{Y}$, and $\hat{\Sigma}_{\tilde{Y}|\Phi^{\intercal}\tilde{X}}$ and $\hat{\Sigma}_{\tilde{X}|\Gamma^{\intercal}\tilde{Y}}$ denote the residual sample variance matrices of the regressions $\tilde{Y}$ vs $\Phi^{\intercal}\tilde{X}$ and $\tilde{X}$ vs $\Gamma^{\intercal}\tilde{Y}$ respectively. That is,

$$\begin{aligned}
\hat{\Sigma}_{\tilde{Y}|\Phi^{\intercal}\tilde{X}} =& \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \hat{\gamma}^{\intercal}\Phi^{\intercal}\tilde{X}_i)(\tilde{Y}_i - \hat{\gamma}^{\intercal}\Phi^{\intercal}\tilde{X}_i)^{\intercal},\\
\hat{\Sigma}_{\tilde{X}|\Gamma^{\intercal}\tilde{Y}} =& \frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i - \hat{\delta}^{\intercal}\Gamma^{\intercal}\tilde{Y}_i)(\tilde{X}_i - \hat{\delta}^{\intercal}\Gamma^{\intercal}\tilde{Y}_i)^{\intercal},
\end{aligned}$$

where $\hat{\gamma}$ and $\hat{\delta}$ are the usual regression coefficients:

$$\begin{aligned}
\hat{\gamma} &= \left(\sum_{i=1}^{n}\Phi^{\intercal}\tilde{X}_i\tilde{X}_i^{\intercal}\Phi\right)^{-1}\sum_{i=1}^{n}\Phi^{\intercal}\tilde{X}_i\tilde{Y}_i^{\intercal},\\
\hat{\delta} &= \left(\sum_{i=1}^{n}\Gamma^{\intercal}\tilde{Y}_i\tilde{Y}_i^{\intercal}\Gamma\right)^{-1}\sum_{i=1}^{n}\Gamma^{\intercal}\tilde{Y}_i\tilde{X}_i^{\intercal}.
\end{aligned}$$

A package for optimization over Grassmann manifold for envelope models can be found on CRAN (R package *Renvlp*. Links to envelope computing packages are available at z.umn.edu/envelopes.).

7 Forms of G and J

For an $a\times a$ symmetric matrix M, we define $C_a \in \mathbb{R}^{a^2\times[a(a+1)/2]}$ and $E_a \in \mathbb{R}^{[a(a+1)/2]\times a^2}$ as the contraction matrix and expansion matrix that connect the vec$(\cdot)$ and vech$(\cdot)$ operator: $\mathrm{vech}(M) = C_a\mathrm{vec}(M)$ and $\mathrm{vec}(M) = E_a\mathrm{vech}(M)$. See Henderson and Searle (1979). Then

$$G = \frac{\partial v_1}{\partial v_2^{\intercal}} = \begin{pmatrix}
\Phi\eta^{\intercal}\otimes I_l & \Phi\otimes\Gamma & (I_k\otimes\Gamma\eta)K_{kt} & 0 & 0 & 0 & 0 & 0\\
G_{21} & 0 & 0 & G_{24} & G_{25} & 0 & 0 & 0\\
0 & 0 & G_{33} & 0 & 0 & G_{36} & G_{37} & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & I_l
\end{pmatrix},$$

where $K_{kt} \in \mathbb{R}^{kt \times kt}$ is a commutation matrix such that, for any $k \times t$ matrix M, $\mathrm{vec}(M^\intercal) = K_{kt}\mathrm{vec}(M)$, and

$$
\begin{aligned}
G_{21} &= 2C_l(\Gamma\Omega \otimes I_l - \Gamma \otimes \Gamma_0\Omega_0\Gamma_0^\intercal), & G_{24} &= C_l(\Gamma \otimes \Gamma)E_s, \\
G_{25} &= C_l(\Gamma_0 \otimes \Gamma_0)E_{l-s}, & G_{33} &= 2C_k(\Phi\Delta \otimes I_k - \Phi \otimes \Phi_0\Delta_0\Phi_0^\intercal), \\
G_{36} &= C_k(\Phi \otimes \Phi)E_t, & G_{37} &= C_k(\Phi_0 \otimes \Phi_0)E_{k-t}.
\end{aligned}
$$

Since ϵ is a Gaussian element in $\mathcal{H}_Y$ and X is a Gaussian element in $\mathcal{H}_X$ independent of ϵ, $\tilde{X}$ and $\tilde{\epsilon}$ are independent Gaussian random vectors. Then the Fisher information for v_1 under the full model ($s = l$, $t = k$) is

$$
J = \begin{pmatrix}
\Sigma_{\tilde{X}} \otimes \Sigma_{\tilde{\epsilon}}^{-1} & 0 & 0 & 0 \\
0 & \frac{1}{2}E_l^\intercal(\Sigma_{\tilde{\epsilon}}^{-1} \otimes \Sigma_{\tilde{\epsilon}}^{-1})E_l & 0 & 0 \\
0 & 0 & \frac{1}{2}E_k^\intercal(\Sigma_{\tilde{X}}^{-1} \otimes \Sigma_{\tilde{X}}^{-1})E_k & 0 \\
0 & 0 & 0 & \Sigma_{\tilde{\epsilon}}^{-1}
\end{pmatrix}.
$$

8 Proof of Theorem 7

To show the consistency of $(\hat{s}, \hat{t})$, we first show that the parameter space of a MELM with response and predictor envelopes $\mathcal{E}_{1,\epsilon}$ and $\mathcal{E}_{1,X}$ is nested within that of a MELM with response and predictor envelopes $\mathcal{E}_{2,\epsilon}$ and $\mathcal{E}_{2,X}$ if $\mathcal{E}_{1,\epsilon} \subseteq \mathcal{E}_{2,\epsilon}$ and $\mathcal{E}_{1,X} \subseteq \mathcal{E}_{2,X}$. To see that, we perform a slight reparameterization on MELM. In the reparameterization, we require that Γ and Γ_0 are chosen such that Ω and Ω_0 are diagonal matrices with their diagonal elements in descending order, and Φ and Φ_0 are chosen such that Δ and Δ_0 are diagonal matrices with their diagonal elements in descending order. We further assume that the first nonzero element in each column of Γ, Γ_0, Φ and Φ_0 are positive. Then Γ, Γ_0, Φ and Φ_0 are unique orthonormal basis of $\mathcal{E}(\beta; \Sigma_{\tilde{\epsilon}})$, $\mathcal{E}(\beta; \Sigma_{\tilde{\epsilon}})^\perp$, $\mathcal{E}(\beta^\intercal; \Sigma_{\tilde{X}})$ and $\mathcal{E}(\beta^\intercal; \Sigma_{\tilde{X}})^\perp$. Under this parameterization, the full parameter vector in MELM (19) with dimensions (s, t) is

$$\Theta(s,t) = (\mathrm{vec}\{(\Gamma,\Gamma_0)\}^\top, \mathrm{vec}(\eta)^\top, \mathrm{vec}\{(\Phi,\Phi_0)\}^\top, \mathrm{diagmv}(\Omega)^\top, \mathrm{diagmv}(\Omega_0)^\top, \\ \mathrm{diagmv}(\Delta)^\top, \mathrm{diagmv}(\Delta_0)^\top, \mu^\top)^\top.$$

For convenience, we assume all the eigenvalues in Ω, Ω_0 are distinct, and all the eigenvalues in Δ, Δ_0 are distinct. We believe this assumption can be relaxed by a more elaborate proof.

Lemma 3. *Suppose the elements of* $(\mathrm{diagmv}(\Omega)^\top, \mathrm{diagmv}(\Omega_0)^\top)$ *are distinct, and the elements of* $(\mathrm{diagmv}(\Delta)^\top, \mathrm{diagmv}(\Delta_0)^\top)$ *are distinct. Then the parameter space of a MELM with response and predictor envelopes* $\mathcal{E}_{1,\epsilon}$ *and* $\mathcal{E}_{1,X}$ *is contained in that of a MELM with response and predictor envelopes* $\mathcal{E}_{2,\epsilon}$ *and* $\mathcal{E}_{2,X}$ *if and only if* $\mathcal{E}_{1,\epsilon} \subseteq \mathcal{E}_{2,\epsilon}$ *and* $\mathcal{E}_{1,X} \subseteq \mathcal{E}_{2,X}$.

PROOF. Let s_1, t_1, s_2 and t_2 denote the dimensions of $\mathcal{E}_{1,\epsilon}$, $\mathcal{E}_{1,X}$, $\mathcal{E}_{2,\epsilon}$ and $\mathcal{E}_{2,X}$. For a generic pair of integers (s, t), let MELM(s, t) denote the MELM with dimensions (s, t). Note that each column of Γ in MELM(s_1, t_1) is either a column of Γ or a column of Γ_0 in MELM(s_2, t_2). And each column of Γ_0 in MELM(s_1, t_1) is either a column of Γ or a column of Γ_0 in the MELM(s_2, t_2). So the parameter $\mathrm{vec}\{(\Gamma,\Gamma_0)\}$ in MELM(s_1, t_1) has a one-to-one correspondence with the parameter $\mathrm{vec}\{(\Gamma,\Gamma_0)\}$ in MELM(s_2, t_2). The same result also holds for the parameter $\mathrm{vec}\{(\Phi,\Phi_0)\}$, $(\mathrm{diagmv}(\Omega)^\top, \mathrm{diagmv}(\Omega_0)^\top)$ and $(\mathrm{diagmv}(\Delta)^\top, \mathrm{diagmv}(\Delta_0)^\top)$. Recall that the η matrix in a MELM contains the coefficients of the regression of $\Gamma^\top Y$ on $\Phi^\top X$. Thus the η matrix in MELM(s_1, t_1) is a submatrix of the η matrix in MELM(s_2, t_2) if and only if $\mathcal{E}_{1,\epsilon} \subseteq \mathcal{E}_{2,\epsilon}$ and $\mathcal{E}_{1,X} \subseteq \mathcal{E}_{2,X}$. In other words, the parameter space of MELM(s_1, t_1) is contained in that of MELM(s_2, t_2) if and only if $\mathcal{E}_{1,\epsilon} \subseteq \mathcal{E}_{2,\epsilon}$ and $\mathcal{E}_{1,X} \subseteq \mathcal{E}_{2,X}$. □

We now prove the consistency of $(\hat{s}, \hat{t})$.

PROOF. Let $\ell(\Theta(s,t))$ denote the log likelihood in (6) for a fixed (s,t), let $\ell_0(\Theta(s,t)) = E[\ell(\Theta(s,t))]$, and let

$$\tilde{\ell}_0(s,t) = \sup_{\Theta(s,t)} \ell_0(\Theta(s,t)).$$

Then $E[\mathrm{BIC}(s,t)] = -2\tilde{\ell}_0(s,t) + \log(n)K(s,t)$. For the rest of the proof, we denote the response and predictor envelopes for MELM(s,t) as $\mathcal{E}_\epsilon$ and $\mathcal{E}_X$, and the true response and predictor envelopes (that is, the envelopes for MELM(s_0, t_0)) as $\mathcal{E}_{0,\epsilon}$ and $\mathcal{E}_{0,X}$. We consider the following two scenarios.

Scenario 1: $\mathcal{E}_{0,\epsilon} \subseteq \mathcal{E}_\epsilon$, $\mathcal{E}_{0,X} \subseteq \mathcal{E}_X$ and $(s_0, t_0) \neq (s, t)$. From Lemma 3, MELM(s_0, t_0) is nested within MELM(s, t). Hence,

$$\tilde{\ell}_0(s_0, t_0) = \tilde{\ell}_0(s, t). \tag{7}$$

By the asymptotic distribution of Wilks' statistic (see, for example, Li and Babu (2019, Chapter 11)), we have, for any $(s, t) \in \Xi$,

$$\hat{l}(s, t) - \hat{l}(s_0, t_0) = O_P(1). \tag{8}$$

Hence

$$\begin{aligned} &\mathrm{BIC}(s, t) - \mathrm{BIC}(s_0, t_0) \\ =& -2[\hat{l}(s, t) - \hat{l}(s_0, t_0)] + \log(n)[K(s, t) - K(s_0, t_0)] \\ =& \, O_P(1) + \log(n)[K(s, t) - K(s_0, t_0)]. \end{aligned} \tag{9}$$

Since $\mathcal{E}_{0,\epsilon} \subseteq \mathcal{E}_\epsilon$, $\mathcal{E}_{0,X} \subseteq \mathcal{E}_X$, $(s_0, t_0) \neq (s, t)$ implies $s_0 \leq s$, $t_0 \leq t$, and at least one of the inequalities is strict, we have $K(s, t) > K(s_0, t_0)$, which, combined with (9), implies $\mathrm{BIC}(s, t) > \mathrm{BIC}(s_0, t_0)$ with probability tending to 1.

Scenario 2: at least one of $\mathcal{E}_{0,\epsilon} \nsubseteq \mathcal{E}_\epsilon$ and $\mathcal{E}_{0,X} \nsubseteq \mathcal{E}_X$ is true. Under this scenario, for some $c > 0$,

$$\frac{1}{n}\tilde{\ell}_0(s, t) > \frac{1}{n}\tilde{\ell}_0(s_0, t_0) + c.$$

Hence, with probability tending to 1,

$$\begin{aligned} &\mathrm{BIC}(s_0, t_0) - \mathrm{BIC}(s, t) \\ &= 2n\left\{\frac{1}{n}\tilde{\ell}_0(s, t) - \frac{1}{n}\tilde{\ell}_0(s_0, t_0)\right\} + \log(n)[K(s_0, t_0) - K(s, t)] \\ &> nc + \log(n)[K(s_0, t_0) - K(s, t)]. \end{aligned}$$

Since $\log(n)/n \to 0$, $\mathrm{BIC}(s_0, t_0) - \mathrm{BIC}(s, t) > nc/2$. Therefore $\mathrm{BIC}(s, t) > \mathrm{BIC}(s_0, t_0)$ with probability tending to 1.

Combining the results for the two scenarios, we see that, whenever $(s, t) \neq (s_0, t_0)$, $\mathrm{BIC}(s, t) > \mathrm{BIC}(s_0, t_0)$ with probability tending to 1. Meanwhile, if $(s, t) = (s_0, t_0)$, then $\mathrm{BIC}(s, t) = \mathrm{BIC}(s_0, t_0)$. Because Ξ is a finite set, we conclude that with probability tending to 1, $\mathrm{BIC}(s, t)$ is minimized uniquely at (s_0, t_0). $\square$

9 Additional simulations

9.1 Larger envelope dimensions

To study the performance of FELM with larger envelope dimensions, we consider the simulation setting with envelope dimensions (s, t) $(3, 3)$, …, $(6, 6)$ and full dimensions. Everything else in the model is generated in the same way as in Section 9 of the manuscript. Note that the dimensions of response envelope and predictor envelope can be different, which can induce many combinations. Here we increase the dimensions s and t together instead of looking into all possible combinations for it better reveals the trend. When $(s, t) = (3, 3)$, the b_{ij}'s in the linear operator B are all zero except that $b_{22} = -1.25$, $b_{24} = -1$, $b_{42} = b_{44} = 0.4$, $b_{53} = 1/28$. When $(s, t) = (4, 4)$, we further set $b_{33} = 5/56$, $b_{35} = 3/80$; when $(s, t) = (5, 5)$, we further set $b_{62} = 0.3$, $b_{66} = -3/32$; when $(s, t) = (6, 6)$, we further set $b_{76} = 1/32$, $b_{77} = -1/24$; when the envelope dimensions are the same as full dimensions, we further set $b_{87} = 1/48$, $b_{88} = 1/32$. The sample size n was fixed at 25. The implementation for each method was the same as that for fixed dimension. The results are shown in Table 1.

Comparing with FFFR, the reduction of the prediction error by FELM decreases as (s, t) increases, from 19.4% when $s = t = 2$, to 13.0%, 8.7%, 5.2%, 2.2% as $s = t$ increases to 3, 4, 5 and 6. This is consistent with the theory that the envelope method is most effective when the dimensions of the envelopes are small. When the envelope subspace is the full space, it reduces to the full model, which is why the last column that the results for FELM and FFFR are the same. Both the direct method and K-L expansion method have the same trend.

Table 1: Comparison on mean squared prediction errors

(s, t)		$(2, 2)$	$(3, 3)$	$(4, 4)$	$(5, 5)$	$(6, 6)$	Full
Direct	FELM	6.38	6.90	7.23	7.61	7.88	8.07
	FFFR	7.92	7.93	7.92	8.03	8.06	8.07
	PCR	12.99	13.62	14.43	10.53	8.99	8.07
	PLS	13.01	11.96	12.02	9.72	8.22	8.07
K-L expansion	FELM	6.38	6.92	7.23	7.61	7.88	8.07
	FFFR	7.92	7.93	7.92	8.03	8.06	8.07
	PCR	12.99	13.62	14.43	10.53	8.99	8.07
	PLS	13.01	11.96	12.02	9.72	8.22	8.07

9.2 Computing time

The computation time for each method in Table 1 of the main text is included below in Table 2. For each case, the computation time is the average of 100 runs performed on a single core of a MacBook Pro laptop with 2.6 GHz 6-core Intel Core i7 processor. For fixed dimension, each method can be computed well under 1 second. FFFR, PCR and PLS take roughly the same amount of computation time. For larger sample sizes, the computation time of FELM is also comparable with the other methods. However, when the sample size is smaller, FELM requires longer computation time than the other methods. This is because FELM involves a nonlinear optimization on a Grassmann manifold, which is typically faster with a larger sample size since one can obtain a better starting value leading to a faster convergence.

With selected dimensions, PCR and PLS take slightly longer to compute than FFFR due to selection of number of components. The computation time of FELM is significantly longer (up to 4 seconds) since it has to fit an envelope model with all possible dimensions and compare the BIC values as discussed in Section 7 on order determination. However, as the sample size increases, the computation time of FELM decreases. Since the computation of BIC can be paralleled, in the high-dimensional setting, parallel computing can be utilized to reduce the computation time. The computation time of FELM we report here is based on a single core.

Table 2: Comparison on computational time (Seconds)

n		25	50	100	200	400	25	50	100	200	400
Method		Fixed dimension					Selected dimension				
Direct	FELM	0.249	0.133	0.163	0.254	0.447	3.919	1.851	1.589	1.705	1.939
	FFFR	0.034	0.066	0.113	0.213	0.410	0.032	0.055	0.111	0.238	0.516
	PCR	0.034	0.067	0.111	0.208	0.413	0.041	0.063	0.114	0.243	0.518
	PLS	0.034	0.066	0.113	0.213	0.412	0.070	0.090	0.145	0.281	0.564
K-L expansion	FELM	0.114	0.130	0.176	0.246	0.512	2.901	1.983	1.887	1.855	1.670
	FFFR	0.036	0.074	0.122	0.209	0.471	0.035	0.069	0.137	0.263	0.444
	PCR	0.036	0.075	0.119	0.209	0.472	0.044	0.078	0.146	0.267	0.454
	PLS	0.036	0.074	0.118	0.206	0.468	0.077	0.112	0.183	0.316	0.498

9.3 Irregular time points

We used the same simulation setting in Section 9 of the paper, except that elements of $\{t_1, \ldots, t_{10}\}$ were generated from uniform distributions on $(0.05, 0.1)$, $(0.15, 0.2)$, $\ldots$, $(0.95, 1)$ for each sample (X_i, Y_i). The results have similar patterns to those in Table 1 of the main text.

Table 3 shows the above mean squared error averaged over the 100 simulated samples.

Table 3: Comparison on mean squared prediction errors

n		25	50	100	200	400	25	50	100	200	400
Method		True dimension					Selected dimension				
Direct	FELM	8.50	6.60	6.27	6.08	6.03	8.41	6.58	6.26	6.07	6.03
	FFFR	9.53	7.27	6.58	6.21	6.10	9.53	7.27	6.58	6.21	6.10
	PCR	12.93	12.11	11.62	11.40	11.33	9.71	7.30	6.58	6.21	6.10
	PLS	12.24	10.23	8.60	7.66	7.24	12.48	11.54	10.64	10.10	9.75
K-L expansion	FELM	8.56	6.60	6.27	6.08	6.04	8.48	6.58	6.26	6.07	6.03
	FFFR	9.53	7.27	6.58	6.21	6.10	9.53	7.27	6.58	6.21	6.10
	PCR	12.93	12.11	11.62	11.40	11.33	9.71	7.30	6.58	6.21	6.10
	PLS	12.24	10.23	8.60	7.66	7.24	12.52	11.54	10.64	10.10	9.75

9.4 Representation errors

To investigate the impact of the representation error caused by using the wrong basis in our simulation in Section 9, we considered the ideal scenario where the basis generating the data agrees with the basis used for estimation. We used the simulation settings in Section 9 but, this time, we set $b_i = \psi_i$, for $i = 1, 2, \ldots$, $c_j = \chi_j$, for $j = 1, 2, \ldots$. In this case, the direct method and K-L expansion method are numerically equivalent. We then reperformed the simulation in Section 9 using the true basis. The results are shown in Table 4 below. Compared with the results in Table 1 of the main paper, where the bases are assumed to be unknown and representation errors are present, the mean squared prediction errors are reduced for both FELM and FFFR. However, the reduction is less noticeable when the sample size increases. The difference in prediction errors between FELM and FFFR also seems to be less significant than the case where the representation errors are present. Take sample size 25 with fixed dimension as an example, the FELM reduces the prediction error from FFFR by 11.4% with known bases, and the reduction is 19.4% with unknown bases. This implies that FELM is less prone to be affected by the representation error.

Table 4: Comparison on mean squared prediction errors

n	25	50	100	200	400	25	50	100	200	400
	Fixed dimension					Selected dimension				
FELM	5.50	5.41	5.37	5.34	5.30	5.49	5.40	5.37	5.34	5.30
FFFR	6.21	5.69	5.50	5.40	5.34	6.21	5.69	5.50	5.40	5.34
PCR	13.00	12.19	11.86	11.79	11.70	6.21	5.69	5.50	5.40	5.34
PLS	10.00	8.09	7.05	6.57	6.35	12.00	10.80	9.84	8.93	8.09

10 Additional details on the analysis fo Covid-19 data

Since the mobility data has periodical trends, it is reasonable to include the Fourier functions in the basis. If we use only cubic splines in the basis functions, the prediction results look like those in Figure 1. The predictions tend to be smooth curves capturing the overall trend, while missing the obvious periodical fluctuation in the data. Therefore, we now use a combination of polynomial functions and Fourier functions as basis. More specifically, we used the basis functions 1, t, t^2, $\sqrt{2}\sin(8\pi t)$, $\sqrt{2}\cos(8\pi t)$, $\sqrt{2}\sin(10\pi t)$ and $\sqrt{2}\cos(10\pi t)$. The results are shown in Figure 2. The predictions are noticeably improved. The prediction for Camden County is one of the best cases in terms of prediction errors for both models. A typical prediction is presented in Figure 3which shows the results of the alphabetically first county, i.e. Atlantic County. The difference between FFFR and FELM is also more noticeable in the Atlantic County data.

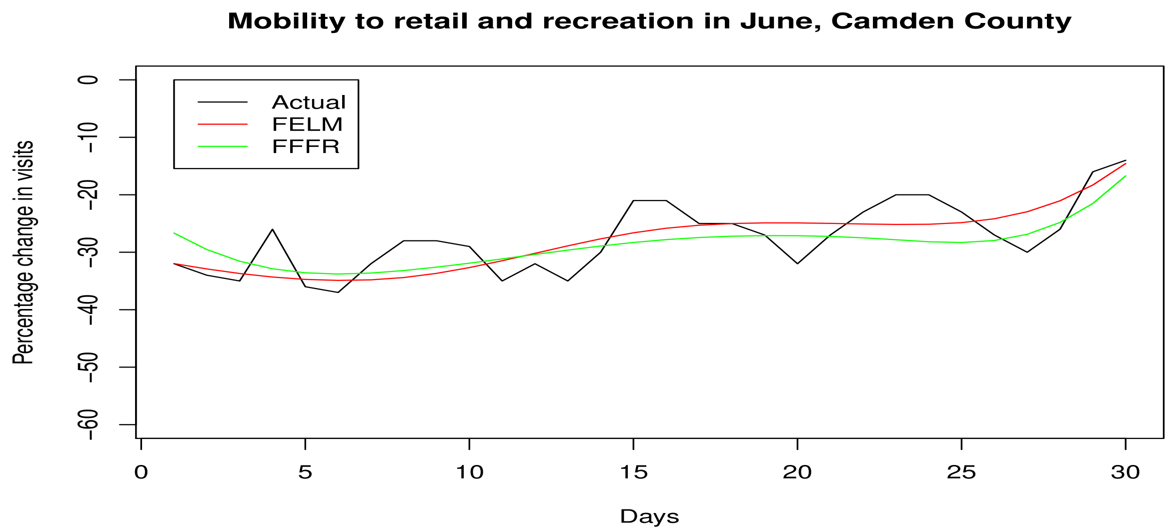

Figure 1: Results using spline basis: Actual data and predictions from FELM and FFFR.

References

Cook, D., Forzani, L., and Liu, L. (2020), "Envelopes for multivariate linear regression with linearly constrained coefficients," *arXiv preprint arXiv:2101.00514*.

Cook, R. D. (2018), *An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics*, Hoboken, NJ: John Wiley & Sons.

Henderson, H. V. and Searle, S. (1979), "Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics," *Canadian Journal of Statistics*, 7, 65–81.

Li, B. and Babu, G. J. (2019), *A Graduate Course on Statistical Inference*, Springer.

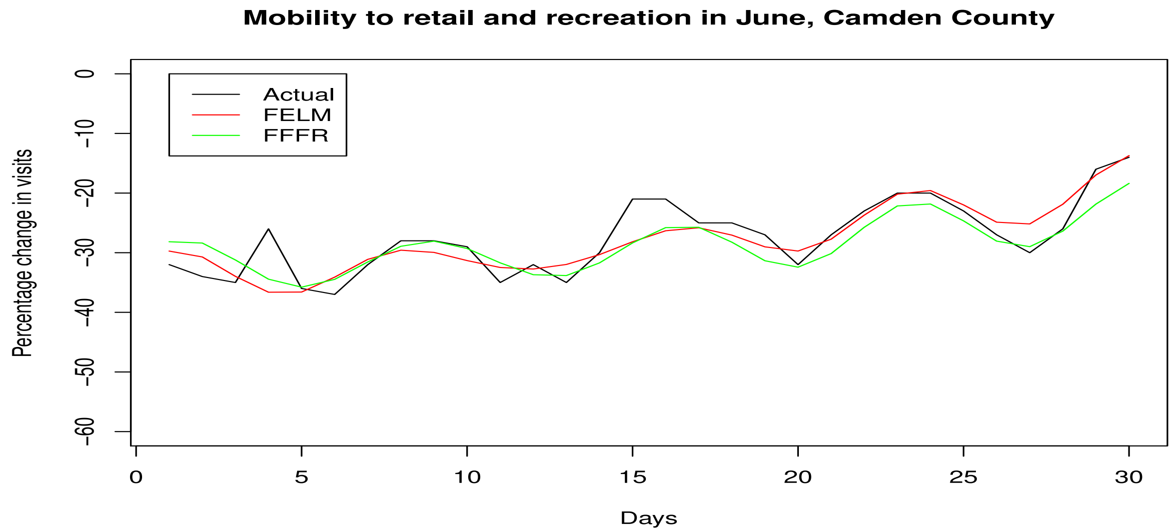

Figure 2: Mobility to retail and recreation for Camden County: Actual data and predictions from FELM and FFFR.

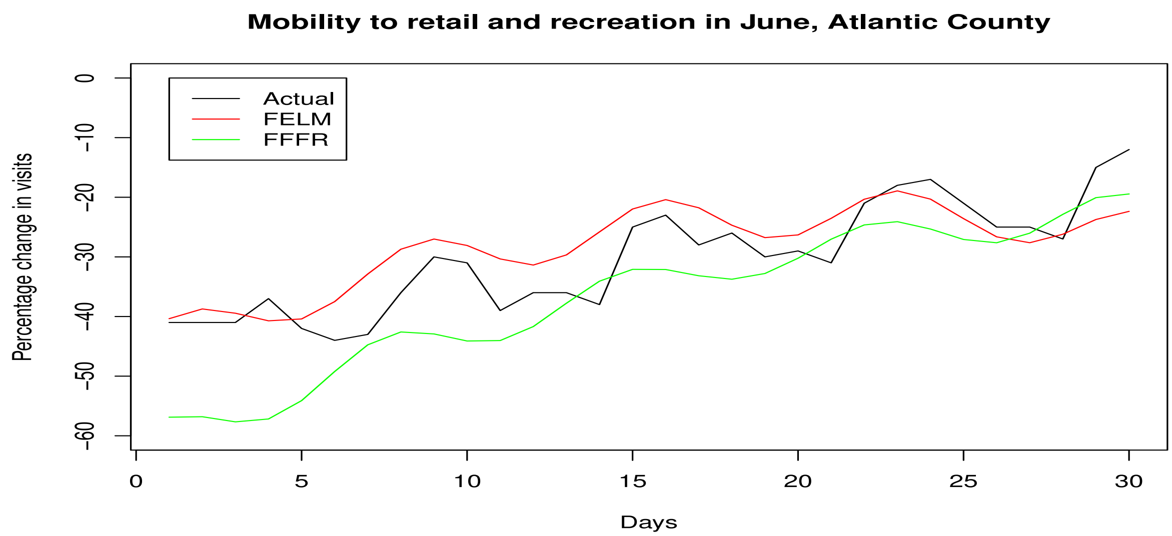

Figure 3: Mobility to retail and recreation for Atlantic County: Actual data and predictions from FELM and FFFR.