Supplemental Material to "Quantile-Adaptive Variable Screening in Ultra-High Dimensional Varying Coefficient Models" *

Junying Zhang^{1,2}, Riquan Zhang^{1,3†} and Zhiping Lu¹

1. School of Finance and Statistics

East China Normal University, Shanghai, 200241, P. R. China

2. Department of Mathematics

Taiyuan University of Technology, Taiyuan, 030024, P. R. China

3. Department of Mathematics

Shanxi Datong University, Datong, 037009, P. R. China

May 19, 2015

The proof of Proposition. There exist $b_0, b_1 \in R^{d_n}$ depending on $\{b_{j_0}(U), b_j(U)\}$ and positive constants c_2, c_3 such that

$$\sup_U |b_{j_0}(U) - \mathbf{B}^T(U)b_0| \leq c_2 d_n^{-l}, \quad (0.1)$$

$$\sup_U |b_j(U)X_j - \mathbf{B}^T(U)b_1X_j| \leq \sup_U |b_j(U) - \mathbf{B}^T(U)b_1||X_j| \leq c_3 d_n^{-l}, a.s. \quad (0.2)$$

which extends the results in Shumaker[3] on condition(vii).

From above (0.1) and (0.2), there exists $c_4 > 0$, such that:

$$\begin{aligned} |(b_{j_0}(U) + b_j(U)X_j) - [\mathbf{B}^T(U), \mathbf{B}^T(U)X_j][b_0^T, b_1^T]^T| \\ &= |(b_{j_0}(U) + b_j(U)X_j) - H_j[b_0^T, b_1^T]^T| \\ &\leq |b_{j_0}(U) - \mathbf{B}^T(U)b_0 + b_j(U)X_j - \mathbf{B}^T(U)b_1X_j| \\ &\leq |b_{j_0}(U) - \mathbf{B}^T(U)b_0| + |b_{j_1}(U)X_j - \mathbf{B}^T(U)b_1||X_j| \\ &\leq c_4 d_n^{-l}. \end{aligned} \quad (0.3)$$

*The research was supported in part by National Natural Science Foundation of China (11171112, 11201190, 11101158), Doctoral Fund of Ministry of Education of China (20130076110004), Program of Shanghai Subject Chief Scientist (14XD1401600) and the 111 Project of China (B14019).

†Correspondence address: School of Finance and Statistics, East China Normal University, Shanghai, 200241, P. R. China. E-mail: zhangriquan@163.com.

The quantile loss function $\rho_\tau(a) = a[\tau - I(a < 0)] = |a|/2 + (\tau - 1/2)a$ satisfies the property: $\rho_\tau(a_1 + a_2) \leq \rho_\tau(a_1) + \rho_\tau(a_2)$, $\forall a_1, a_2 \in R$. For any $b \in R^{2d_n}$, we have

$$\begin{aligned} E_{(Y|X_j, U)} [\rho_\tau(Y - Q_\tau(Y|X_j, U))] \\ \leq E_{(Y|X_j, U)} [\rho_\tau(Y - H_j b)] \\ \leq E_{(Y|X_j, U)} [\rho_\tau(Y - Q_\alpha(Y|X_j, U)) + \\ E_{(Y|X_j, U)} [\rho_\tau(Q(Y|X_j, U) - H_j [b_0^T, b_1^T]^T)], \end{aligned}$$

We define $\tilde{b}_j = [E\tilde{H}_j]^{-1}EY$, so we have

$$\begin{aligned} E_{(Y|X_j, U)} [\rho_\tau(Y - Q_\tau(Y|X_j, U))] \\ \leq E_{(Y|X_j, U)} [\rho_\tau(Y - H_j \tilde{b}_j)] \\ \leq E_{(Y|X_j, U)} [\rho_\tau(Y - Q_\tau(Y|X_j, U))] + \\ E_{(Y|X_j, U)} [\rho_\tau(Q(Y|X_j, U) - H_j [b_0^T, b_1^T]^T)]. \end{aligned}$$

Above all, we have the result of

$$\begin{aligned} E_{(Y|X_j, U)} [\rho_\tau(Y - H_j \tilde{b}_j)] \\ - E_{(Y|X_j, U)} [\rho_\tau(Y - Q_\tau(Y|X_j, U))] \\ \leq c_4 d_n^{-l}. \end{aligned} \tag{0.4}$$

From the identity of Knight (page 758)[4], the left-hand side of above inequality can be written as:

$$\left| \int_0^{H_j \tilde{b}_j - Q_\tau(Y|X_j, U)} [F_{(Y|X_j, U)}(Q_\tau(Y|X_j, U) + t) - F_{(Y|X_j, U)}(Q_\tau(Y|X_j, U))] dt \right|$$

From condition (ii) and inequality (0.3) gets that $H_j \tilde{b}_j - Q_\tau(Y|X_j, U) = o(1)$, as $d_n \rightarrow \infty$, uniformly in U .

We know from Theorem 1 of Angrist, Chernozhukov and Fernandez[1] that

$$\tilde{b}_j = \arg \min_b E[\omega(X_j, U, b)(Q_\tau(Y|X_j, U) - H_j b)^2].$$

Let $\eta_j = Y - Q_\tau(Y|X_j, U)$,

$$\omega(X_j, U, b) = \int_0^1 (1 - u) f_{(\eta_j|X_j, U)}(u[H_j b - Q_\tau(Y|X_j, U)]) du.$$

For the $\bar{b} = [b_0^T, b_1^T]^T$ above. we have

$$E[\omega(X_j, U, \tilde{b})(Q_\tau(Y|X_j, U) - H_j \tilde{b}_j)^2] \leq E[\omega(X_j, U, \bar{b})(Q_\tau(Y|X_j, U) - H_j \bar{b})^2].$$

From above and condition (iii), it is known that there exist some $0 < v_1 < v_2 < \infty$ such that

$$v_1 E[Q_\tau(Y|X_j, U) - H_j \tilde{b}_j]^2 \leq v_2 E[(Q_\tau(Y|X_j, U) - H_j \bar{b})^2].$$

Thus there exists a positive constant c_5 such that

$$E[Q_\tau(Y|X_j, U) - H_j \tilde{b}]^2 \leq c_5 d_n^{-2l}, \quad \forall j.$$

As $\tilde{f}_j = H_j \tilde{b} - Q_\tau(Y|U)$. It follows that

$$E[Q_\tau(Y|X_j, U) - Q_\tau(Y|U) - \tilde{f}_j]^2 \leq c_5 d_n^{-2l}, \quad \forall j.$$

Next we have that

$$\begin{aligned} \min_{j \in \mathcal{M}_*} \|\tilde{f}_j\|^2 &\geq 1/4 \min_{j \in \mathcal{M}_*} E[Q_\tau(Y|X_j, U) - Q_\tau(Y|U)]^2 \\ &\quad - 1/2 \max_{j \in \mathcal{M}_*} E[Q_\tau(Y|X_j, U) - Q_\tau(Y|U) - \tilde{f}_j]^2 \\ &\geq 1/4 c_1 n^{-\iota} - 1/2 c_5 d_n^{-2l} \geq c_6 n^{-\iota}, \end{aligned} \quad (0.5)$$

for all n sufficiently large and some $c_6 > 0$ from any constant a and b , the following inequality holds: $a^2 \geq \frac{1}{4}b^2 - \frac{1}{2}(a - b)^2$, and condition (iv). $\square$

The proof of Theorem 1. We prove part (1) firstly. Consider the $\|\hat{f}_j\|^2 - \|\tilde{f}_j\|^2$ and obtain its exponential tail probability bound.

$$\|\hat{f}_j\|_n^2 - \|\tilde{f}_j\|^2 = S_1 + S_2 + S_3,$$

where

$$\begin{aligned} S_1 &= \frac{1}{n} \hat{b}_j^T (H_{nj}^T H_{nj}) \hat{b}_j - E \tilde{b}_j^T E(H_j^T H_j) E \tilde{b}_j, \\ S_2 &= (F_{Y,n,U}^{-1}(\tau))^2 - Q_\tau^2(Y|U) \end{aligned}$$

and

$$S_3 = 2(Q_\tau(Y|U) E H_j \tilde{b}_j - \frac{1}{n} F_{(Y,n)}^{-1}(\tau) \sum_{i=1}^n H_{ji} \hat{b}_j),$$

where $H_{ji} = (B_1(U_i), \dots, B_{d_n}(U_i), B_1(U_i)X_{ji}, \dots, B_{d_n}(U_i)X_{ji})$.

Note that

$$\begin{aligned} S_1 &= (\hat{b}_j - E \tilde{b}_j)^T \left(\frac{1}{n} H_{nj}^T H_{nj} \right) (\hat{b}_j - E \tilde{b}_j) \\ &\quad + E \tilde{b}_j^T \left(\frac{1}{n} H_{nj}^T H_{nj} - E(H_j^T H_j) \right) E \tilde{b}_j \\ &\quad + 2(\hat{b}_j - E \tilde{b}_j)^T \left(\frac{1}{n} H_{nj}^T H_{nj} \right) E \tilde{b}_j \\ &= S_{11} + S_{12} + S_{13}, \end{aligned}$$

where S_{11}, S_{12}, S_{13} denote the middle three terms respectively.

We bound S_{11} . Note that

$$|S_{11}| \leq \left\| \frac{1}{n} H_{nj}^T H_{nj} \right\| \cdot \|\hat{b}_j - E\tilde{b}_j\|_2^2.$$

By Lemma 3.2 in He, Wang and Hong[5], for any $t_0 > 0$ and some positive constants c_7, c_8 , such that

$$\begin{aligned} P(\|\hat{b}_j - E\tilde{b}_j\|_2^2 \geq t_0 d_n^{-1} n^{-2\iota}) \\ \leq 2s_n \exp(-c_7 n^{1-4\iota}) + s_n \exp(-c_8 d_n^{-2} n^{1-2\iota}). \end{aligned} \quad (0.6)$$

According to the Lemma 7 in Fan, Ma and Dai[2], for some positive constants $t_1 \in (0, 1), t_2$ such that

$$P\left(\left\| \frac{1}{n} H_{nj}^T H_{nj} \right\| \geq t_1 d_n\right) \leq 6d_n^2 \exp(-t_2 d_n^{-3} n). \quad (0.7)$$

From the above two inequalities and based on union bound of probability, we have

$$\begin{aligned} P(|S_{11}| > t_0 t_1 n^{-2\iota}) \leq & 2s_n \exp(-c_7 n^{1-4\iota}) + \\ & s_n \exp(-c_8 d_n^{-2} n^{1-2\iota}) + 6d_n^2 \exp(-t_2 d_n^{-3} n). \end{aligned} \quad (0.8)$$

We next bound S_{12} . Note that

$$|S_{12}| \leq \left\| \frac{1}{n} H_{nj}^T H_{nj} - E(H_j^T H_j) \right\| \cdot \|E\tilde{b}_j\|_2^2. \quad (0.9)$$

From the Lemma 7 in Fan, Ma and Dai[2], for any $t_3 > 0$ we have

$$P\left(\left\| \frac{1}{n} H_{nj}^T H_{nj} - E(H_j^T H_j) \right\| \geq d_n t_3 n^{-1}\right) \leq 6d_n^2 \exp\{-t_3 d_n n^{-1}\}$$

and for some $M > 0$,

$$\|E\tilde{b}_j\|_2^2 = \|EB_i(U)E(Y)\|_2^2 + \|X_j EB_i(U)E(Y)\|_2^2 \leq M. \quad (0.10)$$

It follows above terms and the union bound of probability that

$$P(|S_{12}| \geq d_n M t_3 / n) \leq 6d_n^2 \exp\left(-\frac{t_3}{t_4 d_n^{-1} n + t_5 t_3}\right). \quad (0.11)$$

To bound S_{13} , note that

$$|S_{13}| \leq 2\|\hat{b}_j - E\tilde{b}_j\|_2 \left\| \frac{1}{n} H_{nj}^T H_{nj} \right\| \|E\tilde{b}_j\|_2$$

It from (0.4), (0.5) and (0.8) and the union bound of probability that

$$\begin{aligned} P(|S_{13}| \geq 2t_0^{\frac{1}{2}} t_1 M^{-\frac{1}{2}} d_n^{\frac{1}{2}} n^{-\frac{\iota}{2}}) \\ \leq 2s_n \exp(-c_7 n^{(1-4\iota)}) + s_n \exp(-c_8 d_n^{-2} n^{1-2\iota}) + 6d_n^2 \exp(-t_2 d_n^{-3} n). \end{aligned} \quad (0.12)$$

Thus combining (0.6), (0.9) and (0.10),

$$\begin{aligned}
P(|S_1| \geq t_0 t_1 n^{-2\iota} + d_n M t_3 / n + 2t_0^{\frac{1}{2}} t_1 M^{-\frac{1}{2}} d_n^{\frac{1}{2}} n^{-\iota}) \\
\leq 4s_n \exp(-c_7 n^{1-4\iota}) + 2d_n \exp(-c_8 d_n^{-2} n^{1-2\iota}) \\
+ 12d_n^2 \exp(-t_2 d_n^{-3} n) + 6d_n^2 \exp\{-t_3 d_n^{-2} n^{-\iota}\}.
\end{aligned} \tag{0.13}$$

Similarly, we can get the bound of S_2, S_3 , for positive constant t_6, t_7 ,

$$P(\|S_2\| > t_6 t_0^{1/2} d_n^{-1/2} n^{-\iota} + t_7 n^{-\iota}) \leq 2s_n \exp(-c_7 n^{1-4\iota}) + s_n \exp(-c_8 d_n^{-2} n^{1-2\iota}).$$

and $\|S_3\| = O(n^{-\iota})$.

Let $t_0 t_1 n^{-2\iota} + d_n M t_3 / n + 2t_0^{\frac{1}{2}} t_1 M^{-\frac{1}{2}} d_n^{\frac{1}{2}} n^{-\frac{\iota}{2}} + t_6 t_0^{1/2} d_n^{-1/2} n^{-\iota} + t_7 n^{-\iota} = h_1 d_n n^{-\iota}$, for any given $h_1 > 0$, for $h_2 > 0, h_3 > 0$, we have

$$\begin{aligned}
P(|\|\hat{f}_j\|_n^2 - \|\tilde{f}_j\|^2| > h_1 d_n n^{-\iota}) &< 6s_n \exp(-c_7 n^{1-4\iota}) \\
&+ 4s_n \exp(-c_8 d_n^{-2} n^{1-2\iota}) \\
&+ 12d_n^2 \exp(-t_2 d_n^{-3} n) + 6d_n^2 \exp\{-t_3 d_n^{-2} n^{-\iota}\}.
\end{aligned}$$

Then Theorem 1(1) follows from the union bound of probability.

Next we prove part (2). For the event

$$\mathcal{A} = \{\max_{j \in \mathcal{M}_*} |\|\hat{f}_j\|_n^2 - \|\tilde{f}_j\|^2| \leq c_6 n^{-\iota} / 2\},$$

we have

$$\|\hat{f}_j\|_n^2 \geq c_6 n^{-\iota} / 2, \quad \text{for } j \in \mathcal{M}_*$$

from Proposition 1. Hence, we select $\gamma_n = c_6 n^{-\iota} / 2$, such that $\mathcal{M}_* \subset \mathcal{M}_{\gamma_n}$. The result gains from a simple union of probability ,

$$\begin{aligned}
P(\mathcal{A}^c) &\leq s_n \{6s_n \exp(-c_7 n^{1-4\iota}) \\
&+ 4s_n \exp(-c_8 d_n^{-2} n^{1-2\iota}) \\
&+ 12d_n^2 \exp(-t_2 d_n^{-3} n) + 6d_n^2 \exp(-t_3 d_n^{-2} n^{-\iota}).\}
\end{aligned}$$

The Theorem 1(2) follows. $\square$

The proof of Theorem 2. Let $\|\cdot\|_F$ denote the Frobenius matrix norm, let $\|\cdot\|_2^2$ denote the

spectral matrix norm and $\Sigma = E(H_j^T H_j)$, Note that

$$\begin{aligned}
\sum_{j=1}^p \|\tilde{f}_j\|^2 &= \sum_{j=1}^p (EH_j \tilde{b}_j - Q_\alpha(Y))^2 \\
&\leq \sum_{i=1}^n \lambda_{\max}(E(H_j^T H_j)) \|E \tilde{b}_j\|^2 + c_7 n^{-\iota} \\
&\leq M_1 \sum_{j=1}^p \lambda_{\max}(E(H_j^T H_j)) \\
&\leq M_1 \sum_{j=1}^p \text{trace}(E(H_j^T H_j)) \\
&\leq M_1 E \left[\sum_{j=1}^p \sum_{k=1}^{d_n} (B_k^2(U) + B_k^2(U) X_j^2) \right] \\
&\leq M_1 E(\|H_j\|_F^2) \leq M_1 d_n E\|H_j\|_2^2 = M_1 d_n \lambda_{\max}(\Sigma),
\end{aligned}$$

where the second inequality comes from the result of above S_3 of some $c_7 > 0$, the third from (6.18) of sufficiently large M_1 . This predicated the number of $\{j : \|\tilde{f}_j\|^2 > c_7 d_n n^{-\iota}\}$ cannot exceed $O(n^\iota \lambda_{\max}(\Sigma))$ for any $c_7 > 0$.

Set

$$A_n = \{\max_{1 \leq j \leq p} |\hat{f}_j - \tilde{f}_j| \leq c_7 d_n n^{-\iota}\}.$$

This implies the number of $\{j : \|\hat{f}_j\|^2 > 2c_7 d_n n^{-\iota}\}$ can not exceed the number of $\{j : \|\tilde{f}_j\|^2 > c_7 d_n n^{-\iota}\}$, which is bounded by $O(n^\iota \lambda_{\max}(\Sigma))$. We take $c_7 = c_4/2$, then

$$P\{|\widehat{\mathcal{M}}_{\gamma_n}| \leq O(n^\iota \lambda_{\max}(\Sigma))\} \geq P(A_n).$$

Then the result follows from Theorem 1(1). $\square$

References

- [1] Angrist, J., Chernozhukov, V. and Fernandez-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure, *Econometrica*, 74, 539-563.
- [2] Fan, M., Ma, y. and Dai, W. (2013). Nonparametric Independence Screening in Sparse Ultra-High Dimensional Varying Coefficient Models, arxiv:1303.0458v1.
- [3] Schumaker, L. L. (1981). Spline functions: Basic theory, Wiley, New York.
- [4] Knight, K. (1998). Limiting distributions for L1 regression estimators under general conditions, *Ann. Statist.* 26 755-770.

- [5] He, X., Wang, L. and Hong, H. (2013). Quantile-Adaptive Model-Free Variable Screening for High-Dimensional Heterogeneous Data, *Ann. Statist.*, 41, 342-369.