We approximate Q_0 by Haar wavelets

$$\mathcal{Q} = \Big\{\theta_{(0),0}h_0(X)+\sum_{lk}\theta_{(0),lk}h_{lk}(X)+\Big(\theta_{(0),1}h_0(X)+\sum_{lk}\theta_{(1),lk}h_{lk}(X)\Big)A : \theta_{\cdot,\cdot} \in \mathbb{R}\Big\},$$

where $h_0(x) = 1_{x\in[0,1]}$ and $h_{lk}(x) = 2^{l/2}\big(1_{2^l x\in[k+1/2,k+1)} - 1_{2^l x\in[k,k+1/2)}\big)$ for $l = 0, \ldots, \bar{l}_n$. We choose $\bar{l}_n = \lfloor 3\log_2 n/4\rfloor - 2$. For a given l and sample $(X_i, A_i, R_i)_{i=1}^n$, k takes integer values from $\lfloor 2^l \min_i X_i\rfloor$ to $\lceil 2^l \max_i X_i\rceil - 1$. Then $J_n = 2^{\lfloor 3\log_2 n/4\rfloor} \le n^{3/4}$.

Remark:

In example 4, we allow the number of basis functions J_n to increase with n. The corresponding theoretical result can be obtained by combining Theorem 3.1 and Theorem A.1. Below we demonstrate the validation of the assumptions used in the theorems.

Theorem 3.1 requires that the randomization probability $p(a|x) \ge S^{-1}$ for a positive constant for all (x, a) pairs and the margin condition (3.3) or (3.6) holds. According the generative model, we have that $p(a|x) = 1/2$ and condition (3.6) holds.

Theorem A.1 requires Assumptions A.1 - Assumptions A.4 hold and Θ_n defined in (A.4) is non-empty. Since we consider normal error terms, Assumption A.1 holds. Note that the basis functions used in Haar wavelet are orthogonal. It is also easy to verify that Assumptions A.3 and A.4 hold with $\beta_n = 1$ and Assumption A.2 holds with $U_n = n^{3/8}/2$ and $\eta_{1,n} \le constant + constant \times \|\boldsymbol{\theta}_n^*\|_0$ (since each $|\phi_j \boldsymbol{\theta}_{n,j}^*| = |\phi_j E(\phi_j R)| \le constant \times |\phi_j| E|\phi_j| \le O(1)$). Since Q_0 is piece-wise constant, we can also verify that $\|\boldsymbol{\theta}_n^*\|_0 \le O(\log n)$. Thus for sufficiently large n, Θ_n is non-empty and (A.6) holds. The RHS of (A.5) converges to zero as $n \to \infty$. $\square$

ACKNOWLEDGEMENTS

The authors thank Martin Keller and the investigators of the Nefazodone-CBASP trial for use of their data. The authors also thank John Rush, MD, for the technical support and Bristol-Myers Squibb for helping fund the trial. The authors thank valuable comments from Eric B. Laber and Peng Zhang.

SUPPLEMENTARY MATERIAL

S.1: The overfitting problem

(). This section discusses the problem with over-fitting due to the potentially large number of pretreatment variables (and/or complex approximation space for Q_0) mentioned in Section 4.

S.2: Some modifications of the l_1-PLS estimator $\hat{\boldsymbol{\theta}}_n$

(). This section provides modifications of the l_1-PLS estimator $\hat{\boldsymbol{\theta}}_n$ when some coefficients are not penalized and discusses how to obtain results similar to inequality (A.7) in this case.

S.3: Extra simulation examples

(). This section provides extra four simulation examples based on data from the Nefazodone-CBASP trial [13].

S.3: Proofs of Lemmas A.1 - A.5

(). This section provides proofs of Lemmas A.1 - A.5.

REFERENCES

[1] Bartlett, P. L., Jordan, M. L. and McAuliffe, J. D. (2006). Convexity, classification, and risk bounds. *Journal of the American Statistical Association* **135(3)**, 311–334.

[2] Bartlett, P. L. (2008). Fast rates for estimation error and oracle inequalities for model selection. *Econometric Theory,* **24(2),** 545–552.

[3] Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics,* **37(4),** 1705–1732.

[4] Bunea, F., Tsybakov, A. B. and Wegkamp, M. H. (2007). Sparsity oracle inequalities for the Lasso *Electronic Journal of Statistics,* **1**, 169–194.

[5] Cai, T., Tian, L., Lloyd-Jones, D. M. and Wei, L. J. (2008a). Evaluating Subject-level Incremental Values of New Markers for Risk Classification Rule. *Harvard University Biostatistics Working Paper Series. Working Paper 91.*

[6] Cai, T., Tian, L., Uno, H., Solomon, S. D. and Wei, L. J. (2008b). Calibrating Parametric Subject-specific Risk Estimation. *Harvard University Biostatistics Working Paper Series. Working Paper 92.*

[7] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.),* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

[8] Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81(3)**, 425–455.

[9] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and it oracle properties. *Journal of the American Statistical Association,* **96,** 1348–1360.

[10] Feldstein, M. L., Savlov, E. D. and Hilf, R. (1978). A statistical model for predicting response of breast cancer patients to cytotoxic chemotherapy. *Cancer Research,* **38(8),** 2544–2548.

[11] Insel, T. R. (2009). Translating scientific opportunity into public health impact: a strategic plan for research on mental illness. *Archives of General Psychiatry,* **66(2)**, 128–33.

[12] Ishigooka, J., Murasaki, M., Miura, S. and The Olanzapine Late-Phase II Study Group (2001). Olanzapine optimal dose: Results of an open-label multicenter study in schizophrenic patients. *Psychiatry and Clinical Neurosciences,* **54(4),** 467–478.

[13] Keller, M. B., McCullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J., Markowitz, J. C., Nemeroff. C. B., Russell, J. M., Thase, M. E., Trivedi, M. H. and Zajecka, J. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the