

Online Appendix to “Sparse Quantile Regression” by Le-Yu Chen and Sokbae Lee

B Additional Results of the Empirical Application

In this online appendix, we provide the variable selection results obtained under ℓ_0 -PQR, ℓ_0 -CQR, ℓ_1 -PQR and the other four alternative penalized quantile regression approaches (AL-SCAD, AL-MCP, QR-SCAD, QR-MCP) in our empirical study of Section 6. To facilitate presentation of these results, we give below further details on the five covariate specifications used in this study.

Table B1 lists names and definitions of the basic covariates. These variables constitute the first covariate specification where $p = 21$. The second specification builds on and modifies the first as follows. Let `medu1`, `medu2`, `medu3` be the binary variables indicating respectively whether mother’s years of education were exactly 12, between 12 and 16, or at least 16. Define analogously `fedu1`, `fedu2`, `fedu3` to be the corresponding indicator variables that discretize father’s years of education. For $j = 1, \dots, m + 3$, let $B_j(\text{mage})$, $B_j(\text{fage})$, $B_j(\text{npren})$ and $B_j(\text{mslb})$ denote the cubic B-spline series terms for approximating functions of the variables `mage`, `fage`, `nprenatal` and `months1b` respectively using m interior knots where these approximations do not include B-spline intercept terms.

The second specification consists of all variables of the first specification except that the covariates `medu` and `fedu` are replaced by the six indicator variables that discretize both parents’ years of educations as defined above, and moreover the variables `mage`, `fage`, `nprenatal` and `months1b` are replaced by their corresponding cubic B-spline terms using 4 interior knots. The third covariate specification consists of all variables in the second specification together with those obtained by interacting the B-spline expansion terms with the other explanatory variables. Both the fourth and fifth specifications are constructed using the same procedure as for the third case except that we increase numbers of interior B-spline knots to be 12 and 16 respectively for these two specifications. Accordingly, the covariate vector under the second, third, fourth and fifth specifications has dimension 49, 609, 1281, and 1617 respectively. Finally, for each covariate specification, all stochastic covariates thus constructed are further standardized to have mean zero and variance unity.

We now discuss the variable selection results of our empirical study. Tables B2 - B11 report results of the top 10 most often selected covariates as well as their proportions of being selected and the corresponding average estimated regression coefficient values. From these tables, we note that the regression intercept was always selected under every estimation approach and across all the covariate specifications. In addition, its average estimated value was quite similar in most estimation scenarios. At 5% quantile level, for the case with $p = 21$, Table B2 indicates that the variable for number of prenatal care visits (`nprenatal`) was also always selected and other important predictors were mother’s smoking behavior during pregnancy (`msmoke`) and her race (`mrace`), both being selected with at least 60% incidence rate across all estimation approaches. For each of these three variables, the corresponding estimated coefficient was also of the same sign and had similar mag-

nitude across all the methods. At 95% quantile level, analogous variable selection results emerged in Table B3 though `mrace` was no longer listed among the top 4 most often selected variables under some of the estimation approaches.

For the covariate specification with $p = 49$, Table B4 shows that the B-spline expansion terms $B_1(\text{npre})$, $B_2(\text{npre})$ and $B_4(\text{npre})$ were among the top 4 most often selected variables across most of the quantile regression approaches for the estimation at 5% quantile level. This indicates that the true 5% level conditional quantile function could be nonlinear in the variable `nprenatal`. Yet, at 95% quantile level, we find that, except for $B_5(\text{npre})$, which was selected in at least 60% of the conformal prediction replications under AL-SCAD and AL-MCP, covariates of the B-spline series terms appeared to be less important under the other estimation approaches. By contrast, maternal smoking behavior (`msmoke`) was the most often selected stochastic variable across all the estimation approaches in this setting.

Finally, for higher dimensional cases with $p \in \{609, 1281, 1617\}$, it is evident from Tables B6 - B11 that, except for the ℓ_1 -PQR cases, for each of the stochastic covariates, its incidence of selection was well capped below 60% under all the other estimation approaches. Specifically, at 95% quantile level in the case with $p = 1617$, regression intercept was the only selected covariate under both MIO and FO based implementations of ℓ_0 -PQR. On the whole, while all estimation approaches agreed to the selection of regression intercept, we note that the variable selection results generally appeared to vary to a much larger extent across methods under the high dimensional covariate specifications.

Table B1: Names and definitions of basic covariates

Variable name	Definition
intercept	regression intercept
married	marital status
mage	mother's age
medu	mother's years of education
mhisp	whether mother is hispanic
mrace	whether mother's race is white
fage	father's age
fedu	father's years of education
fhisp	whether father is hispanic
frace	whether father's race is white
foreign	whether mother was born abroad
alcohol	whether mother drank alcohol during pregnancy
msmoke	whether mother smoked during pregnancy
deadkids	whether a newborn died in previous births
monthslb	number of months since last birth
nprenatal	number of prenatal care visits
trimester1	whether the first prenatal care visit was in the first trimester
fbaby	whether the infant was the first born child
season1	whether the infant was born in the winter
season2	whether the infant was born in the spring
season3	whether the infant was born in the summer

Table B2: Top 10 most often selected variables for $p = 21$ at 5% quantile

	ℓ_0 -PQR		ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
	MIO	FO						
1st	intercept (1,2.47)	intercept (1,2.46)	intercept (1,2.46)	intercept (1,2.44)	intercept (1,2.47)	intercept (1,2.48)	intercept (1,2.47)	intercept (1,2.47)
2nd	nprenatal (1,0.16)	nprenatal (1,0.15)	nprenatal (1,0.16)	nprenatal (1,0.14)	nprenatal (1,0.14)	nprenatal (1,0.14)	nprenatal (1,0.17)	nprenatal (1,0.17)
3rd	msmoke (0.7,-0.11)	msmoke (0.6,-0.12)	msmoke (0.7,-0.1)	mrace (0.9,0.11)	mrace (0.7,0.16)	mrace (0.8,0.16)	mrace (0.9,0.19)	mrace (0.9,0.19)
4th	mrace (0.6,0.24)	mrace (0.6,0.18)	mrace (0.7,0.23)	msmoke (0.8,-0.06)	msmoke (0.6,-0.09)	msmoke (0.7,-0.09)	msmoke (0.6,-0.11)	msmoke (0.6,-0.11)
5th	trimester1 (0.3,-0.12)	medu (0.3,0.08)	trimester1 (0.4,-0.09)	foreign (0.6,-0.01)	trimester1 (0.6,-0.05)	trimester1 (0.7,-0.07)	trimester1 (0.6,-0.1)	trimester1 (0.6,-0.1)
6th	medu (0.2,0.12)	frace (0.3,0.15)	medu (0.2,0.12)	frace (0.6,0.06)	medu (0.5,0.06)	medu (0.6,0.06)	mhis (0.3,0.04)	mhis (0.3,0.04)
7th	frace (0.2,0.19)	trimester1 (0.3,-0.11)	frace (0.2,0.19)	married (0.5,0.03)	married (0.3,0.03)	foreign (0.4,-0.02)	foreign (0.3,-0.07)	foreign (0.3,-0.07)
8th	season2 (0.2,-0.07)	season2 (0.2,-0.07)	season2 (0.2,-0.07)	mage (0.5,0.01)	foreign (0.3,-0.02)	married (0.3,0.03)	monthslb (0.3,-0.02)	monthslb (0.3,-0.02)
9th	foreign (0.1,-0.08)	married (0.1,0.07)	married (0.1,0.06)	season1 (0.5,0)	monthslb (0.3,0.02)	deadkids (0.3,0.01)	frace (0.3,0.08)	frace (0.3,0.08)
10th	mage (0.1,-0.05)	foreign (0.1,-0.08)	mhis (0.1,0.11)	season2 (0.5,-0.01)	frace (0.3,0.14)	monthslb (0.3,0.02)	fbaby (0.3,-0.02)	fbaby (0.3,-0.01)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conformalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B3: Top 10 most often selected variables for $p = 21$ at 95% quantile

	ℓ_0 -PQR		ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
	MIO	FO						
1st	intercept (1,4.2)	intercept (1,4.19)	intercept (1,4.2)	intercept (1,4.16)	intercept (1,4.19)	intercept (1,4.2)	intercept (1,4.19)	intercept (1,4.2)
2nd	msmoke (1,-0.11)	msmoke (1,-0.1)	msmoke (1,-0.11)	nprenatal (1,0.06)	msmoke (1,-0.1)	msmoke (1,-0.11)	msmoke (1,-0.11)	msmoke (1,-0.11)
3rd	nprenatal (0.5,0.09)	nprenatal (0.7,0.08)	nprenatal (0.7,0.08)	msmoke (1,-0.08)	nprenatal (0.7,0.09)	mrace (0.8,0.05)	nprenatal (0.4,0.11)	nprenatal (0.5,0.09)
4th	mrace (0.5,0.08)	alcohol (0.5,-0.04)	mrace (0.6,0.07)	mrace (0.9,0.03)	season1 (0.7,0.06)	nprenatal (0.7,0.08)	mrace (0.4,0.07)	mrace (0.5,0.08)
5th	alcohol (0.4,-0.02)	fage (0.4,0.02)	alcohol (0.5,-0.03)	alcohol (0.8,-0.03)	mrace (0.6,0.07)	frace (0.5,-0.01)	fage (0.3,-0.02)	alcohol (0.4,-0.02)
6th	season1 (0.4,0.07)	mrace (0.4,0.08)	fbaby (0.5,-0.06)	medu (0.8,0.02)	fbaby (0.5,-0.05)	fbaby (0.5,-0.05)	trimester1 (0.3,-0.01)	season1 (0.4,0.07)
7th	married (0.3,0)	fbaby (0.4,-0.08)	fage (0.4,0.02)	foreign (0.7,-0.01)	trimester1 (0.5,-0.02)	season1 (0.5,0.08)	season1 (0.3,0.08)	fhispl (0.3,-0.02)
8th	medu (0.3,0.05)	trimester1 (0.4,-0.03)	season1 (0.4,0.07)	fage (0.7,0.01)	married (0.4,-0.02)	married (0.4,-0.01)	married (0.2,0.03)	mage (0.3,0.03)
9th	fage (0.3,0.01)	season1 (0.4,0.08)	married (0.3,-0.01)	fbaby (0.7,-0.04)	fhispl (0.4,-0.03)	fage (0.4,0.01)	fhispl (0.2,-0.03)	frace (0.3,-0.06)
10th	frace (0.3,-0.05)	foreign (0.3,0)	trimester1 (0.3,0)	season1 (0.7,0.04)	frace (0.4,-0.01)	trimester1 (0.4,-0.02)	alcohol (0.2,-0.02)	trimester1 (0.3,0.01)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conformalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B4: Top 10 most often selected variables for $p = 49$ at 5% quantile

	ℓ_0 -PQR		ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
	MIO	FO						
1st	intercept (1,2.47)	intercept (1,2.45)	intercept (1,2.47)	intercept (1,2.44)	intercept (1,2.46)	intercept (1,2.47)	intercept (1,2.45)	intercept (1,2.44)
2nd	B4(npres) (1,0.3)	B1(npres) (0.8,-0.26)	B4(npres) (1,0.38)	B4(npres) (1,0.11)	B4(npres) (1,0.25)	B4(npres) (1,0.22)	B4(npres) (0.9,0.27)	B4(npres) (0.8,0.27)
3rd	frace (0.6,0.12)	B4(npres) (0.8,0.21)	frace (0.7,0.11)	B1(npres) (0.9,-0.14)	B1(npres) (0.8,-0.26)	B1(npres) (0.8,-0.19)	B1(npres) (0.7,-0.33)	B1(npres) (0.7,-0.32)
4th	B2(npres) (0.6,0.23)	frace (0.5,0.1)	trimester1 (0.6,-0.11)	mrace (0.7,0.06)	B2(npres) (0.6,0.23)	B2(npres) (0.5,0.17)	B2(npres) (0.6,0.22)	B2(npres) (0.5,0.23)
5th	trimester1 (0.5,-0.11)	msmoke (0.4,-0.11)	B2(npres) (0.6,0.25)	B3(mslb) (0.7,-0.03)	mrace (0.5,0.12)	mrace (0.4,0.13)	B4(mslb) (0.4,0.18)	mrace (0.3,0.1)
6th	msmoke (0.4,-0.12)	B1(fage) (0.4,-0.11)	B3(npres) (0.6,0.31)	frace (0.6,0.08)	fbaby (0.4,0.06)	trimester1 (0.4,-0.06)	msmoke (0.3,-0.12)	B1(fage) (0.3,-0.09)
7th	mrace (0.4,0.14)	mrace (0.3,0.13)	B1(npres) (0.4,-0.3)	B1(fage) (0.6,-0.06)	B1(fage) (0.4,-0.04)	B1(fage) (0.4,-0.05)	mrace (0.3,0.16)	msmoke (0.2,-0.11)
8th	B1(npres) (0.4,-0.38)	B3(npres) (0.3,0.11)	msmoke (0.3,-0.1)	msmoke (0.5,-0.05)	B3(mslb) (0.4,-0.07)	B3(mslb) (0.4,-0.07)	fbaby (0.3,0.09)	frace (0.2,0.11)
9th	B3(npres) (0.4,0.28)	medu1 (0.2,-0.09)	B5(npres) (0.3,0.18)	trimester1 (0.5,-0.03)	B4(mslb) (0.4,0.15)	B4(mslb) (0.4,0.13)	trimester1 (0.3,-0.13)	fbaby (0.2,0.09)
10th	mhisps (0.2,0.09)	B4(mage) (0.2,0)	mrace (0.2,0.15)	B4(mslb) (0.5,0.03)	msmoke (0.3,-0.1)	msmoke (0.3,-0.09)	B3(mslb) (0.3,-0.11)	trimester1 (0.2,-0.1)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conformalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B5: Top 10 most often selected variables for $p = 49$ at 95% quantile

	ℓ_0 -PQR		ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
	MIO	FO						
1st	intercept (1,4.19)	intercept (1,4.19)	intercept (1,4.19)	intercept (1,4.15)	intercept (1,4.19)	intercept (1,4.19)	intercept (1,4.19)	intercept (1,4.2)
2nd	msmoke (0.9,-0.1)	msmoke (1,-0.1)	msmoke (1,-0.11)	msmoke (1,-0.07)	msmoke (0.8,-0.1)	msmoke (0.9,-0.09)	msmoke (0.8,-0.11)	msmoke (0.7,-0.11)
3rd	mrace (0.4,0.07)	mrace (0.3,0.06)	mrace (0.3,0.08)	alcohol (0.8,-0.02)	B5(npri) (0.6,0.08)	B5(npri) (0.8,0.07)	B5(npri) (0.5,0.1)	B5(npri) (0.4,0.12)
4th	B5(npri) (0.3,0.12)	B1(npri) (0.3,-0.06)	fedu2 (0.2,0.06)	mrace (0.8,0.03)	season1 (0.4,0.04)	fedu2 (0.5,0.04)	B6(npri) (0.4,-0.08)	B6(mslb) (0.4,-0.09)
5th	fbaby (0.2,-0.06)	season1 (0.2,0.07)	B1(fage) (0.2,-0.08)	fedu2 (0.7,0.02)	B1(fage) (0.4,-0.06)	mrace (0.4,0.03)	fedu1 (0.3,0.08)	B3(npri) (0.3,0.1)
6th	B6(mage) (0.2,0.09)	B4(mage) (0.2,-0.07)	B1(npri) (0.2,-0.07)	B1(fage) (0.7,-0.02)	B5(fage) (0.4,0.02)	season1 (0.4,0.03)	fedu2 (0.3,0.07)	B6(npri) (0.3,-0.08)
7th	B1(fage) (0.2,-0.1)	B1(fage) (0.2,-0.06)	B4(npri) (0.2,0.09)	B1(npri) (0.7,-0.04)	married (0.3,-0.01)	fedu1 (0.4,0.03)	B6(mage) (0.3,0.09)	B7(mslb) (0.3,0.12)
8th	B3(fage) (0.2,0.01)	B5(fage) (0.2,0.01)	frace (0.1,0.08)	B5(npri) (0.7,0.02)	season2 (0.3,0.02)	B1(fage) (0.4,-0.07)	B3(npri) (0.3,0.11)	married (0.2,-0.02)
9th	B1(npri) (0.2,-0.02)	married (0.1,0.09)	season1 (0.1,0.09)	B4(mslb) (0.7,0.02)	fedu2 (0.3,0.05)	B5(fage) (0.4,0.01)	B4(npri) (0.3,0.14)	mrace (0.2,0.09)
10th	B6(npri) (0.2,-0.07)	fbaby (0.1,-0.04)	medu2 (0.1,0.06)	season1 (0.5,0.02)	B3(mage) (0.3,0.01)	B6(npri) (0.4,-0.07)	B6(mslb) (0.3,-0.08)	fedu1 (0.2,0.11)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conformalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B6: Top 10 most often selected variables for $p = 609$ at 5% quantile

	MIO	ℓ_0 -PQR	FO	ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
1st	intercept (1,2.44)	intercept (1,2.46)	intercept (1,2.45)	intercept (1,2.42)	intercept (1,2.45)	intercept (1,2.45)	intercept (1,2.45)	intercept (1,2.44)	intercept (1,2.44)
2nd	B1(npree) (0.2,-0.45)	trimester1*B1(npree) (0.4,-0.23)	trimester1*B1(npree) (0.5,-0.26)	trimester1*B1(npree) (0.8,-0.09)	foreign*B7(npree) (0.3,-0.22)	foreign*B7(npree) (0.2,0.05)	foreign*B7(npree) (0.2,0.05)	trimester1*B1(npree) (0.2,-0.19)	trimester1*B1(npree) (0.2,-0.26)
3rd	B4(npree) (0.2,0.23)	frace*B4(npree) (0.3,0.2)	frace*B4(npree) (0.4,0.21)	frace*B4(npree) (0.6,0.04)	frace*B4(npree) (0.2,0.03)	foreign*B7(npree) (0.2,-0.34)	foreign*B7(npree) (0.2,-0.34)	season2 (0.1,-0.4)	B4(npree) (0.1,0.23)
4th	married*B2(npree) (0.1,-0.01)	foreign*B7(npree) (0.2,-0.75)	B4(npree) (0.3,0.17)	B4(npree) (0.6,0.06)	trimester1*B1(npree) (0.2,-0.18)	trimester1*B1(npree) (0.2,-0.21)	trimester1*B1(npree) (0.2,-0.21)	B1(fage) (0.1,-0.15)	foreign*B7(npree) (0.1,-0.69)
5th	frace*B4(npree) (0.1,0.29)	frace*B4(npree) (0.1,0.14)	B1(fage) (0.1,-0.16)	frace*B4(npree) (0.6,0.06)	B4(npree) (0.1,0.13)	B1(npree) (0.1,-0.01)	B1(npree) (0.1,-0.01)	B1(npree) (0.1,-0.47)	alcohol*B7(npree) (0.1,0)
6th	frace*B5(mslb) (0.1,0.04)	frace (0.1,0.16)	married*B2(npree) (0.1,0.15)	frace (0.4,0.05)	mhispp*B7(fage) (0.1,0.01)	B4(npree) (0.1,0.13)	B4(npree) (0.1,0.13)	B4(npree) (0.1,0.23)	mrace*B2(npree) (0.1,0.22)
7th	fedu1*B1(npree) (0.1,-0.35)	B1(fage) (0.1,-0.11)	married*B4(npree) (0.1,0.08)	mrace*B4(npree) (0.4,0.02)	deadkids*B1(npree) (0.1,-0.05)	married*B3(mslb) (0.1,-0.01)	married*B3(mslb) (0.1,-0.01)	B4(mslb) (0.1,0.13)	mrace*B4(npree) (0.1,0.27)
8th	B4(npree) (0.1,0.12)	foreign*B6(npree) (0.1,0)	foreign*B6(npree) (0.1,0)	fedu1*B1(npree) (0.4,-0.07)	mmsmoke*B4(mage) (0.1,-0.09)	mhispp*B7(fage) (0.1,0.01)	mhispp*B7(fage) (0.1,0.01)	married*B3(mslb) (0.1,-0.16)	fedu1*B1(npree) (0.1,-0.35)
9th	deadkids*B1(npree) (0.1,-0.1)	alcohol*B5(npree) (0.1,0)	alcohol*B5(npree) (0.1,0)	B1(fage) (0.3,-0.03)	mrace*B3(mage) (0.1,0.11)	alcohol*B7(mslb) (0.1,-0.15)	alcohol*B7(mslb) (0.1,-0.15)	fhispp*B7(mslb) (0.1,0.31)	fedu1*B4(npree) (0.1,0.19)
10th	mmsmoke*B3(mage) (0.1,-0.15)	mmsmoke*B4(mage) (0.1,0.25)	mmsmoke*B4(mage) (0.1,0.25)	mmsmoke*B4(mage) (0.3,-0.02)	mmsmoke*B4(mage) (0.1,0.13)	deadkids*B1(npree) (0.1,-0.06)	deadkids*B1(npree) (0.1,-0.06)	foreign*B7(npree) (0.1,-0.3)	fedu2*B1(npree) (0.1,-0.18)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conormalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B7: Top 10 most often selected variables for $p = 609$ at 95% quantile

	ℓ_0 -PQR		FO	ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
1st	MIO intercept (1,4.22)		intercept (1,4.21)	intercept (1,4.2)	intercept (1,4.03)	intercept (1,4.17)	intercept (1,4.17)	intercept (1,4.18)	intercept (1,4.16)
2nd	msmoke (0.1,-0.08)		msmoke (0.1,-0.06)	msmoke (0.3,-0.12)	msmoke (0.6,-0.05)	mhispp*B7(mage) (0.2,-1.55)	mhispp*B7(mage) (0.2,-2.1)	married*B5(npres) (0.2,0.08)	mhispp*B7(mage) (0.2,-9.11)
3rd	B1(npres) (0.1,-0.06)		B4(mage) (0.1,-0.16)	married*B4(npres) (0.2,0.06)	mhispp*B7(npres) (0.4,0)	season1*B7(npres) (0.2,-0.2)	season1*B7(npres) (0.2,-0.21)	mhispp*B7(mage) (0.2,0.89)	married*B5(npres) (0.1,0.01)
4th	deadkids*B2(fage) (0.1,-0.03)		B6(mage) (0.1,0.04)	msmoke*B3(fage) (0.2,-0.15)	mrace (0.3,0.01)	trimester1 (0.1,0.09)	trimester1 (0.1,0.07)	season1*B7(npres) (0.2,-0.29)	frace*B4(npres) (0.1,0.01)
5th	season1*B5(npres) (0.1,0.01)		B4(npres) (0.1,0.2)	mrace*B4(mslb) (0.2,0.09)	married*B4(npres) (0.3,-0.01)	B3(mage) (0.1,0.12)	B3(mage) (0.1,0.14)	fbaby (0.1,-0.07)	season1*B3(npres) (0.1,0.09)
6th	season1*B7(mslb) (0.1,-0.06)		married*B3(fage) (0.1,0.07)	frace*B4(fage) (0.2,0.03)	married*B5(npres) (0.3,0.03)	B4(mage) (0.1,-0.11)	B4(mage) (0.1,-0.1)	B1(mage) (0.1,0.19)	season1*B7(npres) (0.1,-0.38)
7th			married*B5(fage) (0.1,-0.12)	frace*B4(npres) (0.2,0.1)	mrace*B4(mslb) (0.3,0.01)	B4(fage) (0.1,-0.01)	B4(fage) (0.1,-0.03)	B3(mage) (0.1,0.15)	medu2*B3(mslb) (0.1,0)
8th			married*B1(npres) (0.1,-0.07)	fbaby*B4(npres) (0.2,0.06)	season2*B3(mslb) (0.3,0.02)	B5(fage) (0.1,-0.04)	B5(fage) (0.1,0)	B2(fage) (0.1,-0.14)	
9th			married*B4(npres) (0.1,-0.11)	season1*B5(fage) (0.2,0.07)	trimester1 (0.2,0)	B6(fage) (0.1,-0.04)	B6(fage) (0.1,-0.06)	B3(fage) (0.1,-0.09)	
10th			married*B5(npres) (0.1,0.07)	B6(fage) (0.1,0.02)	B1(npres) (0.2,-0.04)	B1(npres) (0.1,0.07)	B1(npres) (0.1,0.07)	B3(npres) (0.1,0.08)	

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conformalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B8: Top 10 most often selected variables for $p = 1281$ at 5% quantile

	ℓ_0 -PQR		FO	ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
1st	MIO intercept (1,2.44)		intercept (1,2.44)	intercept (1,2.44)	intercept (1,2.03)	intercept (1,2.03)	intercept (1,2.04)	intercept (1,2.03)	intercept (1,2.03)
2nd	trimester1*B2(npred) (0.3,-0.26)		trimester1*B2(npred) (0.3,-0.26)	trimester1*B1(npred) (0.3,-0.32)	mrace (0.7,0.05)	fedu3*B1(fage) (0.5,-10)	fedu3*B1(fage) (0.5,-10)	fedu3*B1(fage) (0.5,-10)	fedu3*B1(fage) (0.5,-10)
3rd	mrace (0.2,0.13)		mrace (0.2,0.13)	B1(npred) (0.2,-0.26)	frace (0.7,0.04)	mrace (0.3,0.12)	trimester1*B2(npred) (0.4,-0.14)	fedu3*B2(fage) (0.3,-10)	fedu3*B2(fage) (0.3,-10)
4th	trimester1*B1(npred) (0.2,-0.28)		trimester1*B1(npred) (0.2,-0.28)	B12(fage) (0.1,-0.35)	trimester1*B2(npred) (0.7,-0.08)	trimester1*B2(npred) (0.3,-0.16)	fedu1*B1(npred) (0.4,-0.14)	mrace (0.2,0.14)	fhisp*B15(mage) (0.2,-10)
5th	medu2*B15(mage) (0.2,-0.15)		fedu1*B1(npred) (0.2,-0.23)	B2(npred) (0.1,-0.24)	fedu3*B1(fage) (0.5,-10)	fedu3*B2(fage) (0.3,-10)	mrace (0.3,0.12)	fhisp*B15(mage) (0.2,-10)	mrace (0.1,0.16)
6th	frace (0.1,0.11)		medu2*B15(mage) (0.2,-0.15)	B13(mslb) (0.1,-0.12)	msmoke*B7(mage) (0.4,-0.01)	married*B1(npred) (0.2,-0.1)	fedu3*B2(fage) (0.3,-10)	fhisp*B15(fage) (0.2,-5)	married*B1(npred) (0.1,-0.28)
7th	foreign*B1(fage) (0.1,-1.3)		frace (0.1,0.11)	married*B9(mage) (0.1,0.07)	mrace*B7(mslb) (0.4,-0.04)	fhisp*B15(mage) (0.2,-10)	married*B1(npred) (0.2,-0.13)	trimester1*B2(npred) (0.2,-0.27)	mhhisp*B14(mage) (0.1,5.41)
8th	foreign*B15(npred) (0.1,-1.31)		foreign*B1(fage) (0.1,-1.3)	mhhisp*B6(npred) (0.1,0.06)	trimester1*B1(npred) (0.4,-0.09)	season2*B6(fage) (0.2,-0.07)	fhisp*B15(mage) (0.2,-10)	msmoke (0.1,-0.16)	mhhisp*B15(mage) (0.1,-10)
9th	alcohol*B6(mage) (0.1,-0.13)		foreign*B15(npred) (0.1,-1.31)	fhisp*B4(fage) (0.1,-0.04)	fedu1*B1(npred) (0.4,-0.11)	fedu1*B1(npred) (0.2,-0.17)	msmoke (0.1,-0.01)	B9(mage) (0.1,-0.35)	mhhisp*B10(npred) (0.1,-10)
10th	deadkids*B2(npred) (0.1,-0.21)		alcohol*B6(mage) (0.1,-0.13)	alcohol*B7(mage) (0.1,-0.02)	married (0.3,0.01)	B2(fage) (0.1,-0.05)	B2(fage) (0.1,-0.07)	married*B9(mage) (0.1,0.32)	mhhisp*B15(npred) (0.1,-10)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conformalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B9: Top 10 most often selected variables for $p = 1281$ at 95% quantile

	MIO	ℓ_0 -PQR	FO	ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
1st	intercept (1,4.22)	intercept (1,4.22)	intercept (1,4.21)	intercept (1,4.22)	intercept (1,3.56)	intercept (1,3.75)	intercept (1,3.66)	intercept (1,3.76)	intercept (1,3.76)
2nd	msmoke (0.1,-0.16)	msmoke (0.1,-0.15)	msmoke (0.1,-0.16)	msmoke (0.6,-0.06)	msmoke (0.6,-0.06)	fedu3*B1(fage) (0.5,-10)	fedu3*B1(fage) (0.5,-10)	fedu3*B1(fage) (0.5,-10)	fedu3*B1(fage) (0.5,-10)
3rd	season1*B12(mslb) (0.1,0.06)	msmoke (0.1,-0.03)	msmoke (0.1,-0.03)	B8(mage) (0.1,-0.03)	fedu3*B1(fage) (0.5,-10)	fedu3*B2(fage) (0.3,-10)	fedu3*B2(fage) (0.3,-10)	fedu3*B2(fage) (0.3,-10)	fedu3*B2(fage) (0.3,-10)
4th		B2(fage) (0.1,-0.05)	B2(fage) (0.1,-0.05)	married*B12(npri) (0.4,0.03)	married*B12(npri) (0.4,0.03)	fhisp*B15(mage) (0.2,-10)	fhisp*B15(mage) (0.2,-10)	fhisp*B15(mslb) (0.2,-5.73)	fhisp*B15(mage) (0.2,-10)
5th		alcohol*B8(mage) (0.1,0)	alcohol*B8(mage) (0.1,0)	fedu2*B13(mage) (0.4,0)	fedu2*B13(mage) (0.4,0)	fhisp*B11(mage) (0.1,-0.51)	medu2*B10(npri) (0.2,0.02)	fhisp*B15(mage) (0.2,-10)	fhisp*B11(mage) (0.1,-0.51)
6th		alcohol*B11(mslb) (0.1,-0.01)	alcohol*B11(mslb) (0.1,-0.01)	medu2*B9(mslb) (0.4,0.02)	medu2*B9(mslb) (0.4,0.02)	fhisp*B14(mage) (0.1,-3.81)	mrace (0.1,0.11)	fhisp*B15(mslb) (0.2,1.37)	fhisp*B14(mage) (0.1,-3.81)
7th		deadkids*B4(mage) (0.1,-0.01)	deadkids*B4(mage) (0.1,-0.01)	mrace (0.3,0.03)	mrace (0.3,0.03)	fhisp*B15(mage) (0.1,-10)	B9(npri) (0.1,0.03)	mrace (0.1,0.12)	fhisp*B15(mage) (0.1,-10)
8th		deadkids*B13(npri) (0.1,-0.03)	deadkids*B13(npri) (0.1,-0.03)	B1(npri) (0.3,-0.03)	B1(npri) (0.3,-0.03)	fhisp*B10(npri) (0.1,-10)	B11(mslb) (0.1,0.09)	married*B10(mage) (0.1,-0.11)	fhisp*B10(npri) (0.1,-10)
9th		mrace*B10(mslb) (0.1,0.03)	mrace*B10(mslb) (0.1,0.03)	fhisp*B15(npri) (0.3,-3.33)	fhisp*B15(npri) (0.3,-3.33)	fhisp*B15(npri) (0.1,-10)	married*B1(mage) (0.1,-0.01)	married*B13(fage) (0.1,0.08)	fhisp*B15(npri) (0.1,-10)
10th		frace*B15(fage) (0.1,-0.05)	frace*B15(fage) (0.1,-0.05)	deadkids*B12(mslb) (0.3,-0.01)	deadkids*B12(mslb) (0.3,-0.01)	fhisp*B15(mslb) (0.1,-10)	married*B3(mage) (0.1,0.04)	fhisp*B11(mage) (0.1,-0.51)	fhisp*B15(mslb) (0.1,-10)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conormalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B10: Top 10 most often selected variables for $p = 1617$ at 5% quantile

	ℓ_0 -PQR		FO	ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
1st	MIO intercept (1,2.46)	intercept (1,2.46)	intercept (1,2.46)	intercept (1,2.44)	intercept (0.9,1.76)	intercept (0.9,1.79)	intercept (0.9,1.79)	intercept (0.9,1.79)	intercept (0.9,1.79)
2nd	frace (0.2,0.11)	frace (0.2,0.11)	frace (0.2,0.11)	mmsoke*B5(mage) (0.2,-0.04)	frace (0.6,0.04)	trimester1*B2(npri) (0.3,-0.1)	trimester1*B2(npri) (0.4,-0.13)	trimester1*B2(npri) (0.3,-0.28)	trimester1*B2(npri) (0.3,-0.27)
3rd	trimester1*B2(npri) (0.2,-0.28)	trimester1*B2(npri) (0.2,-0.28)	trimester1*B2(npri) (0.2,-0.28)	B11(mage) (0.1,0.07)	trimester1*B2(npri) (0.6,-0.09)	fedu3*B2(fage) (0.3,-10)	mrace (0.3,0.1)	fedu3*B2(fage) (0.3,-10)	fedu3*B2(fage) (0.3,-10)
4th	fedu1*B1(npri) (0.2,-0.14)	fedu1*B1(npri) (0.2,-0.14)	fedu1*B1(npri) (0.2,-0.14)	B12(mage) (0.1,-0.02)	mrace (0.5,0.04)	medu3*B2(mage) (0.3,-10)	mrace*B9(mon) (0.3,-0.03)	medu3*B2(mage) (0.3,-10)	medu3*B2(mage) (0.3,-10)
5th	fedu1*B2(npri) (0.2,-0.24)	fedu1*B2(npri) (0.2,-0.24)	fedu1*B2(npri) (0.2,-0.24)	B15(mage) (0.1,-0.12)	B15(npri) (0.5,0.02)	mrace (0.2,0.12)	fedu3*B2(fage) (0.3,-10)	mhisp*B19(mage) (0.2,-10)	mhisp*B19(mage) (0.2,-10)
6th	fhisp (0.1,0.05)	fhisp (0.1,0.05)	fhisp (0.1,0.05)	married*B11(mage) (0.1,0.18)	mrace*B9(mslb) (0.4,-0.04)	mhisp*B19(mage) (0.2,-10)	medu3*B2(mage) (0.3,-10)	mhisp*B13(npri) (0.2,-10)	mhisp*B13(npri) (0.2,-10)
7th	fhisp*B14(mage) (0.1,-0.08)	fhisp*B14(mage) (0.1,-0.08)	fhisp*B14(mage) (0.1,-0.08)	married*B12(mage) (0.1,0.3)	trimester1*B1(npri) (0.4,-0.06)	mhisp*B13(npri) (0.2,-10)	frace (0.2,0.03)	fhisp*B14(mage) (0.2,-10)	fhisp*B14(mage) (0.2,-10)
8th	foreign*B17(mage) (0.1,-0.08)	foreign*B17(mage) (0.1,-0.08)	foreign*B17(mage) (0.1,-0.08)	married*B19(fage) (0.1,-0.1)	fedu1*B2(npri) (0.4,-0.08)	fhisp*B14(mage) (0.2,-10)	mhisp*B19(mage) (0.2,-10)	foreign*B19(npri) (0.2,-0.68)	foreign*B19(npri) (0.2,-0.68)
9th	foreign*B2(fage) (0.1,-0.24)	foreign*B2(fage) (0.1,-0.24)	foreign*B2(fage) (0.1,-0.24)	mhisp*B4(mage) (0.1,0)	fedu3*B5(mage) (0.4,-0.01)	frace (0.1,0.02)	mhisp*B13(npri) (0.2,-10)	fedu1*B1(npri) (0.2,-0.25)	mrace (0.1,0.22)
10th	foreign*B19(npri) (0.1,-1.25)	foreign*B19(npri) (0.1,-1.25)	foreign*B19(npri) (0.1,-1.25)	mhisp*B8(mage) (0.1,0.02)	fhisp*B14(mage) (0.3,-6.69)	B2(fage) (0.1,-0.09)	fhisp*B14(mage) (0.2,-10)	mrace (0.1,0.22)	B2(fage) (0.1,-0.05)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conformalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.

Table B11: Top 10 most often selected variables for $p = 1617$ at 95% quantile

	ℓ_0 -PQR		ℓ_0 -CQR	ℓ_1 -PQR	AL-SCAD	AL-MCP	QR-SCAD	QR-MCP
	MIO	FO	intercept	intercept	intercept	intercept	intercept	intercept
1st	(1,4.23)	(1,4.22)	intercept (1,4.22)	intercept (1,3.27)	intercept (1,3.38)	intercept (1,3.37)	intercept (1,3.41)	intercept (1,3.43)
2nd			married (0.2,0.03)	msmoke (0.5,-0.06)	medu3*B2(mage) (0.4,-10)	medu3*B2(mage) (0.4,-10)	medu3*B2(mage) (0.4,-10)	medu3*B2(mage) (0.4,-10)
3rd			deadkids*B15(mage) (0.2,-0.02)	married*B16(npri) (0.5,0.04)	fedu3*B1(mage) (0.3,-6.41)	fedu3*B1(mage) (0.3,-6.45)	fedu3*B1(mage) (0.3,-6.62)	fedu3*B1(mage) (0.3,-6.46)
4th			married*B2(fage) (0.1,-0.01)	fhisp*B15(mage) (0.5,-1.99)	fedu3*B2(fage) (0.3,-10)	fedu3*B2(fage) (0.3,-10)	fedu3*B2(fage) (0.3,-10)	fedu3*B2(fage) (0.3,-10)
5th			married*B13(npri) (0.1,0.02)	married*B17(npri) (0.4,0)	B9(mage) (0.2,0.04)	B9(mage) (0.2,0.05)	married*B16(npri) (0.2,0.07)	mhisp*B19(mage) (0.2,-10)
6th			mhisp*B8(mage) (0.1,-0.01)	mhisp*B13(npri) (0.4,-4.99)	married*B15(mage) (0.2,-0.06)	married*B15(mage) (0.2,-0.06)	mhisp*B19(mage) (0.2,-10)	mhisp*B13(npri) (0.2,-10)
7th			mhisp*B15(npri) (0.1,0)	fedu2*B9(mslb) (0.4,0.01)	married*B9(mslb) (0.2,0.05)	married*B16(npri) (0.2,0.04)	mhisp*B13(npri) (0.2,-10)	fhisp*B14(mage) (0.2,-10)
8th			alcohol*B10(mage) (0.1,0)	fedu3*B19(mslb) (0.4,-0.28)	mhisp*B19(mage) (0.2,-10)	married*B9(mslb) (0.2,0.05)	fhisp*B14(mage) (0.2,-10)	deadkids*B9(mage) (0.2,0.02)
9th			alcohol*B11(mslb) (0.1,0.03)	medu3*B2(mage) (0.4,-10)	mhisp*B13(npri) (0.2,-10)	married*B12(mslb) (0.2,0.04)	msmoke*B19(mage) (0.2,-5.06)	deadkids*B14(fage) (0.2,0.04)
10th			alcohol*B14(mslb) (0.1,-0.02)	married*B5(mage) (0.3,0)	fhisp*B14(mage) (0.2,-10)	mhisp*B19(mage) (0.2,-10)	fbaby*B14(mage) (0.2,0.05)	msmoke*B19(mage) (0.2,-5.12)

For each parenthesized pair of values, the left value shows the proportion of the variable being selected across the sample splitting replications in the conormalized quantile regression procedure. The right value shows the corresponding averaged estimated regression coefficient value over those cases where the variable has been selected.