Appendix A Proof of Theorems

Proof of Theorem 1

Proof. Let $\boldsymbol{\beta}_{S_n}$ be a $k_n \times 1$ vector in which S_n entries are nonzero, $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_j, j = 1, \ldots, k_n\}$ be the estimated sample wavelet coefficient magnitude, and ρ_{k_n} be the smallest eigenvalue of $\boldsymbol{\Sigma}_{k_n} = \frac{1}{n}\boldsymbol{Z}_{k_n}^T \boldsymbol{Z}_{k_n}$. Without loss of generality, we assume the first S_n entries in $\boldsymbol{\beta}_{S_n}$ are nonzero, and assume columns of $\boldsymbol{Z}_{k_n}$ are standardized such that each column has mean of 0 and standard deviation of 1, and the first S_n elements of $\boldsymbol{\theta}$ are nonzero. By Parseval's theorem, we have

$$||\hat{\eta}_n - \eta||_{L^2}^2 = \| \hat{\boldsymbol{\beta}}_{k_n} - \boldsymbol{\beta}_{S_n}\|^2 + \sum_{j=S_n+1}^{k_n} \beta_j^2 + \sum_{j=k_n+1}^{N_n} \beta_j^2 + \sum_{j=N_n+1}^{\infty} \beta_j^2. \tag{A.1}$$

The first term on right hand side of equation (A.1) stands for model estimation error, the second term is due to thresholding error, the third term is due to screening error, and the fourth term is due to finite sampling error which depends on how densely we sample the functional predictor. By assumption (a4) and Theorem 9.5 of Mallat (2008),

$$\sum_{j=N_n+1}^{\infty} \beta_j^2 = o(N_n^{-2q}). \tag{A.2}$$

By assumption (a5) and Theorem 9.10 of Mallat (2008),

$$\sum_{j=S_n+1}^{k_n} \beta_j^2 + \sum_{j=k_n+1}^{N_n} \beta_j^2 = \sum_{j=S_n+1}^{N_n} \beta_j^2 = o(S_n^{1-2/r}). \tag{A.3}$$

We show the convergence rate of $\| \hat{\boldsymbol{\beta}}_{k_n} - \boldsymbol{\beta}_{S_n}\|^2$ below. Let $l(\boldsymbol{\beta}) = 1/n||Y - \boldsymbol{Z}_{k_n}^T \boldsymbol{\beta}||^2 + \sum_{j=1}^{k_n} \lambda_n \hat{\theta}_j^{-1} |\beta_j|$, $\delta = ||\hat{\boldsymbol{\beta}}_{k_n} - \boldsymbol{\beta}_{S_n}||$, and $\hat{\boldsymbol{\beta}}_{k_n} - \boldsymbol{\beta}_{S_n} = \delta \mathbf{u}$ with $||\mathbf{u}|| = 1$. Given equation (4), we have

$$l(\hat{\boldsymbol{\beta}}_{k_n}) - l(\boldsymbol{\beta}_{S_n}) = -2\delta n^{-1} \boldsymbol{\varepsilon}^{*T} \boldsymbol{Z}_{k_n}^T \mathbf{u} + \delta^2 n^{-1} \mathbf{u}^T \boldsymbol{Z}_{k_n}^T \boldsymbol{Z}_{k_n} \mathbf{u} + \sum_{j=1}^{k_n} \lambda_n \hat{\theta}_j^{-1} (|\hat{\beta}_j| - |\beta_j|) \leq 0. \tag{A.4}$$

We know $\sum_{j=1}^{k_n} \lambda_n \hat{\theta}_j^{-1}(|\hat{\beta}_j| - |\beta_j|) \geq \sum_{j \in H_n} \lambda_n \hat{\theta}_j^{-1}(|\hat{\beta}_j| - |\beta_j|)$. By reverse triangle inequality and Cauchy-Schwarz inequality, we have

$$\sum_{j \in H_n} \lambda_n \hat{\theta}_j^{-1}(|\hat{\beta}_j| - |\beta_j|) \geq -\sum_{j \in H_n} \lambda_n \hat{\theta}_j^{-1}(|\hat{\beta}_j - \beta_j|) \geq -\delta \sqrt{\sum_{j \in H_n} (\lambda_n \hat{\theta}_j^{-1})^2}. \tag{A.5}$$

Combine (A.4) and (A.5), we have

$$\delta \rho_{k_n} \leq \delta \mathbf{u}^T \boldsymbol{\Sigma}_{k_n} \mathbf{u} \leq 2||n^{-1} \boldsymbol{Z}_{H_n}^T \boldsymbol{\varepsilon}^*|| + \sqrt{\sum_{j \in H_n} (\lambda_n \hat{\theta}_j^{-1})^2}. \tag{A.6}$$

We next show the convergence rate of $||n^{-1} \boldsymbol{Z}_{k_n}^T \boldsymbol{\varepsilon}^*||$ in equation (A.6). Given equation (4), a constant C, and $\xi_i^2 = (\sum_{i=k_n+1}^{N_n} z_{ij}\beta_j)^2$,

$$\begin{aligned}
E(||n^{-1/2}\boldsymbol{Z}_{k_n}^T\boldsymbol{\varepsilon}^*||^2) &= E(n^{-1}\boldsymbol{\varepsilon}^T\boldsymbol{Z}_{k_n}^T\boldsymbol{Z}_{k_n}\boldsymbol{\varepsilon}) + n^{-1}\boldsymbol{\xi}^T\boldsymbol{Z}_{k_n}^T\boldsymbol{Z}_{k_n}\boldsymbol{\xi} \\
&\leq \sigma^2 tr(\boldsymbol{\Sigma}_{k_n}) + C/n\sum_{i=1}^{n}\xi_i^2 \\
&\leq \sigma^2 k_n + 1/n\sum_{i=1}^{n}\left(\sum_{j=k_n+1}^{N_n} z_{ij}^2\right)\left(\sum_{j=k_n+1}^{N_n}\beta_j^2\right)
\end{aligned}$$

By assumption (a4), $E(||n^{-1/2}\boldsymbol{Z}_{k_n}^T\boldsymbol{\varepsilon}^*||^2) = O(k_n) + o(k_n^{1-2/r})$. By Markov's inequality,

$$||n^{-1}\boldsymbol{Z}_{H_n}^T\boldsymbol{\varepsilon}^*|| = O_p((k_n/n)^{1/2}) + o_p(k_n^{1/2-1/r}/\sqrt{n}). \tag{A.7}$$

By assumption (a2), we have $\sqrt{\sum_{j\in H_n}(\lambda_n\hat{\theta}_j^{-1})^2} = o_p(n^{-1/2})$. Combining this with (A.7) and (A.6), we have

$$\delta = \|\ \hat{\beta}_{k_n} - \beta_{S_n}\|\ = O_p\left(\frac{k_n^{1/2}}{n^{1/2}\rho_{k_n}}\right). \tag{A.8}$$

By (A.2), (A.3), and (A.8), this implies

$$\|\ \hat{\eta}_n - \eta\|\ _{L^2}^2 = O_p\left(\frac{k_n}{n\rho_{k_n}^2}\right) + o(k_n^{1-2/r}) + o(N_n^{-2q})$$

□

Proof of Theorem 2

Proof. Following Johnstone and Lu (2009), we assume, without loss of generality, that wavelet coefficient population magnitude $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_N > 0$. We also assume the coefficient of variation for each wavelet coefficient is bounded by a constant C_0 (i.e., $C_h = \sigma_h/\theta_h, 0 < C_h < C_1, h = 1, 2, \ldots, N$), where $\sigma_h^2, h = 1, \ldots, N$ is the population variance of wavelet coefficient. Let γ_n be a suitably chosen small positive number, and d be a suitably chosen constant. Let $z_h \sim N(\mu_h, \sigma_h^2), \theta_h = |\mu_h|\ \forall h$, and $Z \sim N(0, 1)$. For any fixed constant t and $l \in M$,

$$\hat{\theta}_h \leq t \text{ for } h \geq k, h \neq l \text{ and } \hat{\theta}_l \geq t \Rightarrow \hat{\theta}_l \geq \hat{\theta}_{(k)}.$$

If we let $t = \theta_k + \gamma_n$ and $\theta_l \geq d \times t$ with $d > 1$, we have

$$\begin{aligned}
P(\hat{\theta}_l < \hat{\theta}_{(k)}) &\leq \sum_{h \geq k} P(\hat{\theta}_h > t) + P(\hat{\theta}_l < t) = \sum_{h \geq k} \left\{ P\left(Z > \frac{\sqrt{n}(t - \theta_h)}{\sigma_h}\right) + P\left(Z < \frac{\sqrt{n}(-t - \theta_h)}{\sigma_h}\right) \right\} \\
&\quad + P\left(\frac{\sqrt{n}(-t - \theta_l)}{\sigma_l} < Z < \frac{\sqrt{n}(t - \theta_l)}{\sigma_l}\right) \\
&= \sum_{h \geq k} \left\{ P\left(Z > \frac{\sqrt{n}(t/\theta_h - 1)}{\sigma_h/\theta_h}\right) + P\left(Z > \frac{\sqrt{n}(t/\theta_h + 1)}{\sigma_h/\theta_h}\right) \right\} \\
&\quad + P\left(Z > \frac{\sqrt{n}(1 - t/\theta_l)}{\sigma_l/\theta_l}\right) - P\left(Z > \frac{\sqrt{n}(1 + t/\theta_l)}{\sigma_l/\theta_l}\right) \\
&\leq \sum_{h \geq k} \left\{ P\left(Z > \frac{\sqrt{n}(\gamma_n/\theta_k)}{C_0}\right) + P\left(Z > \frac{\sqrt{n}(\gamma_n/\theta_k + 2)}{C_0}\right) \right\} + P\left(Z > \frac{\sqrt{n}(1 - 1/d)}{C_0}\right) \\
&= (N - k + 1)\,\Phi\left(-\frac{\sqrt{n}\gamma_n/\theta_k}{C_0}\right) + (N - k + 1)\,\Phi\left(-\frac{\sqrt{n}(2 + \gamma_n/\theta_k)}{C_0}\right) + \Phi\left(\frac{\sqrt{n}(1/d - 1)}{C_0}\right)
\end{aligned}$$

The bound $P(FE) \leq (N - k + 1)\,\Phi\left(-\frac{\sqrt{n}}{C_0 b}\right) + (N - k + 1)\,\Phi\left(-\frac{\sqrt{n}(2b+1)}{C_0 b}\right) + \Phi\left(\frac{\sqrt{n}(1-d)}{dC_0}\right)$ follows from $\gamma_n = \sqrt{log(n)/n}$, $\theta_k = b\gamma_n$ with $b > 0$, and a suitably chosen constant $d > 1$. $\square$

References

Johnstone, I. and Lu, A. (2009). On Consistency and Sparsity for Principal Components Analysis in High Dimensions. *Journal of American Statistical Association* 104 682-693.

Mallat, S. (2008). A Wavelet Tour of Signal Processing: The Sparse Way. *Academic Press*