Supplementary Material for "A minimum discrepancy approach with Fourier transform in sufficient dimension reduction

Bentley University and University of Kentucky

Supplementary Material

S1 Proofs

Proof of Theorem 1: Let $\hat{\boldsymbol{\varphi}}_{\boldsymbol{\omega}} = \frac{1}{n}\sum_{k=1}^{n} e^{i\boldsymbol{\omega}^T\mathbf{y}_k}\mathbf{x}_k$, $\boldsymbol{\varphi}_{\boldsymbol{\omega}} = \mathrm{E}(e^{i\boldsymbol{\omega}^T\mathbf{Y}}\mathbf{X})$, $\mathbf{C}_{\boldsymbol{\omega}} = \mathrm{E}(e^{i\boldsymbol{\omega}^T\mathbf{Y}})$, $\hat{\mathbf{C}}_{\boldsymbol{\omega}} = \frac{1}{n}\sum_{k=1}^{n} e^{i\boldsymbol{\omega}^T\mathbf{y}_k}$, and $\mathbf{z}_k = \Sigma^{-1/2}(\mathbf{x}_k - \boldsymbol{\mu})$. We need a lemma in Li et al. (2003), that is

$$\hat{\Sigma}^{-1} - \Sigma^{-1} = -n^{-1}\Sigma^{-1/2}\sum_{k=1}^{n}(\mathbf{z}_k\mathbf{z}_k^T - I)\Sigma^{-1/2} + O_p(n^{-1}).$$

We consider

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\xi}}_{\boldsymbol{\omega}} - \boldsymbol{\xi}_{\boldsymbol{\omega}}) &= \sqrt{n}\hat{\Sigma}^{-1}(\hat{\boldsymbol{\varphi}}_{\boldsymbol{\omega}} - \hat{\mathbf{C}}_{\boldsymbol{\omega}}\bar{\mathbf{x}}) - \sqrt{n}\Sigma^{-1}(\boldsymbol{\varphi}_{\boldsymbol{\omega}} - \mathbf{C}_{\boldsymbol{\omega}}\boldsymbol{\mu}) \\ &= \sqrt{n}(\hat{\Sigma}^{-1} - \Sigma^{-1})(\boldsymbol{\varphi}_{\boldsymbol{\omega}} - \mathbf{C}_{\boldsymbol{\omega}}\boldsymbol{\mu}) \\ &\quad + \sqrt{n}\Sigma^{-1}[(\hat{\boldsymbol{\varphi}}_{\boldsymbol{\omega}} - \hat{\mathbf{C}}_{\boldsymbol{\omega}}\bar{\mathbf{x}}) - (\boldsymbol{\varphi}_{\boldsymbol{\omega}} - \mathbf{C}_{\boldsymbol{\omega}}\boldsymbol{\mu})] + O_p(n^{-1/2}).\end{aligned} \tag{S1.1}$$

The first term can be written as:

$$\begin{aligned}\sqrt{n}(\hat{\Sigma}^{-1}-\Sigma^{-1})(\boldsymbol{\varphi}_{\boldsymbol{\omega}}-\mathbf{C}_{\boldsymbol{\omega}}\boldsymbol{\mu}) &= -n^{-1/2}\Sigma^{-1/2}\textstyle\sum_{k=1}^{n}(\mathbf{z}_k\mathbf{z}_k^T-I)\Sigma^{-1/2}(\boldsymbol{\varphi}_{\boldsymbol{\omega}}-\mathbf{C}_{\boldsymbol{\omega}}\boldsymbol{\mu})+O_p(n^{-1/2}) \\ &= -n^{-1/2}\Sigma^{-1/2}\textstyle\sum_{k=1}^{n}(\mathbf{z}_k\mathbf{z}_k^T-I)\mathrm{E}(e^{i\boldsymbol{\omega}^T\mathbf{Y}}\mathbf{Z})+O_p(n^{-1/2}).\end{aligned} \tag{S1.2}$$

Then

$$\begin{aligned}\hat{\boldsymbol{\varphi}}_{\boldsymbol{\omega}}-\hat{\mathbf{C}}_{\boldsymbol{\omega}}\bar{\mathbf{x}} &= \tfrac{1}{n}\textstyle\sum_{k=1}^{n}e^{i\boldsymbol{\omega}^T\mathbf{y}_k}\mathbf{x}_k-\tfrac{1}{n}\sum_{k=1}^{n}e^{i\boldsymbol{\omega}^T\mathbf{y}_k}\bar{\mathbf{x}} \\ &= \tfrac{1}{n}\textstyle\sum_{k=1}^{n}(e^{i\boldsymbol{\omega}^T\mathbf{y}_k}-\mathrm{E}e^{i\boldsymbol{\omega}^T\mathbf{Y}})(\mathbf{x}_k-\boldsymbol{\mu})-\tfrac{1}{n}(\bar{\mathbf{x}}-\boldsymbol{\mu})\textstyle\sum_{k=1}^{n}(e^{i\boldsymbol{\omega}^T\mathbf{y}_k}-\mathrm{E}e^{i\boldsymbol{\omega}^T\mathbf{Y}}) \\ &= \tfrac{1}{n}\textstyle\sum_{k=1}^{n}(e^{i\boldsymbol{\omega}^T\mathbf{y}_k}-\mathrm{E}e^{i\boldsymbol{\omega}^T\mathbf{Y}})(\mathbf{x}_k-\boldsymbol{\mu})+O_p(n^{-1}).\end{aligned}$$

Therefore, the second term can be simplified as

$$\begin{aligned}\sqrt{n}\Sigma^{-1}[(\hat{\boldsymbol{\varphi}}_{\boldsymbol{\omega}}-\hat{\mathbf{C}}_{\boldsymbol{\omega}}\bar{\mathbf{x}})-(\boldsymbol{\varphi}_{\boldsymbol{\omega}}-\mathbf{C}_{\boldsymbol{\omega}}\boldsymbol{\mu})] &= n^{-1/2}\Sigma^{-1/2}\textstyle\sum_{k=1}^{n}[\Sigma^{-1/2}(e^{i\boldsymbol{\omega}^T\mathbf{y}_k}-\mathrm{E}e^{i\boldsymbol{\omega}^T\mathbf{Y}})(\mathbf{x}_k-\boldsymbol{\mu})] \\ &\quad -\sqrt{n}\Sigma^{-1}(\boldsymbol{\varphi}_{\boldsymbol{\omega}}-\mathbf{C}_{\boldsymbol{\omega}}\boldsymbol{\mu})+O_p(n^{-1/2}) \\ &= n^{-1/2}\Sigma^{-1/2}\textstyle\sum_{k=1}^{n}[\mathbf{z}_k(e^{i\boldsymbol{\omega}^T\mathbf{y}_k}-\mathrm{E}e^{i\boldsymbol{\omega}^T\mathbf{Y}})-\mathrm{E}(e^{i\boldsymbol{\omega}^T\mathbf{Y}}\mathbf{Z})] \\ &\quad +O_p(n^{-1/2}).\end{aligned} \tag{S1.3}$$

Then we put equations (S1.2) and (S1.3) into (S1.1):

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\xi}}_{\boldsymbol{\omega}}-\boldsymbol{\xi}_{\boldsymbol{\omega}}) &= n^{-1/2}\Sigma^{-1/2}\textstyle\sum_{k=1}^{n}[\mathbf{z}_k(e^{i\boldsymbol{\omega}^T\mathbf{y}_k}-\mathrm{E}e^{i\boldsymbol{\omega}^T\mathbf{Y}})-\mathrm{E}(e^{i\boldsymbol{\omega}^T\mathbf{Y}}\mathbf{Z})-(\mathbf{z}_k\mathbf{z}_k^T-I)\mathrm{E}(e^{i\boldsymbol{\omega}^T\mathbf{Y}}\mathbf{Z})] \\ &\quad +O_p(n^{-1/2}) \\ &= n^{-1/2}\Sigma^{-1/2}\textstyle\sum_{k=1}^{n}\{\mathbf{z}_k[e^{i\boldsymbol{\omega}^T\mathbf{y}_k}-\mathrm{E}e^{i\boldsymbol{\omega}^T\mathbf{Y}}-\mathbf{z}_k^T\mathrm{E}(e^{i\boldsymbol{\omega}^T\mathbf{Y}}\mathbf{Z})]\}+O_p(n^{-1/2}) \\ &= n^{-1/2}\Sigma^{-1/2}\textstyle\sum_{k=1}^{n}\mathbf{z}_k\varepsilon_{\boldsymbol{\omega},k}+O_p(n^{-1/2}),\end{aligned}$$

where $\varepsilon_{\boldsymbol{\omega},k}=e^{i\boldsymbol{\omega}^T\mathbf{y}_k}-\mathrm{E}e^{i\boldsymbol{\omega}^T\mathbf{Y}}-\mathbf{z}_k^T\mathrm{E}(e^{i\boldsymbol{\omega}^T\mathbf{Y}}\mathbf{Z})$. Let $\boldsymbol{\varepsilon}_k=(\varepsilon^R_{\boldsymbol{\omega}_1,k},\varepsilon^I_{\boldsymbol{\omega}_1,k},...,\varepsilon^R_{\boldsymbol{\omega}_m,k},\varepsilon^I_{\boldsymbol{\omega}_m,k})^T$

with $k = 1, ..., n$. Then we have $\sqrt{n}[\text{vec}(\hat{\boldsymbol{\xi}}) - \text{vec}(\beta\nu)] = n^{-1/2}\sum_{k=1}^{n}(\Sigma^{-1/2}\mathbf{z}_k\boldsymbol{\varepsilon}_k^T) + O_p(n^{-1/2})$. Thus

$$\sqrt{n}[\text{vec}(\hat{\boldsymbol{\xi}}) - \text{vec}(\beta\nu)] \xrightarrow{D} N(0, \Gamma),$$

where $\Gamma = \text{Cov}[\text{vec}(\Sigma^{-1/2}\mathbf{Z}\boldsymbol{\epsilon}^T)] \in \mathbb{R}^{2pm\times 2pm}$. □

Proof of Theorem 2: Because $\hat{\Gamma}$ converges to Γ in probability, the asymptotic distribution of $n\hat{F}_d$ is the same as that of $n\hat{H}_d$ using Lemma A.3 of Cook and Ni (2005), where $H_d(B, C) = [\text{vec}(\hat{\boldsymbol{\xi}}) - \text{vec}(BC)]^T\Gamma^{-1}[\text{vec}(\hat{\boldsymbol{\xi}}) - \text{vec}(BC)]$. We use one full-rank reparameterization of (β, ν). Let $\beta = (\beta_1^T, \beta_2^T)^T$, where $\beta_1 = I_d \in \mathbb{R}^{d\times d}$ and $\beta_2 \in \mathbb{R}^{(p-d)\times d}$. The new parameterization brings a full-rank Jacobian matrix and an open parameter space in $\mathbb{R}^{d(2m+p-d)}$.

Let $\theta = (\text{vec}(B)^T, \text{vec}(C)^T)^T \in \mathbb{R}^{d(p+2m)}$ $\theta_0 = (\text{vec}(\beta)^T, \text{vec}(\nu)^T)^T$, $g(\theta) = \text{vec}(BC) \in \mathbb{R}^{2pm}$, and $g(\theta_0) = \text{vec}(\beta\nu)$. Then $\Delta = \frac{\partial}{\partial\theta}g(\theta)|_{\theta=\theta_0} = (\nu^T \otimes I_p, I_{2m} \otimes \beta)$. Based on Proposition 3.1 in Shapiro (1986) by checking all the condistion for θ_0 and discrepancy function H_d, $n\hat{H}_d \xrightarrow{D} \chi^2_k$. Here $k = 2pm - rank(\Delta)$ and $rank(\Delta) = d(2m + p - d)$, so $k = (2m - d)(p - d)$. The conclusion 2 is then proved. Also, $g(\theta)$ is one-to-one, bicontinuous, and twice continuously differentiable function. Based on Lemma A.4 in Cook and Ni (2005) and Theorem 1, we can get $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, (\Delta^T\Gamma^{-1}\Delta)^-)$,

and

$$\sqrt{n}[\text{vec}(\hat{\beta}\hat{\nu}) - \text{vec}(\beta\nu)] \xrightarrow{D} N(0, \Delta(\Delta^T\Gamma^{-1}\Delta)^-\Delta^T).$$

Finally, consistency of $\text{Span}(\hat{\beta})$ follows directly from conclusion 1. □

Proof of Theorem 3: Follow the proof in Appendix D in Cook and Ni (2005). Let $V = \Gamma_D^{-1} = \text{diag}\{\Gamma_l^{-1}\}$ and $V_n = \hat{\Gamma}_D^{-1} = \text{diag}\{\hat{\Gamma}_l^{-1}\}$. The discrepancy function $F_d(B, C, \hat{\Gamma}_D^{-1})$ can be written as

$$F_d(B, C; V_n) = [\text{vec}(\hat{\xi}) - \text{vec}(BC)]^T V_n [\text{vec}(\hat{\xi}) - \text{vec}(BC)].$$

First, the consistent property can be derived as Theorem 2. Then, $n\hat{F}_d$ is asymptotically distributed as a linear combination of independent chi-squared random variables each with one degree of freedom. The coefficient of the chi-squared variables are the eigenvalues of $Q_\Phi \Omega Q_\Phi$, where $\Omega = V^{1/2}\Gamma V^{1/2}$ and $\Phi = V^{1/2}\Delta$. The dimension of $\dim(Q_\Phi \Omega Q_\Phi)$ =$\dim(Q_\Phi)$ = $2pm - \dim(\Delta) = (p-d)(2m-d)$, which is the number of terms in linear combination. □

Proof of Lemma 1: First, we know that $\text{Span}(\hat{\beta})$ is a consistent estimator of $\sum_{j=1}^m \text{Span}\{\xi_j\}$ by Theorem 3 Part 1. Under coverage condition, $\sum_{j=1}^m \text{Span}\{\xi_j\} \subseteq \mathcal{S}_{Y|\mathbf{X}}$. Second, under the linearity condition, $\text{Span}(\hat{u}_1, \cdots, \hat{u}_d) \subseteq \mathcal{S}_{Y|\mathbf{X}}$. All the spaces: $\text{Span}(\hat{\beta})$, $\text{Span}(\hat{u}_1, \cdots, \hat{u}_d)$, and $\mathcal{S}_{Y|\mathbf{X}}$ have dimension d. Hence, $\text{Span}(\hat{\beta}) \subseteq \text{Span}(\hat{u}_1, \cdots, \hat{u}_d)$. □

Proof of Theorem 4: The Proof of this theorem is similar to that of

Theorem 3, by replacing $V = \text{diag}\{\Sigma\}$ and $\hat{V} = \text{diag}\{\hat{\Sigma}\}$. $\square$

S2 Additional Algorithm and Simulations

S2.1 Detailed Algorithm for FT-IRE

1. Choose an initial value for $B \in \mathbb{R}^{p\times d}$. One of the choices could be $\boldsymbol{e}_i = (0, ..0, 1, 0, ..0)^T$ with i^{th} place 1 and other places 0s. Alternatively, we use the spectral decomposition result from FT (Weng and Yin, 2018).

2. Fixed B, update C by minimizing $F_d(B, C; V)$. We fit linear regression $V^{1/2}\text{vec}(\hat{\xi})$ on $V^{1/2}(I_{2m} \otimes B)$, then $\text{vec}(C) = [(I_{2m} \otimes B^T)V(I_{2m} \otimes B)]^{-1}(I_{2m} \otimes B^T)V\text{vec}(\hat{\xi})$.

3. Fixed C, minimize $F_d(B, C; V)$ with respect to one column of B, subject to unit norm and orthogonal to other columns (keeping them constants). For this partial minimization problem, the quadratic discrepancy function is $F(b) = (\boldsymbol{\alpha}_k - (\boldsymbol{c}_k^T \otimes I_p)Q_{B_{(-k)}}\boldsymbol{b})^T V(\boldsymbol{\alpha}_k - (\boldsymbol{c}_k^T \otimes I_p)Q_{B_{(-k)}}\boldsymbol{b})$, where $\boldsymbol{\alpha}_k = \text{vec}(\hat{\xi} - B_{(-k)}C_{(-k)})$, $\boldsymbol{c}_k$ is kth column of C, $C_{(-k)}$ (or $B_{(-k)}$) are deleting k^{th} column from C (or B) and $Q_{B_{(-k)}}$ is orthogonal complement of $\text{Span}(B_{(-k)})$. For $k = 1, ..., d$:

   (a) Denote $B = (\boldsymbol{b}_1, ..., \boldsymbol{b}_{k-1}, \boldsymbol{b}_k, \boldsymbol{b}_{k+1}, ..., \boldsymbol{b}_d)$ and update $\hat{\boldsymbol{b}}_k = Q_{B_{(-k)}}[Q_{B_{(-k)}}(\boldsymbol{c}_k^T \otimes I_p)V(\boldsymbol{c}_k^T \otimes I_p)Q_{B_{(-k)}}]^- Q_{B_{(-k)}}(\boldsymbol{c}_k^T \otimes I_p)V\boldsymbol{\alpha}_k$, then normalize $\hat{\boldsymbol{b}}_k$ using

$\hat{\boldsymbol{b}}_k/||\hat{\boldsymbol{b}}_k||$.

(b) Update B by replace $\boldsymbol{b}_k$ with $\hat{\boldsymbol{b}}_k$ and update C like step 2.

4. Return to step 3 until $\max\{|B_{(t+1)} - B_{(t)}|^2, |C_{(t+1)} - C_{(t)}|^2\} < 10^{-6}$..

S2.2 Other simulation results

In Model 1, we consider three other different ways to calculate the precision matrix of Γ: the QL decomposition of sample covariance $\widehat{\Gamma}$ (Figure S2.1), the generalized inverse of sample covariance $\widehat{\Gamma}$ (Figure S2.2), and the generalized inverse of soft-thresholding covariance $\widetilde{\Gamma}$ (Figure S2.3). The soft-thresholding covariance (Figure 1) approach produces higher accuracy and more stable r_2 values than the sample covariance $\widehat{\Gamma}$ (Figures S2.1 and S2.2 in Section S2.2). The generalized inverse matrix results in less accurate estimates, especially for sub-optimal estimators (Figure S2.3 in Section S2.2). Thus, we use QL decomposition of soft-thresholding covariance to present our results.

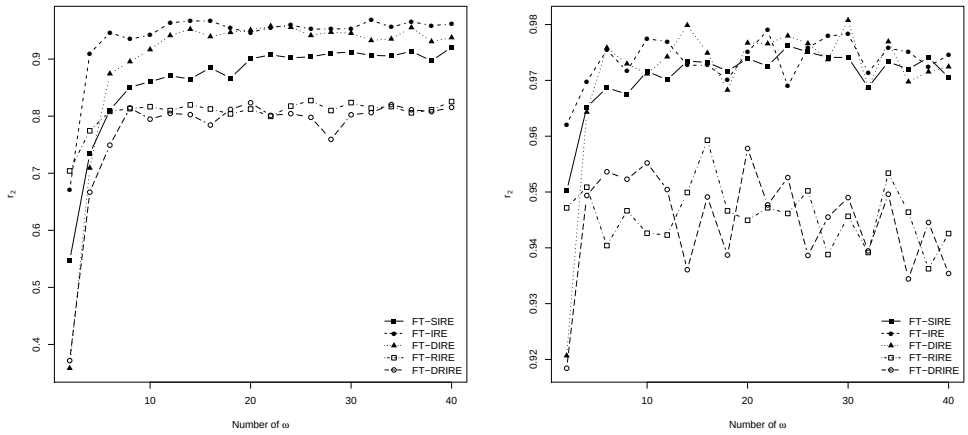
For Model 2, we also present same results when $m = 6, 10, 16$ in Figure S2.4.

In addition to the results in Model 3, we add FT-DIRE and FT-DRIRE into our analysis in Figure S2.5 and don't recommend using the degenerated estimators to detect structural dimension and predictor hypothesis

tests. Also, we compare QL decomposition of soft-thresholding covariance with two other different methods in calculating inverse matrices of Γ, that is generalized inverse matrix of sample covariance matrix (GI) and generalized inverse matrix of soft-thresholding covariance (GS), while fixing the sample size $n = 500$ and varying the number of Fourier transforms varies among $\{2, 4, ..., 40\}$ (Figures S2.6 and S2.7). Also, the results in Figures S2.6 and S2.7 support our conclusion that FT-IRE and FT-RIRE with the QL decomposition soft-thresholding covariance have the highest percentages even when m is small.

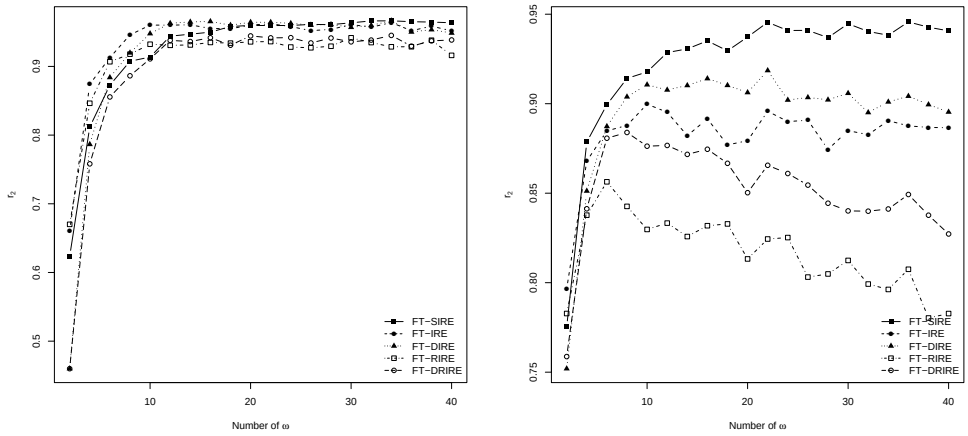
S2.3 Real data results

For the realdata analysis, Table S2.1 shows the respective estimates from SIR, IRE, FIRE, DIRE, FT-IRE, and FT-DIRE. To evaluate the accuracy, we plot six scatter plots using the first six reduced predictors ($\beta_1^T\mathbf{X}$) from these methods versus the outcome in the first and second rows of Figure S2.8, revealing that the variables found by our proposal have a stronger linear association with the response. We also present the second reduced predictors ($\beta_2^T\mathbf{X}$) versus the outcome for IRE, FIRE, and DIRE in the last row of Figure S2.8, indicating that the second reduced predictors do not provide meaningful information.



(a) Example 1.1

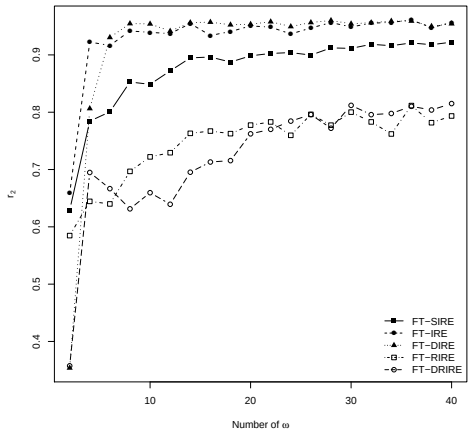
(b) Example 1.2



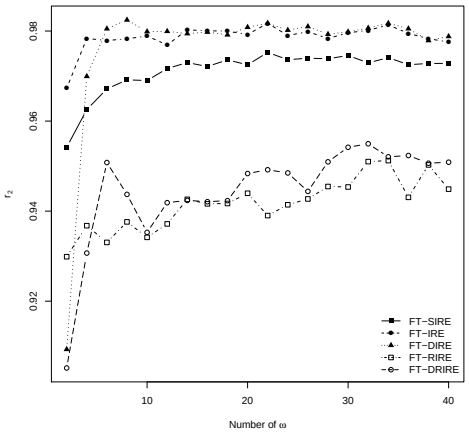
(c) Example 1.3

(d) Example 1.4

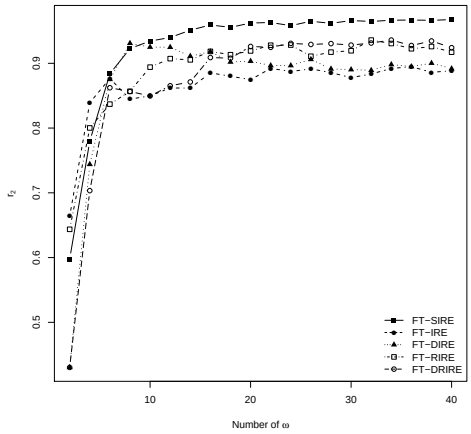
Figure S2.1: Using QL decomposition of sample covariance $\widehat{\Gamma}$: Mean values of r_2 over 100 simulated data vs. different sizes of $\boldsymbol{\omega}$: $\{2, 4, ..., 40\}$ in Model 1.



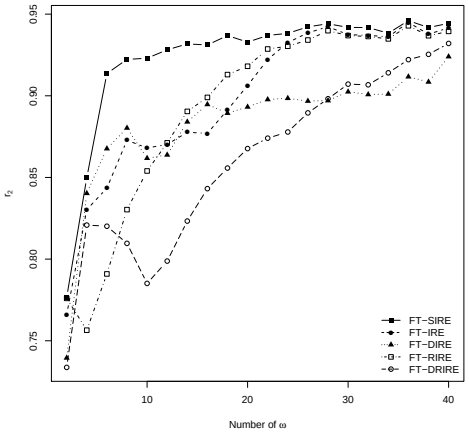
(a) Example 1.1



(b) Example 1.2

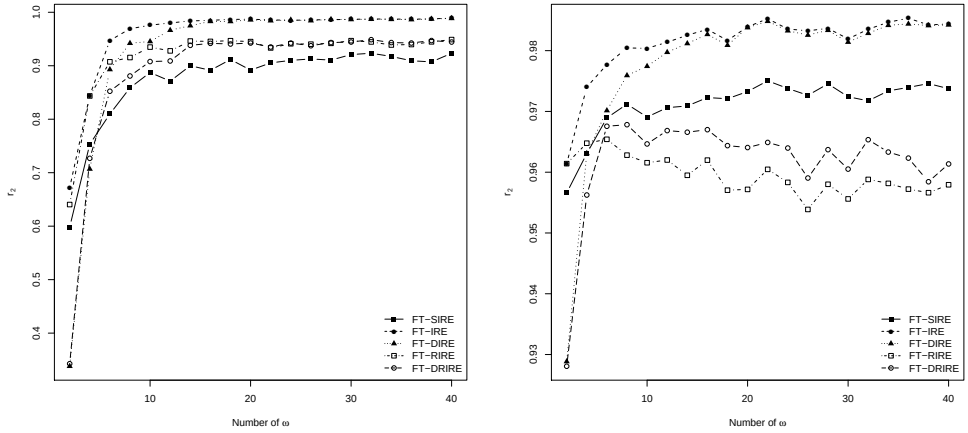


(c) Example 1.3



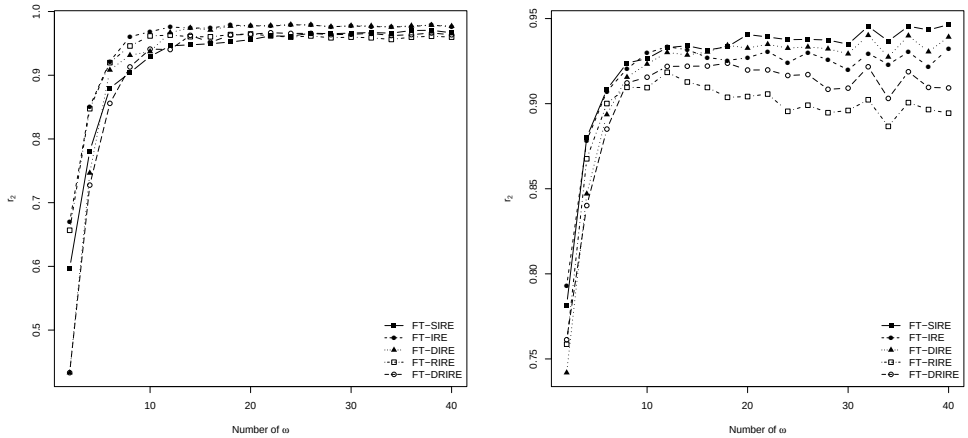
(d) Example 1.4

Figure S2.2: Using generalized inverse matrix of sample covariance matrix: Mean values of r_2 over 100 simulated data vs. different sizes of $\boldsymbol{\omega}$: $\{2, 4, ..., 40\}$ in Model 1.



(a) Example 1.1

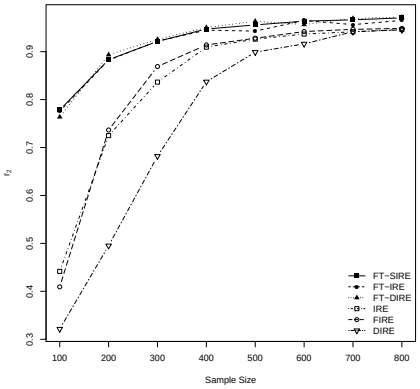
(b) Example 1.2



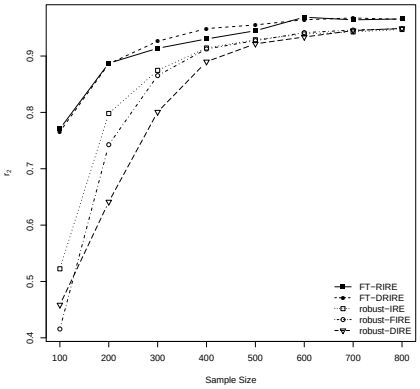
(c) Example 1.3

(d) Example 1.4

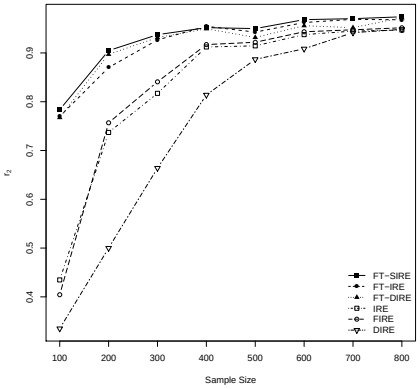
Figure S2.3: Using generalized inverse matrix of soft-thresholding covariance $\widetilde{\Gamma}$: Mean values of r_2 over 100 simulated data vs. different sizes of $\boldsymbol{\omega}$: $\{2, 4, ..., 40\}$ in Model 1.



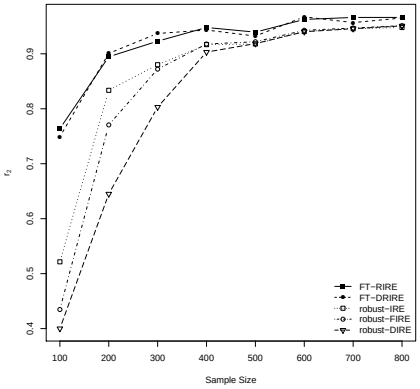
(a) Number of ω: 6



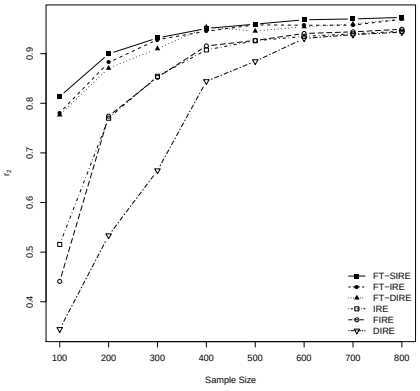
(b) Number of ω: 6



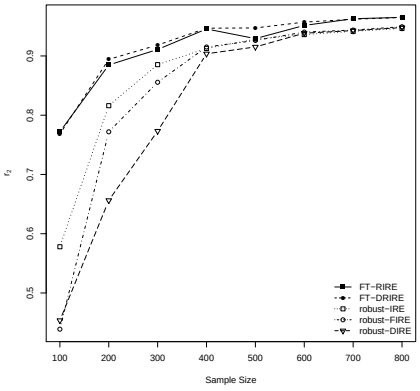
(c) Number of ω: 10



(d) Number of ω: 10

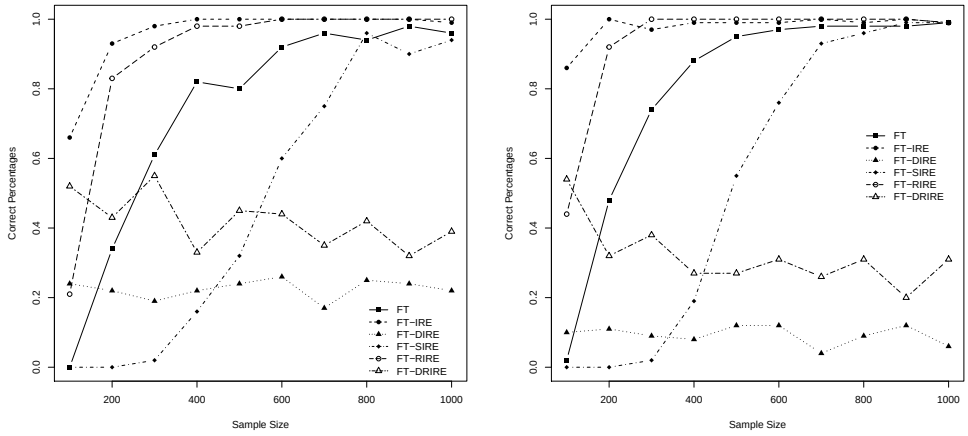


(e) Number of ω: 16



(f) Number of ω: 16

Figure S2.4: Mean values of r_2 over 100 simulated data vs. sample sizes from 100 to 800 at increments of 100 in Model 2.



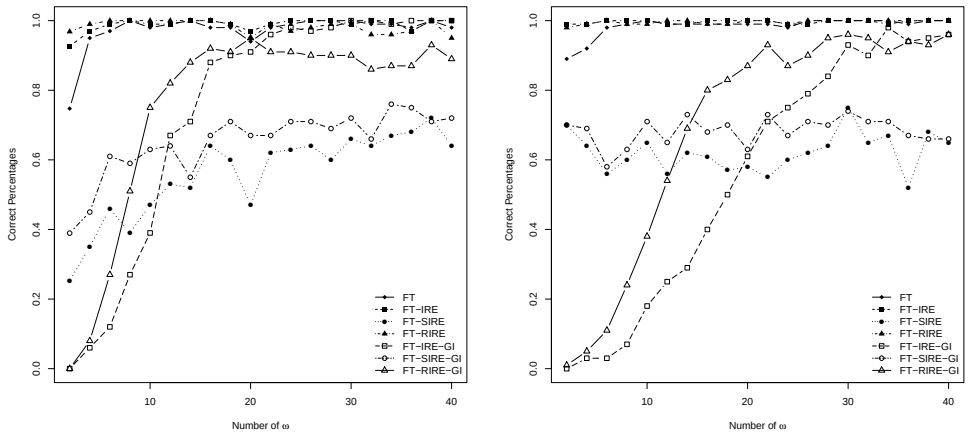
(a) Example 3.1

(b) Example 3.2

(c) Example 3.3

(d) Example 3.4

Figure S2.5: Percentages of correctly detecting dimensions ($d = 2$) over 100 simulated data vs. sample sizes n: $\{100, ..., 1000\}$ in Model 3.



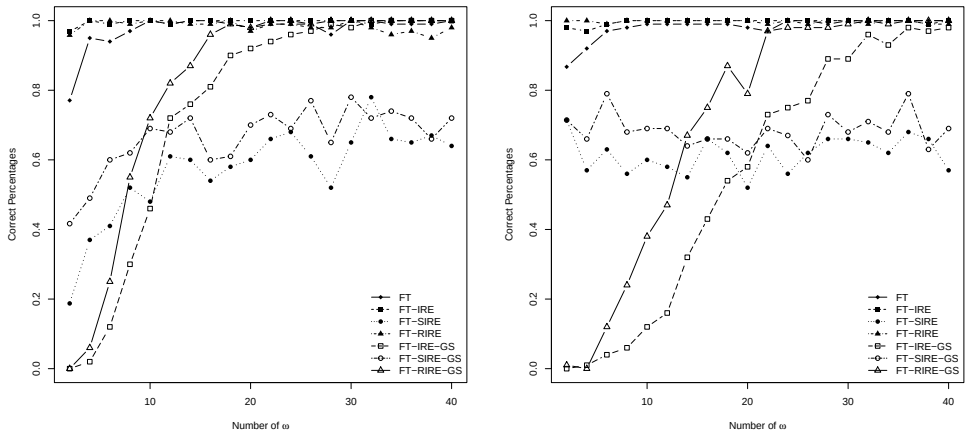
(a) Example 3.1

(b) Example 3.2

(c) Example 3.3

(d) Example 3.4

Figure S2.6: Percentages of correctly detecting dimensions ($d = 2$) over 100 simulated data vs. m: $\{2, 4, ..., 40\}$ in Model 3 using generalized inverse matrix of sample covariance matrix.



(a) Example 3.1

(b) Example 3.2

(c) Example 3.3

(d) Example 3.4

Figure S2.7: Percentages of correctly detecting dimensions ($d = 2$) over 100 simulated data vs. m: $\{2, 4, ..., 40\}$ in Model 3 using generalized inverse matrix of soft-thresholding covariance.

We also conduct similar simulations using the SIR estimate with $d = 1$, then fit a simple linear model, resulting in $\hat{\sigma}$ and $\hat{Y} = \alpha_1 X_1^*$. Again, we generate 100 data sets from the model $Y = \hat{Y} + \epsilon$ with different sample sizes 50, 100, 200, and 400, where $\epsilon \sim N(0, \hat{\sigma})$. In Table S2.2, it is not surprising that SIR is the best but Fourier transform approaches are very close to SIR, and they all have higher accuracy compared to IRE, FIRE, and DIRE.

Table S2.1: The estimation using IRE, FIRE and DIRE with $d = 2$ and SIR, FT-IRE, and FT-DIRE with $d = 1$.

Var.	IRE1	IRE2	FIRE1	FIRE2	DIRE1	DIRE2	SIR	FT-IRE	FT-DIRE
lcavol	-0.8857	-0.1424	0.7929	0.1161	0.0870	0.2186	0.5726	-0.5745	-0.6007
lweight	-0.2979	0.3025	0.3064	0.2752	0.1059	0.3851	0.4973	-0.4929	-0.5351
age	-0.0236	-0.1188	-0.0383	0.0549	-0.0141	0.0735	-0.0183	0.0175	0.0160
lbph	-0.1451	0.5337	0.1193	-0.0712	0.0884	-0.1808	0.1187	-0.0878	-0.1011
svi	-0.2232	-0.4763	0.3835	-0.6643	0.9843	-0.0004	0.6157	-0.6053	-0.5309
lcp	-0.0322	-0.0071	-0.2822	0.1245	0.0355	0.1280	-0.1559	0.1961	0.1730
gleason	-0.1757	0.5658	-0.1854	-0.6677	0.0543	-0.8656	0.0829	-0.1183	-0.1747
pgg45	0.1535	0.2054	0.0233	0.0124	0.0081	0.0107	0.0087	-0.0077	-0.0083

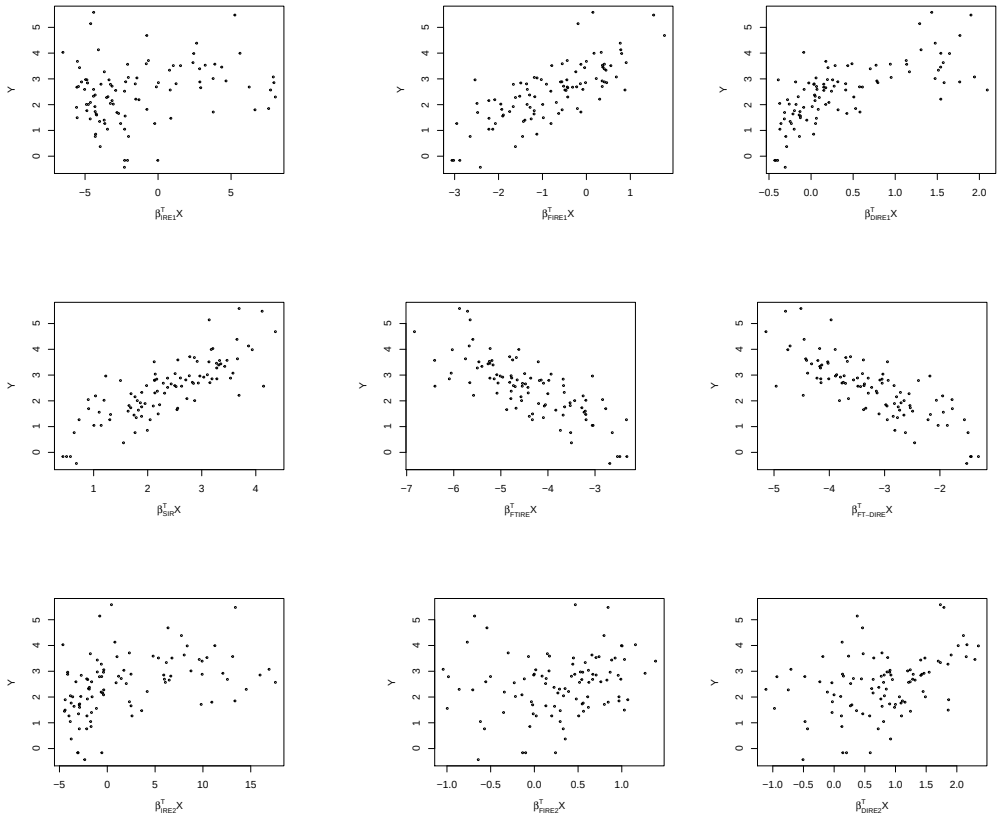


Figure S2.8: Scatter plots of $\beta_1^T \mathbf{X}$ vs. Y (the first two rows) and scatter plots of $\beta_2^T \mathbf{X}$ vs. Y (the third row).

Table S2.2: Comparing six methods using SIR estimation with $d = 1$.

n	Criteria	IRE	FIRE	DIRE	SIR	FT-IRE	FT-DIRE
50	MSE	1.0500	1.0171	1.0027	0.6756	0.6800	0.6680
	r_2	0.4616	0.4314	0.4421	0.8755	0.8627	0.8701
	Norm	1.1986	1.209	1.1999	0.6228	0.6235	0.6259
100	MSE	1.0109	0.944	0.8591	0.6735	0.6755	0.6720
	r_2	0.4671	0.4864	0.6475	0.9248	0.9097	0.9092
	Norm	1.1930	1.1603	0.9952	0.4959	0.5275	0.5270
200	MSE	1.0150	0.7662	0.7646	0.6807	0.6901	0.6839
	r_2	0.5251	0.7106	0.7544	0.9653	0.9319	0.9508
	Norm	1.1408	0.8899	0.8391	0.3447	0.4489	0.4051
400	MSE	1.0281	0.7249	0.7275	0.6831	0.6963	0.6884
	r_2	0.4635	0.8123	0.8522	0.9830	0.9511	0.9605
	Norm	1.1962	0.7804	0.6959	0.2391	0.3663	0.3519

References

Cook, R. D. and L. Ni (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association 100*(470), 410–428.

Li, B., R. D. Cook, and F. Chiaromonte (2003). Dimension reduction for the conditional mean in regressions with categorical predictors. *The Annals of Statistics 31*(5), 1636–1668.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association 81*(393), 142–149.

Weng, J. and X. Yin (2018). Fourier transform approach for inverse dimension reduction method. *Journal of Nonparametric Statistics 30*(4), 1049–1071.