THE DANTZIG SELECTOR IN COX'S PROPORTIONAL HAZARDS MODEL

Anestis Antoniadis,

Laboratoire Jean Kuntzmann, Department de Statistique,

Université Joseph Fourier, B.P. 53

38041 Grenoble CEDEX 9, France.

Piotr Fryzlewicz,

Department of Statistics, London School of Economics

Columbia House, Houghton Street, London WC2A 2AE, UK.

and

Frédérique Letué,

Laboratoire Jean Kuntzmann, Department de Statistique,

Université Pierre Mendès France, B.P. 53

38041 Grenoble CEDEX 9, France.

1 Appendix: Proofs

Proof of Theorem 1. To prove the result, we will also need the following Lemma which we state with no proof since it is a straightforward generalization of Lemma 3.1 in Candès & Tao (2007).

Lemma 1.1 *Let A be an $n \times p$ matrix and suppose $T_0 \subset \{1, \ldots, p\}$ is a set of cardinality S. For a vector $h \in \mathbb{R}^p$, let T_1 be the S' largest positions of h outside of T_0 and put $T_{01} = T_0 \cup T_1$.*

Then

$$\begin{aligned} \|h_{T_{01}}\|_2 &\leq \frac{1}{\delta_{S+S'}}\|A_{T_{01}}^T Ah\|_2 + \frac{\theta_{S',S+S'}}{\delta_{S+S'}(S')^{1/2}}\|h_{T_0^c}\|_1 \\ \|h\|_2^2 &\leq \|h_{T_{01}}\|_2^2 + (S')^{-1}\|h_{T_0^c}\|_1^2. \end{aligned}$$

To prove the Theorem we need to establish that $\|U(\beta_0)\|_\infty \leq \gamma$ implies that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2^2 \leq 64S(\frac{\gamma}{\delta_{2S}-\theta_{S,2S}})^2$. Assume that $\|U(\beta_0)\|_\infty \leq \gamma$ where

$$\|U(\boldsymbol{\beta}_0)\|_\infty = \sup_j \left| \frac{1}{n}\sum_{i=1}^n \int_0^\tau dM_i(u) \left[\sum_{k=1}^n \left\{Z_{ij} - Z_{kj}\right\} w_k(u)\right]\right|.$$

Let us first prove the consistency of SDS. We prove it first for the non-zero components of $\boldsymbol{\beta}$, and then for the zero components of $\boldsymbol{\beta}$. Let T_0 be the support of $\boldsymbol{\beta}_0$.

- In this item, we work in the subset generated by the non-zero components of $\boldsymbol{\beta}$, but we omit the T_0 index for the sake of readability. Recall that $U(.)$ converges to some function $u(.)$, uniformly on any compact subset (Andersen & Gill (1982)) and that from our hypotheses, the limit u admits a unique zero at point $\boldsymbol{\beta}_0$. Hence, since the matrix $I(\boldsymbol{\beta}_0, \tau)$ is positive definite, there exists some η s.t. $\inf_{\boldsymbol{\beta}\notin\mathcal{B}(\boldsymbol{\beta}_0,\eta)} \|u(\boldsymbol{\beta})\|_\infty = \rho > 0$, where $\mathcal{B}(\boldsymbol{\beta}_0, \eta)$ is the ball centered at $\boldsymbol{\beta}$ with radius η. Let $\varepsilon < \eta$. For n large enough, $\sup_{\boldsymbol{\beta}} \|U(\boldsymbol{\beta}) - u(\boldsymbol{\beta})\|_\infty < \rho/2$. Therefore, for any $\boldsymbol{\beta}$ outside the ball $\mathcal{B}(\boldsymbol{\beta}_0, \varepsilon)$ and n large enough, $\|U(\boldsymbol{\beta})\|_\infty > \rho/2$. Finally, by definition, $\|U(\hat{\boldsymbol{\beta}})\|_\infty < \gamma < \rho/2$, for n large enough. Therefore, $\hat{\boldsymbol{\beta}} \in \mathcal{B}(\boldsymbol{\beta}_0, \varepsilon)$, for n large enough, which proves the consistency for the non-zero components of $\boldsymbol{\beta}_0$.
- To prove the consistency of the zero-components, just remark that $\|h_{T_0}\|_1$ tends to zero when n tends to infinity in inequality (16) of the initial paper.

Recall here that for any consistent estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$, we may write:

$$\begin{aligned} J(\tilde{\boldsymbol{\beta}}, \tau) - I(\boldsymbol{\beta}_0, \tau) &= \int_0^\tau (V_n(\tilde{\boldsymbol{\beta}}, u) - v(\tilde{\boldsymbol{\beta}}, u))\frac{d\bar{N}(u)}{n} &(1)\\ &+ \int_0^\tau (v(\tilde{\boldsymbol{\beta}}, u) - v(\boldsymbol{\beta}_0, u))\frac{d\bar{N}(u)}{n} &(2)\\ &+ \int_0^\tau v(\boldsymbol{\beta}_0, u)\frac{d\bar{M}(u)}{n} &(3)\\ &+ \int_0^\tau v(\boldsymbol{\beta}_0, u)(\frac{S_n(\boldsymbol{\beta}_0, u)}{n} - s(\boldsymbol{\beta}_0, u))\alpha_0(u)du, &(4) \end{aligned}$$

where $V_n(\boldsymbol{\beta},u)=\frac{S_n^2}{S_n}(\boldsymbol{\beta},u)-(\frac{S_n^1}{S_n})^{\otimes 2}(\boldsymbol{\beta},u)$ and $v(\boldsymbol{\beta},u)=\frac{s^2}{s}(\boldsymbol{\beta},u)-(\frac{s^1}{s})^{\otimes 2}(\boldsymbol{\beta},u)$. Since $\boldsymbol{\beta}_0$ is a nonzero S-sparse vector with S independent of n and since the true information matrix $\mathcal{I}(\boldsymbol{\beta}_0,\tau)$ is positive definite at $\boldsymbol{\beta}_0$, for any $\boldsymbol{\beta}^*$ in an Euclidian ball $B_r=B(\boldsymbol{\beta}_0,r)$ centered at $\boldsymbol{\beta}_0$ and of radius at most $r=8\sqrt{S}\frac{\gamma}{\delta_{2S}-\theta_{S,2S}}$, the regularity conditions of Theorem 3.4 in Huang (1996) hold and it follows that

$$\sup_{\tilde{\boldsymbol{\beta}}\in B_r}\|J(\boldsymbol{\beta}^\star,\tau)-I(\boldsymbol{\beta}_0,\tau)\|_\infty=O_P(n^{-1/2}) \tag{5}$$

as n tends to ∞.

Define $h=\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0$. According to Lemma 1 of the initial paper, we have $\|\hat{\boldsymbol{\beta}}\|_1\leq\|\boldsymbol{\beta}_0\|_1$ and this inequality implies that $\|h_{T_0^c}\|_1\leq\|h_{T_0}\|_1$, which yields, by Cauchy inequality,

$$\|h_{T_0^c}\|_1\leq\|h_{T_0}\|_1\leq S^{1/2}\|h_{T_0}\|_2. \tag{6}$$

By assumption, we have $\|U(\boldsymbol{\beta}_0)\|_\infty\leq\gamma$ and by construction of the estimator, $\|U(\hat{\boldsymbol{\beta}})\|_\infty\leq\gamma$. Adding up the two inequalities (triangle inequality)

$$\|U(\boldsymbol{\beta})-U(\hat{\boldsymbol{\beta}})\|_\infty\leq 2\gamma$$

By Andersen & Gill (1982), formula (2.6), we have, Taylor-expanding the LHS of the above,

$$\left\|J(\boldsymbol{\beta}^*,\tau)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0})\right\|_\infty\leq 2\gamma, \tag{7}$$

where $\boldsymbol{\beta}^*$ lies within the segment between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$.

Now, using the consistency of SDS and our remark (5) on the behavior of the matrix $I(\boldsymbol{\beta}_0,\tau)$ at the neighborhood of $\boldsymbol{\beta}_0$ we have

$$\begin{aligned}\left\|I(\boldsymbol{\beta_0},\tau)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0})\right\|_\infty &\leq \left\|(J(\boldsymbol{\beta}^*,\tau)-I(\boldsymbol{\beta_0},\tau))(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0})\right\|_\infty+\left\|J(\boldsymbol{\beta}^*,\tau)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0})\right\|_\infty\\ &\leq Dn^{-1/2}\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0}\right\|_1+2\gamma,\\ &\leq 4\gamma,\end{aligned}$$

for n large enough, since $\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta_0}\right\|_1\leq\left\|\hat{\boldsymbol{\beta}}\right\|_1+\|\boldsymbol{\beta_0}\|_1\leq 2\|\boldsymbol{\beta_0}\|_1$. Hence, if $A=I(\beta_0,\tau)^{1/2}$ denotes the squared root of the (semi)definite positive matrix $I(\beta_0,\tau)$, we have

$$\|AAh\|_\infty\leq 4\gamma.$$

This, again by Cauchy inequality, implies $\|A^T_{T_{01}} Ah\|_2 \leq 4(S+S')^{1/2}\gamma$. Take $S' = S$. By the first inequality of Lemma 1.1 and inequality (6), we have

$$\begin{aligned} \|h_{T_{01}}\|_2 &\leq \frac{4}{\delta_{2S}}(2S)^{1/2}\gamma + \frac{\theta_{S,2S}}{\delta_{2S}S^{1/2}} S^{1/2}\|h_{T_0}\|_2 \\ &\leq \frac{4}{\delta_{2S}}(2S)^{1/2}\gamma + \frac{\theta_{S,2S}}{\delta_{2S}}\|h_{T_{01}}\|_2. \end{aligned}$$

Rearranging for $\|h_{T_{01}}\|_2$, we get

$$\begin{aligned} \|h_{T_{01}}\|_2 \left(1 - \frac{\theta_{S,2S}}{\delta_{2S}}\right) &\leq \frac{4}{\delta_{2S}}(2S)^{1/2}\gamma \\ \|h_{T_{01}}\|_2 &\leq \frac{4}{\delta_{2S} - \theta_{S,2S}}(2S)^{1/2}\gamma. \end{aligned}$$

By the second inequality of Lemma 1.1 and inequality (6), we have

$$\|h\|_2^2 \leq \|h_{T_{01}}\|_2^2 + S^{-1}S\|h_{T_0}\|_2^2 \leq 2\|h_{T_{01}}\|_2^2 \leq 64S(\frac{\gamma}{\delta_{2S} - \theta_{S,2S}})^2,$$

which completes the proof of the Theorem.

References

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. Annals of Statistics **10**, 1100–1120.

Candès, E. & Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. Annals of Statistics **35**, 2313–2351.

Huang, J. (1996). Efficient estimation for the Cox model with interval censoring. Annals of Statistics **24**, 540–568.