Pseudo-Bayesian Approach for Quantile Regression Inference: Adaptation to Sparsity

Yuanzhi Li and Xuming He

Department of Statistics, University of Michigan

Supplementary Material

The Supplementary Material contains eight sections. Section S1 contains some auxiliary details for the simulation in the paper. Section S2 presents a new simulation setting with increasing covariate dimensions. The remaining parts are technical analysis. We first review some notations and preliminary lemmas in Section S3. We prove the results in Section 3 in Sections S4 and S5 where the covariate dimension p is fixed. In Section S6, we prove Theorem 3 in the asymptotic regime where p diverges with the sample size n. The proofs for some auxiliary technical lemmas are collected in Sections S7 and S8.

Contents

S1 Supplementary details for the simulation in Section 5

In this section, we present some further results for the simulation study in Section 5 of the paper, as well as some figures to corroborate the conclusions in the paper.

S1.1 Implementation details

First, we provide our implementations of the competing approaches in the paper. For posterior sampling of our Bayesian approach, we use the Gibbs sampler given in Alhamzawi et al. (2012) with a burn-in period of $2,000$ and the chain length of $12,000$. The effective sample size from the posterior chain is over $1,000$ for the slope coefficients. For the rank-score method, we rely on the `R` package '*quantreg*' (Koenker, 2018). For the wild bootstrap approach, we follow Wang et al. (2018) to use the two-point weight distribution (see Remark 1 therein) with 800 bootstrap samples.

Next, we consider the tuning parameter λ for both the shrinkage prior and the adaptive lasso model selection. To make a fair comparison, in Table 1 of the paper we compare the performances of those methods under a same value of λ that is also kept fixed across all simulated data sets. The ad-hoc value of λ used there is obtained as follows. From the 10-fold

cross-validation for adaptive lasso (Wu and Liu, 2009) on 100 data sets generated under the same model, we select λ at the 0.4th quantile of those values. This tuning parameter is then used in all simulated data sets in Table 1. On the other hand, in Figure 2 and S3 of the paper, we vary λ at a wide range of values and compare the performances in Figure 2 and also Figure S3. Note, however, that for each value of λ, we keep it fixed across all the shrinkage-based approaches and the 2,000 Monte Carlo data sets.

S1.2 Stability of the pseudo-Bayesian approach

Here we give a few supporting figures for the simulation results in the main paper; the purpose is to demonstrate the stability of our approach with respect to tuning parameter selection. When the tuning parameter λ is kept fixed, i.e., in the setting of Table 1, Figure S1 depicts the finite-sample stability of the adjusted posterior inference approach, as observed from the standard errors in Table 1 of the paper. Focusing on an inactive coefficient β_6, Figure S1 shows the average interval lengths in two scenarios, depending on whether the adaptive lasso picks up X_6 or not. The approach based on the adaptive-lasso (AL) selected model shows distinct interval lengths in those scenarios: they are either zero or similar to those based on the full model. On the other hand, the adjusted posterior inference is relatively

stable across two scenarios. Thus, the finite-sample stability of the adjusted posterior inference reflects its avoidance of pursuing dichotomous variable selection.

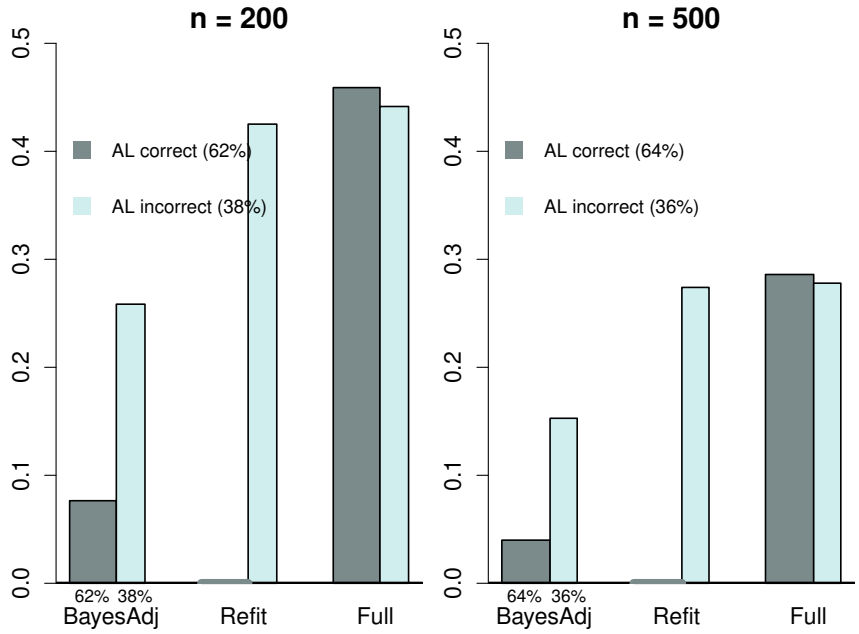

Figure S1: The average interval lengths for an inactive coefficient β_6 in the setting of Table 1 of the paper, where we separate the scenarios into two cases: (i) the Adaptive Lasso (AL) correctly identifies β_6 as 0; and (ii) the AL selection is incorrect for β_6. The numbers in the parentheses show the percentage of time that the AL selection is correct for β_6. For a descriptions of the abbreviated methods in the comparison (Full, BayesAdj, Refit), see the caption of Table 1 in the paper.

Next we present the performance of the pseduo-Bayesian approach, together with the approach based on frequentist variable selection, when λ

varies. Figure S2 shows the model selection performances of the adaptive lasso for a wide range of λ. The dashed vertical line marks the λ value used in Table 1 of the paper. We observe that the performance varies significantly with different choices of λ. In addition, Figure S3 presents the counterpart to Figure 2 in the main paper when the sample size $n = 200$; we observe similar pattern as in Figure 2. Our pseudo-Bayesian approach, 'BayesAdj' appears to be relatively stable compared to other methods.

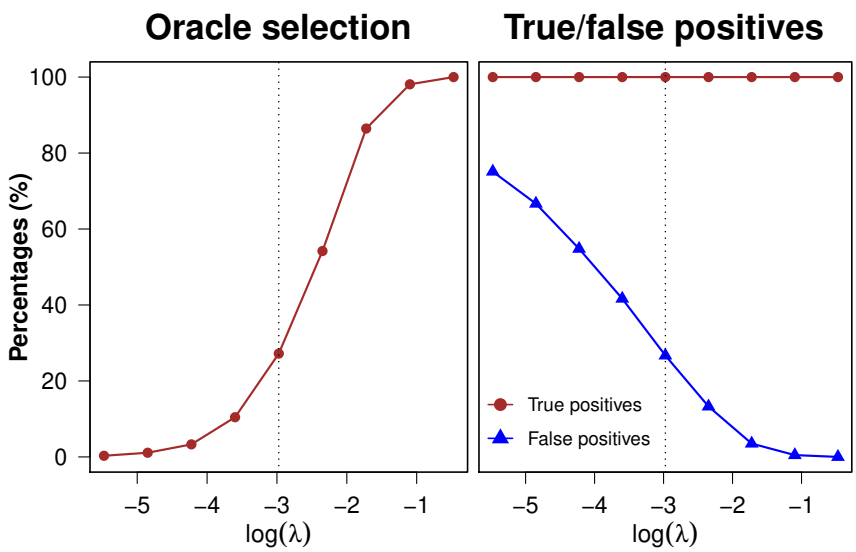

Figure S2: Variable selection performances of the adaptive lasso quantile regression when $n = 500$. Left panel: the percentage of time among the $2,000$ simulation that achieves exact oracle selection. Right panel: the average true positive rates (out of 2 active coefficients) and false positive rates (out of 4 inactive ones). The value of λ marked by a vertical broken line is used in Table 1 of the paper, which corresponds to an oracle selection percentage of 27.2% and an average false positive rate of 26.7%.

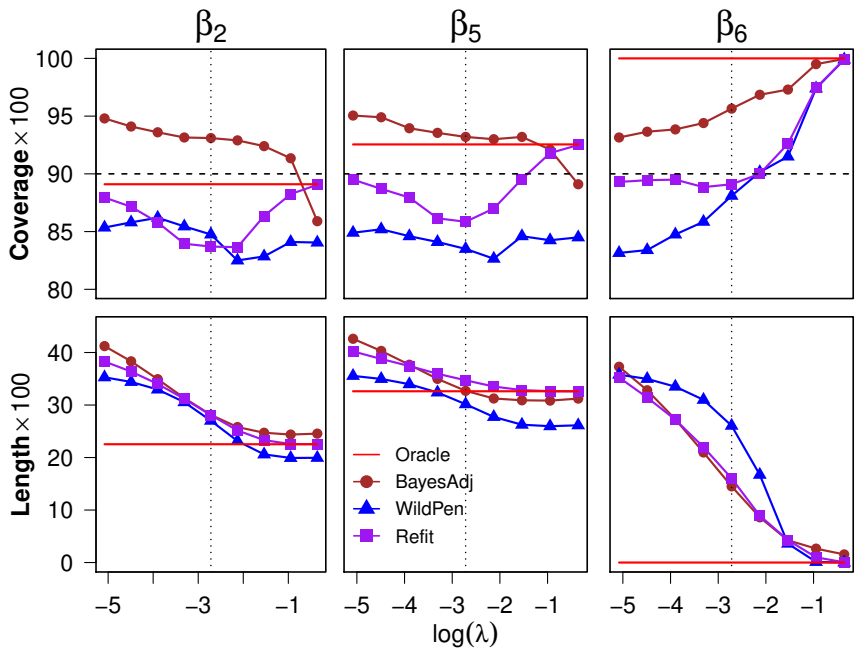

Figure S3: Empirical coverage probabilities and average lengths for 90% confidence intervals with different λ when $n = 200$. The true regression coefficients are $\beta_2^0 = 3$, $\beta_5^0 = -5$ and $\beta_6^0 = 0$. Refer to Table 1 in the paper for the value of λ marked by a vertical broken line and the abbreviated methods' names.

Table 1: Frequentist variable selection performances.

	Oracle (%)	False Pos.	False Neg.	# Selected
$n = 200$				
AL	53.80	0.93 (1.21)	0.00 (0.00)	3.93 (1.21)
SCAD	65.20	0.76 (1.25)	0.00 (0.00)	3.76 (1.25)
Lasso	11.00	2.19 (1.33)	0.00 (0.00)	5.19 (1.33)
$n = 500$				
AL	55.70	0.88 (1.19)	0.00 (0.00)	3.88 (1.19)
SCAD	67.65	0.66 (1.17)	0.00 (0.00)	3.66 (1.17)
Lasso	12.55	2.15 (1.35)	0.00 (0.00)	5.15 (1.35)

For the measures, 'Oracle' refers to the empirical probability that a method selects the true model $\{x_2, x_5\}$. 'False Pos.' is the average number of false positives, i.e., inactive covariates but selected into the model; 'False Neg.' means the reverse. '# Selected' is the average size of the selected model. For the methods, 'AL' stands for the adaptive lasso penalty and 'SCAD' for the smoothly clipped absolute deviation penalty. The tuning parameter is selected by 10-fold cross-validation. The numbers shown in the parentheses are the empirical standard deviations.

S1.3 More on variable selection performances

As noted by referees, the performance of the 'Refit' approach in Section 5 of the paper hinges on the success of variable selection. Under the same simulation setting in the paper, here we present some more results regarding the 'Refit' approach when we use other approaches for variable selection, including the lasso, adaptive lasso, and SCAD. To pursue better finite-sample performances, here, the tuning parameter is selected by 10-fold cross-validation as in Wu and Liu (2009), instead of using a fixed one in the paper. These approaches are implemented in the `R` package *rqPen* (Sherwood and Maidman, 2022).

Table 1 suggests that in practice, the variable selection approaches are often conservative and cannot achieve 'oracle' selection for quantile regression with heterogeneity in data. Such observations are consistent with the empirical findings in Wu and Liu (2009), even though the adaptive lasso and SCAD penalty were shown to achieve oracle selection in theory. In addition, there is non-negligible uncertainty induced by the variable selection procedure, as shown from the empirical standard errors in Table 1. We observe similar phenomena if we have used the same fixed tuning parameter as in the paper.

Figure S4 presents the performance of the 90% confidence intervals from different versions of the 'Refit' approach in the paper, where we apply the robust rank-score inference after different variable selections in Table 1. All the 'Refit' approaches are quite different from the 'Oracle' approach; in particular, the coverage probabilities are insufficient for the active coefficient β_2.

S1.4 About other priors

As a referee pointed out, many other shrinkage priors exist in the Bayesian literature. Under the same simulation setting in Section 5 of the paper, we report the inference performance under three other Bayesian priors: the

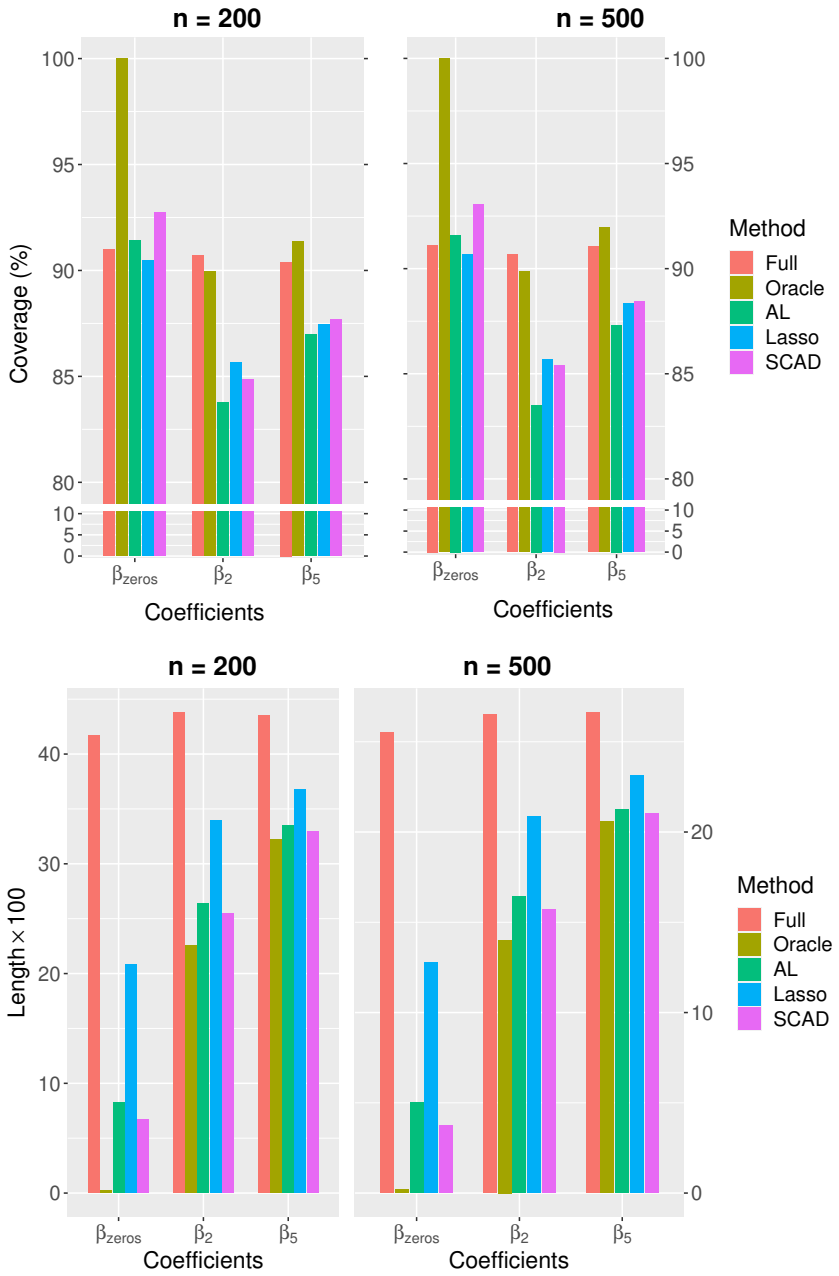

Figure S4: The empirical coverage probabilities (top) and average lengths (bottom) of the 90% confidence intervals from the 'Refit' approach, where we use different variable selection approaches as in Table 1 and the tuning parameter selected by 10-fold cross-validation. The result for β_{zeros} is an average over β_1, β_3, β_4, and β_6.

horseshoe (HS) prior (Kohns and Szendrei, 2020), the hierarchical adaptive Lasso prior (HAL) (Alhamzawi et al., 2012), and the discrete spike-and-slab prior (SSVS) (Chen et al., 2013), all of which use the same asymmetric Laplace working likelihood as in our paper. We include two methods for constructing confidence intervals under each of the prior choice: (i) we use direct normal approximation for the posterior distribution and; (ii) we use the adjusted posterior variance to form Wald-type intervals as in Section 3.3 of our paper. Note that we do not pursue variable selection with the priors. The HS approach is implemented in Matlab using the code provided by the author of Kohns and Szendrei (2020), the HAL and the SSVS approaches are available from `R` packages *Brq* (Alhamzawi, 2018) and *MCMCpack* (Martin et al., 2011), respectively. We include $12,000$ MCMC samples with a burn-in period of $2,000$ for all the MCMC sampling algorithms.

Table 2 shows that other Bayesian priors are not directly suitable for adjusted posterior inference. We also include two benchmark approaches taken from Table 1 in the paper, where we conduct frequentist inference under the full model and the oracle model, separately. Except for the unadjusted HAL prior, all other Bayesian approaches do not give accurate inference when judged by frequentist measures. The SSVS prior tends to be too conservative while the HS prior tends to be too liberal; the phenomena

hold with or without the variance adjustment. The unadjusted HAL prior provides sufficient coverage probability in this example, yet we demonstrate in Section S2 that unadjusted posterior inference is not valid in general. These empirical results suggest that not all shrinkage priors are ready to use in the pseudo-Bayesian framework for quantile regression inference. For the posterior inference to be valid in the frequentist sense, we need to carefully study the asymptotic property of both the posterior mean and the variance adjustment. While we established such property under the simple priors in the paper, it remains a challenging problem to systematically derive the theory under general shrinkage priors.

Here we further comment on the performance of adjusted posterior inference under the priors in Table 2. The relatively poor inferential performances originate from the mismatch of shrinkage on the posterior mean and the posterior variance. The relative scale of the posterior variance can be measured by the lengths of intervals in Table 2; the accuracy of the posterior mean is reported in Table 3. Under the SSVS prior, the sampling variation of the posterior mean is relatively small, close to the frequentist oracle estimator; however the adjusted posterior variance is still relatively large, similar to that using the full model. On the other hand, the HS and HAL priors shrink the posterior variance too much; the adjusted intervals in

Table 2: Empirical coverage probabilities (%) and average lengths ($\times 100$) for 90% confidence intervals

	Empirical coverage			Average length (s.e.)		
			$n = 200$			
	β_2	β_5	β_{zeros}	β_2	β_5	β_{zeros}
Full	92	91	90	43.7 (0.25)	43.9 (0.25)	41.7 (0.10)
Oracle	89	93	100	22.5 (0.13)	32.6 (0.21)	0.0 (0.00)
SSVS	100	100	100	60.4 (0.13)	63.6 (0.13)	37.6 (0.12)
SSVSAdj	100	100	100	45.3 (0.16)	52.1 (0.18)	17.5 (0.13)
HS	53	53	61	22.1 (0.24)	21.8 (0.23)	19.5 (0.16)
HSAdj	24	23	25	9.3 (0.23)	9.1 (0.22)	8.2 (0.17)
HAL	93	94	95	39.6 (0.11)	39.6 (0.11)	34.4 (0.07)
HALAdj	55	55	57	16.2 (0.08)	16.3 (0.08)	14.0 (0.06)
			$n = 500$			
	β_2	β_5	β_{zeros}	β_2	β_5	β_{zeros}
Full	91	91	90	26.7 (0.12)	26.6 (0.12)	25.6 (0.06)
Oracle	90	93	100	14.0 (0.06)	20.5 (0.11)	0.0 (0.00)
SSVS	100	100	100	33.8 (0.06)	36.4 (0.07)	18.4 (0.07)
SSVSAdj	100	100	100	24.8 (0.07)	30.0 (0.09)	6.9 (0.06)
HS	47	43	52	13.3 (0.15)	13.0 (0.14)	11.7 (0.10)
HSAdj	20	18	21	5.5 (0.15)	5.3 (0.14)	4.8 (0.10)
HAL	91	92	92	24.7 (0.05)	24.8 (0.05)	21.8 (0.03)
HALAdj	52	50	53	10.0 (0.04)	10.0 (0.04)	8.9 (0.03)

'SSVS', 'HS' and 'HAL' refers to direct posterior inference using normal approximation; The corresponding methods with 'Adj' refer to the confidence interval constructed operationally from the adjustment in Section 3.3 of the paper; For the acronym of other methods, refer to Table 1 in the main paper. The column β_{zeros} averages over all inactive coefficients β_1, β_3, β_4 and β_6. The numbers shown in the parentheses are the estimated standard errors. The Table can be viewed as an supplement to Table 1 in the main paper with more methods.

Table 2 are even narrower than those from the Oracle method. Therefore, it would require more careful study and variance adjustments for general shrinkage priors to achieve valid inference.

S1.5 On estimation accuracy

The focus of the paper is on inference but here we briefly examine the performance of posterior mean as the point estimator under the same simulation settings in Section 5 of the paper. We include two frequentist quantile regression estimators under the full model and the oracle model as benchmarks.

Table 3 reports the root mean squared error (RMSE) and mean absolute error (MAE) for the point estimates. We observe that the posterior mean under our choice of AL prior shows good estimation accuracy. The SSVS prior and HAL priors are also well-suited for point estimation; the posterior means are more efficient than the frequentist estimator under the full model whereas the SSVS prior outperforms the AL prior. The HS prior is less efficient than the full model estimates in this example.

S1.6 On weighted posterior inference

Here we examine the performance of weighted posterior inference discussed in Section 3.4 of the paper, where we use a weighted likelihood as in Equation (3.8) with the optimal weight

$$\zeta_i^* = f_{\varepsilon|X=x_i}(0).$$

Table 3: Estimation accuracy for frequentist and Bayesian approaches

	RMSE (×100)			MAE (×100)		
			$n = 200$			
	β_2	β_5	β_{zeros}	β_2	β_5	β_{zeros}
Full	12.58	12.48	11.81	9.95	9.89	9.39
Oracle	6.81	8.83	0.00	5.35	6.95	0.00
AL	8.41	8.83	5.62	6.55	7.02	3.40
SSVS	7.31	8.31	3.11	5.74	6.58	2.13
HS	14.89	15.21	13.41	11.59	12.07	10.02
HAL	10.76	10.77	9.15	8.48	8.54	7.15
			$n = 500$			
	β_2	β_5	β_{zeros}	β_2	β_5	β_{zeros}
Full	8.01	7.82	7.52	6.35	6.25	5.93
Oracle	4.20	5.65	0.00	3.35	4.50	0.00
AL	5.00	5.57	3.24	3.95	4.39	1.91
SSVS	4.30	5.31	1.55	3.41	4.25	0.96
HS	10.61	10.60	9.57	8.31	8.47	7.25
HAL	7.14	7.06	6.37	5.64	5.65	4.97

The 'Full' and 'Oracle' methods refer to the frequentist estimators as in Table 1 of the main paper. The 'AL' method refers to the posterior mean under our pseudo-Bayesian approach with AL prior in Section 2 of the main paper. 'SSVS', 'HS' and 'HAL' refer to the posterior means under other priors in Table 2. The column β_{zeros} averages over all inactive coefficients β_1, β_3, β_4 and β_6.

Since those weights are unknown in practice, we consider the following approach to estimating the weights:

$$\hat{\zeta}_i = \frac{2h}{\widehat{Q}_{\tau+h}(Y \mid X = x_i) - \widehat{Q}_{\tau-h}(Y \mid X = x_i)}, \tag{S1.1}$$

where $\widehat{Q}_{\tau\pm h}(Y \mid X = x)$ is an estimator of the conditional $\tau \pm h$ quantile of Y given $X = x$, and h is a bandwidth parameter; see Chapter 3.4 of Koenker (2005) for more discussions. We consider two ways of estimating $\widehat{Q}_{\tau\pm h}(Y \mid X = x)$ in our simulation setting: (i) linear quantile regression with all covariates $X_1, \ldots, X_6$ included; (ii) quantile regression with all the covariates and the quadratic function X_6^2 included. As for the bandwidth h, we use the one selected by the the `R` package '*quantreg*' (Koenker, 2018). To improve numerical stability, we left-winsorize the denominator in (S1.1) by a small constant 0.05. We note that the performance can be sensitive to the choice of the threshold, and further studies are needed for using estimated weights in posterior inference.

Table 4 shows that weighted posterior inference remains useful. However, the efficiency depends on how well the weights are estimated. While using the (unknown) optimal weights leads to improved efficiency, there are variations when the weights are estimated. In this case, the conditional quantile functions (other than the median) involve X_6^2, so if linear quantile regression is used to estimate the weights, we will see a clear loss of

efficiency. On the other hand, estimating the weights with the term X_6^2 included leads to improved inferential efficiency.

Table 4: Empirical coverage probabilities (%) and average lengths ($\times 100$) for 90% confidence intervals under different weighting schemes.

	Empirical coverage (%)			Average length (s.e.)		
			$n = 200$			
	β_2	β_5	β_{zeros}	β_2	β_5	β_{zeros}
No weight	92	92	96	28.5 (0.12)	32.8 (0.14)	11.8 (0.12)
True weight	95	94	97	24.9 (0.10)	28.5 (0.13)	10.6 (0.10)
Est. weight-L	94	91	97	41.6 (0.32)	46.5 (0.37)	18.8 (0.25)
Est. weight-Q	93	92	96	26.4 (0.18)	29.8 (0.18)	11.5 (0.13)
			$n = 500$			
	β_2	β_5	β_{zeros}	β_2	β_5	β_{zeros}
No weight	91	91	95	16.2 (0.05)	19.2 (0.06)	6.1 (0.06)
True weight	92	91	96	14.3 (0.05)	16.6 (0.05)	5.5 (0.05)
Est. weight-L	92	90	98	24.7 (0.13)	28.3 (0.17)	10.2 (0.11)
Est. weight-Q	91	90	95	14.8 (0.06)	17.1 (0.06)	5.8 (0.06)

All reported results are based on weighting of our adjusted posterior inference with AL prior, where all intervals are operationally constructed using the same procedure in Section 3.3 of the paper. 'No weight' refers to the original adjusted posterior inference in Table 1 of the main paper, and 'True weights' refers to the optimally-weighted posterior inference as in Section 3.4 of the paper. The 'Est. weight-L' method uses linear quantile regression to estimate the weights in (S1.1) and 'Est. weight-Q' includes the term X_6^2 in the quantile regression for estimating the weights. The column β_{zeros} averages over all inactive coefficients β_1, β_3, β_4 and β_6.

S2 An additional simulation setting

Here we use another Monte Carlo experiment to demonstrate the finite-sample performance of our approach in higher dimensions and more extreme quantile levels. We also compare with posterior inference under other priors in Section S1.4.

We consider a setting where the number of available covariates p grows with the sample size n. We consider the following model:

$$Y = 1 + X^T \beta^0 + \varepsilon, \tag{S2.2}$$

where $X \in \mathbb{R}^p$ and $X \sim \mathrm{N}(\mathbf{0}, \Sigma)$ with $\Sigma(i,j) = 0.5^{|i-j|}$, $\varepsilon \sim N(0, 3^2)$ independent of X, and

$$\beta^0 = (2,\ 0\underbrace{,\ldots,\ 0}_{\lfloor p/2 \rfloor - 2},\ 3,\ -5,\ 2,\ 0\underbrace{,\ldots,\ 0}_{\lceil p/2 \rceil - 2})^T.$$

Besides the intercept, there are 4 active covariates in the quantile regression model, with the other $p-4$ covariates being inactive. We focus on three different combinations of the pair (n, p): $(150, 8)$, $(400, 15)$ and $(800, 25)$; at each sample size we consider three quantile levels $\tau = 0.3$, 0.5 and 0.9. We generate $2,000$ Monte Carlo replications for each setting.

We first compare the results of our adjusted posterior inference with frequentist inferential methods using the full or oracle model; the results are summarized in Table 5. We find that our adjusted posterior inference approach can still achieve valid and adaptive inference across different covariate dimensions and quantile levels. The performance of our approach is reasonably close to oracle approach as if we knew the true model.

Next, we examine the posterior inference under the HAL and the SSVS priors in Section S1.4, where we report two methods for constructing con-

Table 5: Empirical coverage probabilities and average lengths ($\times 10$) for 90% confidence intervals under Model (S2.2).

		Empirical coverage			Average length (s.e.)		
		β_0	β_{active}	β_{zeros}	β_0	$\beta_{actives}$	β_{zeros}
$n = 150$, $p = 8$							
$\tau = 0.3$	Full	91	90	91	11.4 (0.06)	14.1 (0.09)	14.1 (0.09)
	Oracle	90	89	100	10.9 (0.05)	12.5 (0.08)	0.0 (0.00)
	BayesAdj	92	89	98	11.2 (0.05)	13.1 (0.07)	4.2 (0.09)
$\tau = 0.5$	Full	91	90	89	10.6 (0.05)	13.2 (0.08)	13.2 (0.08)
	Oracle	90	90	100	10.3 (0.05)	11.8 (0.07)	0.0 (0.00)
	BayesAdj	92	89	97	10.5 (0.04)	12.5 (0.06)	4.2 (0.09)
$\tau = 0.9$	Full	94	93	92	15.6 (0.09)	20.1 (0.15)	20.2 (0.16)
	Oracle	92	91	100	14.7 (0.09)	17.4 (0.14)	0.0 (0.00)
	BayesAdj	95	91	97	15.0 (0.07)	18.2 (0.12)	6.9 (0.13)
$n = 400$, $p = 15$							
$\tau = 0.3$	Full	92	91	90	6.9 (0.03)	8.5 (0.04)	8.6 (0.04)
	Oracle	90	90	100	6.5 (0.03)	7.5 (0.04)	0.0 (0.00)
	BayesAdj	92	90	98	6.6 (0.02)	7.8 (0.03)	1.8 (0.04)
$\tau = 0.5$	Full	92	90	90	6.6 (0.03)	8.0 (0.04)	8.1 (0.04)
	Oracle	91	90	100	6.2 (0.02)	7.2 (0.03)	0.0 (0.00)
	BayesAdj	92	89	99	6.3 (0.02)	7.4 (0.03)	1.5 (0.04)
$\tau = 0.9$	Full	93	92	92	9.5 (0.05)	11.8 (0.07)	12.0 (0.07)
	Oracle	91	90	100	8.6 (0.04)	9.9 (0.06)	0.0 (0.00)
	BayesAdj	93	91	98	8.8 (0.03)	10.3 (0.05)	2.7 (0.06)
$n = 800$, $p = 25$							
$\tau = 0.3$	Full	93	90	91	5.0 (0.02)	6.0 (0.02)	6.1 (0.03)
	Oracle	90	90	100	4.6 (0.02)	5.3 (0.02)	0.0 (0.00)
	BayesAdj	91	90	99	4.6 (0.01)	5.4 (0.02)	0.9 (0.02)
$\tau = 0.5$	Full	91	90	90	4.6 (0.02)	5.6 (0.02)	5.7 (0.02)
	Oracle	89	89	100	4.4 (0.01)	5.0 (0.02)	0.0 (0.00)
	BayesAdj	90	89	99	4.4 (0.01)	5.1 (0.02)	0.9 (0.02)
$\tau = 0.9$	Full	93	93	92	6.8 (0.03)	8.3 (0.04)	8.5 (0.04)
	Oracle	90	90	100	6.0 (0.03)	6.9 (0.04)	0.0 (0.00)
	BayesAdj	92	90	99	6.1 (0.02)	7.2 (0.03)	1.5 (0.03)

The column β_0 is for the intercept term, $\beta_{actives}$ averages over the 4 active coefficients, and β_{zeros} averages over other inactive (zero) coefficients. The numbers shown in the parentheses are the estimated standard errors. For the coverage estimates, their standard errors are all below 0.9. The abbreviation of methods' names are the same as Table 1 in the main paper. Using the procedure in Section S1.1, we fix the tuning parameter λ across all simulated data sets for our BayesAdj method.

fidence intervals as discussed in Section S1.4. We focus on the setting with $n = 400$ and $p = 15$ and examine the performance at different quantile levels in Table 6; we also include the frequentist method under the full or oracle model as benchmarks. In this numerical setting, we find that posterior inference under the HAL prior, either adjusted or unadjusted, is not valid in general; its performance varies significantly across quantile level. On the other hand, the adjusted posterior inference under the SSVS prior is still conservative, similar to what we find in Section S1.4. To conclude, the results here and those in Section S1.4 suggest that the proper use of common shrinkage priors requires more careful study in our setting.

S3 Review of some preliminary results

S3.1 Notations

Throughout the Supplementary Materials, we shall assume that: (i) there is no intercept term in the quantile regression model, and (ii) the covariate vector X is centered and has mean 0. Correspondingly, the covariate vector $X = (X_1, \ldots, X_p)^T$ and the true regression coefficient $\boldsymbol{\beta}^0 = (\beta_1^0, \ldots, \beta_p^0)^T \in \mathbb{R}^p$, instead of $\mathbb{R}^{p+1}$ presented in the main paper. We work under these two assumptions merely to simplify the presentation of the technical details.

Table 6: Empirical coverage probabilities and average lengths ($\times 10$) for 90% confidence intervals under Model (S2.2) with $n = 400$ and $p = 15$.

		Empirical coverage			Average length (s.e.)		
		β_0	β_{active}	β_{zeros}	β_0	$\beta_{actives}$	β_{zeros}
$\tau = 0.3$	Full	92	91	90	6.9 (0.03)	8.5 (0.04)	8.6 (0.04)
	Oracle	90	90	100	6.5 (0.03)	7.5 (0.04)	0.0 (0.00)
	SSVS	90	90	98	6.1 (0.01)	7.4 (0.02)	4.7 (0.04)
	SSVSAdj	100	99	100	10.6 (0.03)	12.9 (0.05)	5.9 (0.10)
	HAL	80	83	88	5.0 (0.01)	6.2 (0.01)	5.8 (0.01)
	HALAdj	93	94	96	7.2 (0.02)	8.9 (0.04)	7.6 (0.03)
$\tau = 0.5$	Full	92	90	90	6.6 (0.03)	8.0 (0.04)	8.1 (0.04)
	Oracle	91	90	100	6.2 (0.02)	7.2 (0.03)	0.0 (0.00)
	SSVS	90	89	98	5.7 (0.01)	7.0 (0.01)	4.4 (0.04)
	SSVSAdj	100	99	100	10.2 (0.03)	12.3 (0.05)	5.7 (0.10)
	HAL	86	84	89	5.0 (0.01)	6.3 (0.01)	5.8 (0.01)
	HALAdj	97	96	98	7.9 (0.02)	9.7 (0.04)	8.3 (0.03)
$\tau = 0.9$	Full	93	92	92	9.5 (0.05)	11.8 (0.07)	12.0 (0.07)
	Oracle	91	90	100	8.6 (0.04)	9.9 (0.06)	0.0 (0.00)
	SSVS	94	94	99	8.7 (0.01)	10.5 (0.02)	6.8 (0.05)
	SSVSAdj	100	99	100	14.1 (0.04)	17.2 (0.08)	7.8 (0.11)
	HAL	70	70	79	4.9 (0.01)	6.2 (0.02)	5.7 (0.02)
	HALAdj	68	68	77	4.8 (0.02)	6.2 (0.04)	5.2 (0.03)

The format of the table is the same as Table 5. For acronym and implementation of the priors, see Section S1.4.

Since we do not impose a shrinkage prior on the intercept term in the paper, the explicit form of the prior distribution does not change when we omit the intercept term. In fact, all the proofs can go through if we add the intercept term $X_0 = 1$ back to the covariate vector.

Among the p-dimensional covariates, we assume $\mathcal{S} = \{1, \ldots, s\}$ is the index set of the active (non-zero) coefficients. For a vector $\boldsymbol{v} = (v_1, \ldots, v_p)^T$, we let $\boldsymbol{v}_{\mathcal{S}} = \{v_j : \ j \in \mathcal{S}\}$ and $\boldsymbol{v}_{\mathcal{S}^c} = \{v_j : \ j \notin \mathcal{S}\}$. For a matrix $A \in \mathbb{R}^{p \times p}$, we partition

$$A = \begin{bmatrix} A_{\mathcal{S}} & A_{\mathcal{S},\mathcal{S}^c} \\ A_{\mathcal{S}^c,\mathcal{S}} & A_{\mathcal{S}^c} \end{bmatrix},$$

where $A_{\mathcal{S}} \in \mathbb{R}^{s \times s}$; we write $A(i,j)$ as the (i,j)th entry of A.

We introduce the following set of notation in the Supplementary Materials. We use bold lowercase letters to denote (column) vectors, and we reserve bold capital letters for matrices. Specifically, let $\boldsymbol{I}_q$ be the q by q identity matrix, and $\boldsymbol{e}_j$ be the jth unit vector where only the jth entry is 1. Let $\boldsymbol{X}$ be such that $\boldsymbol{X}^T = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \mathbb{R}^{p \times n}$ be the design matrix, and let $\boldsymbol{X}_{\mathcal{S}} \in \mathbb{R}^{n \times s}$ denote the design matrix for the active covariates. Recall the quantile-loss function is $\rho_\tau(u) = u\{\tau - 1[u \leq 0]\}$ and we write $L_n(\boldsymbol{\beta}) = \sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})$. Let $\phi_\tau(u) = \tau - 1[u \leq 0]$ and we write $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^T$, where $\phi_i = \phi_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}^0)$. We define $\Delta_p = G^{-1} \boldsymbol{X}^T \boldsymbol{\phi}$ and $\Delta_s = G_{\mathcal{S}}^{-1} \boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{\phi}$, where the matrix G is defined in Section 3.1 of the

paper.

In addition, we shall continue to use the notations in Section 3.1 of the paper, and we review some of them here. Let pr^* be the true data generating probability, and let E^* be the expectation under pr^*. The posterior probability is $\Pi(\cdot \mid \mathbb{D}_n)$. Conversely, we shall use $\mathrm{pr}(\cdot)$ and $\mathrm{E}(\cdot)$ for generic probability calculations. Recall $\hat{\boldsymbol{\beta}} \in \mathbb{R}^p$ is the classic quantile regression estimator and $\tilde{\boldsymbol{\beta}}_{\mathcal{S}} \in \mathbb{R}^s$ is the oracle estimator. For any symmetric matrix A, we define $\theta_{\max}(A)$ and $\theta_{\min}(A)$ as the maximal/minimal eigenvalue of A. For stochastic sequences A_n and B_n, we define $A_n \leq_{\mathrm{pr}^*} B_n$ if $A_n \leq B_n$ holds with pr^*-probability going to 1.

S3.2 Some standard probabilistic results

In this section, we state and prove some useful lemmas. Let $\chi^2_d(\nu)$ represent the chi-square distribution with d degrees of freedom and non-centrality parameter ν; let Laplace(b) represent the Laplace distribution with density function

$$f_b(x) = \frac{1}{2b} \exp\left\{-\frac{|x|}{b}\right\}, \quad x \in \mathbb{R};$$

and let $N(\mu, \Sigma)$ represent the multivariate normal distribution. We first present Lemma 1–3 regarding the properties for those distributions.

Lemma 1. *Let $X \sim \chi^2_d(\nu)$, then we have for some $C_0, C_1 > 0$,*

$$\mathrm{E}[X^k] \leq C_0 \times (v^k + C_1 d^k), \quad k \in \{0, \ldots, 4\},$$

and for all $x \geq 4(d + 2\nu)$,

$$P(X \geq x) \leq \exp(-x/4)$$

Furthermore, let $Z \sim \mathrm{N}(\mu, \Sigma) \in \mathbb{R}^d$, then if $x^2 \geq 4\theta_{\max}(\Sigma) \cdot (d + 2\mu^T \Sigma^{-1} \mu)$, we have

$$P(\|Z\| \geq x) \leq \exp\left(-\frac{x^2}{4\theta_{\max}(\Sigma)}\right).$$

Proof. The first two inequalities follows from standard bounds for chi-square distributions; see e.g., Lemma 8.1 of Birgé et al. (2001). To show the third inequality, note that $Z^T \Sigma^{-1} Z \sim \chi^2_d\left(\mu^T \Sigma^{-1} \mu\right)$; the conclusion follows from the second inequality. □

Lemma 2. *Given $a_1, \ldots, a_m \geq a_{\min} > 0$, let $X_1, \ldots, X_m$ be independent random variables with*

$$X_j \sim \textit{Laplace}\left(\frac{1}{a_j}\right), \quad j = 1, \ldots m,$$

Then, for all $x > 0$, we have

$$\mathrm{pr}\left(\max_{j=1,\ldots m} |X_j| \geq x\right) \leq m \cdot \exp\{-a_{\min} \cdot x\},$$

and for some constant $C_0 > 0$,

$$\mathrm{E}\left(\sum_{j=1}^{m} X_j^2\right)^k \leq C_0 \times \left(\frac{m}{a_{\min}^2}\right)^k, \quad k \in \{0, 1, 2\}.$$

Proof. The first inequality follows from a standard union bound, since

$$\mathrm{pr}(|X_j| \geq x) = 2\int_x^\infty \frac{a_j}{2} \exp\{-a_j|u|\}\,\mathrm{d}u \leq \exp\{-a_{\min} \cdot x\}.$$

The second inequality for $k = 0$ is straightforward; for $k = 1$ and 2, note $\mathrm{E}[X_j^2] = 2a_j^{-2}$, and $\mathrm{E}[X_j^4] = 24a_j^{-4}$. The inequality then follows by independence across X_j, $j \in \{1, \ldots, m\}$. □

Lemma 3. *Let $w = (w_1, \ldots, w_s)^T$, and let $X \in \mathbb{R}^s$ be distributed as*

$$X \sim N\left(\mu, \frac{\sigma_0^2}{n} I_s\right).$$

For any $0 < \varepsilon < 1/2$, if

$$\|w\| \leq \varepsilon \times \left\{\frac{\sqrt{2n}}{\sigma_0} \wedge \frac{1}{\|\mu\|}\right\},$$

then we have

$$\left|\mathrm{E}_X\left(\exp\{-w^T X\}\right) - 1\right| \leq 4\varepsilon.$$

Furthermore, if K satisfies $K \geq 3\sigma_0$ and $K^2 s \geq 16n\|\mu\|^2$, we have

$$\mathrm{E}_X\left(\exp\{-w^T X\} \cdot \mathbf{1}\left[\|X\| \geq K\sqrt{\frac{s}{n}}\right]\right) \lesssim \exp\left(-\frac{K^2 s}{8\sigma_0^2}\right).$$

Proof. By leveraging the moment generating function of the normal distribution, and using the upper bound for $\|w\|$,

$$\begin{aligned} \mathrm{E}\left(\exp\{-w^T X\}\right) &= \exp\left\{\mu^T w + \frac{\sigma_0^2}{2n}\|w\|^2\right\} \\ &\leq \exp\{2\varepsilon\} \\ &\leq 1 + 4\varepsilon, \end{aligned}$$

for $0 < \varepsilon < 1/2$. In a similar manner, we can establish the lower bound for $\mathrm{E}\left(\exp\{-w^T X\}\right)$, which shows the first result.

For the second inequality, using Cauchy-Schwartz inequality gives

$$\begin{aligned} & \mathrm{E}^2\left(\exp\{-w^T X\} \cdot 1\left[\|X\| \geq K\sqrt{\frac{s}{n}}\right]\right) \\ \leq\ & \mathrm{E}\left(\exp\{-2w^T X\}\right) \cdot \mathrm{pr}\left(\|X\| \geq K\sqrt{\frac{s}{n}}\right) \\ \leq\ & \exp\{4\varepsilon\} \cdot \exp\left\{-\frac{K^2 s}{4\sigma_0^2}\right\}, \end{aligned}$$

where the tail probability is bounded by Lemma 1. □

The following results are variants of the Bernstein inequality, and we include them here for completeness. They are Theorem 2.10 of Boucheron et al. (2013) and Theorem 2.8.2 of Vershynin (2018), respectively.

Lemma 4. *Let $X_1, \ldots, X_n$ be independent random variables. Suppose there exist positive constants $c, v > 0$ such that $\sum_{i=1}^n \mathrm{E}(X_i^2) \leq v$, and*

$$\sum_{i=1}^n \mathrm{E}\left[(X_i^q)_+\right] \leq \frac{q!}{2} v c^{q-2} \quad \textit{for all integers } q > 3.$$

Then

$$\mathrm{P}\left(\sum_{i=1}^{n}[X_i - \mathrm{E}(X_i)] \geq t\right) \leq 2\exp\left(-\frac{t^2}{2(v+ct)}\right).$$

Lemma 5. *Let $X_1, \ldots, X_n$ be independent, mean zero random variables that satisfy*

$$\sup_{i=1,\ldots,n} \mathrm{pr}\left(|X_i| \geq x\right) \leq \exp(-x/\sigma_0),$$

for some constant σ_0. Then there is a universal constant C_2, such that for every $t \geq 0$ and $a = (a_1, \ldots, a_n)$,

$$\mathrm{pr}\left(\left|\sum_{i=1}^{n} a_i X_i\right| \geq t\right) \leq 2\exp\left(-C_2 \cdot \min\left\{\frac{t^2}{\sigma_0^2\|a\|_2^2}, \frac{t}{\sigma_0\|a\|_\infty}\right\}\right).$$

The following lemma is simple but useful; we will implicitly use the lemma in the upcoming proofs.

Lemma 6. *Let $f(z;\theta)$ be a probability density function indexed by $\theta \in \Theta \subset \mathbb{R}^k$. We write $Z \sim f(z;\theta)$ and define $\mathrm{pr}_\theta(Z \geq A) = \int_{z \geq A} f(z;\theta)\mathrm{d}z$, where Z is independent of the data. Let $g(\cdot,\cdot)$ be a bivariate function of $\Theta \times \mathbb{R} \to \mathbb{R}$; suppose we have*

$$\sup_{\theta:\, g(\theta,A) \leq B} \mathrm{pr}_\theta(X \geq A) \leq a,$$

for some real numbers a, A, and B. For any statistic θ_n that satisfies $g(\theta_n, A) \leq_{\mathrm{pr}^*} B$, it holds that

$$\mathrm{pr}_{\theta_n}(X \geq A) \leq_{\mathrm{pr}^*} a.$$

Proof. Let $W = \{\theta \in \Theta : g(\theta, A) \leq B\}$. Note

$$\mathrm{pr}_{\theta_n}(X \geq A) \leq \mathrm{pr}_{\theta_n}(X \geq A) 1_{[\theta_n \in W]} + 1_{[\theta_n \notin W]}.$$

Since $\mathrm{pr}^*(\theta_n \in W) \to 1$, as $n \to \infty$, the proof is complete. □

S3.3 Discussion on the improper CA prior

Here we give some results under the CA prior (2.5) in the paper. Since the CA prior itself is improper, the posterior distribution is not automatically valid. The following Proposition formally shows that the posterior, which is constructed operationally with the Bayes rule, is still a valid distribution.

Proposition S1. *Consider the posterior density $p(\boldsymbol{\beta} \mid \mathbb{D}_n)$ under the CA prior. For fixed $p + 1 < n$, and the tuning parameter $\lambda > 0$. If $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ expands $\mathbb{R}^{p+1}$, then we have that*

$$\int_{\mathbb{R}^{p+1}} p(\boldsymbol{\beta} \mid \mathbb{D}_n) \mathrm{d}\boldsymbol{\beta} < +\infty.$$

Proof. Since the CA prior is uniformly bounded by 1, we have that

$$\begin{aligned} p(\boldsymbol{\beta} \mid \mathbb{D}_n) &\propto \exp\left\{ -\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}) - n \sum_{j=1}^{p} p_\lambda(\beta_j) \right\} \\ &\leq \exp\left\{ -\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}) \right\}. \end{aligned}$$

The last line corresponds to the posterior density under the improper flat prior (i.e., $\pi(\beta_j) \propto 1$) for all $p + 1$ coefficients. From Theorem 1 of Yu and

Moyeed (2001) we have

$$\int_{\mathbb{R}^{p+1}} \exp\left\{-\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\right\} \mathrm{d}\boldsymbol{\beta} < +\infty,$$

which completes the proof with the inequality for $p(\boldsymbol{\beta} \mid \mathbb{D}_n)$. □

S3.4 Other useful lemmas with increasing dimensions

In this subsection, we present some technical lemmas in the context of quantile loss function. We consider the asymptotic regime in Section 4 of the paper, i.e., the covariate dimension p grows with the sample size n. We rely on Assumptions 1, 2',3, and 4' in the paper, and we make any dependency on n and p explicit. For the asymptotic regime where p is fixed, note Assumptions 1 – 4 are implied by Assumptions 1, 2', 3 and 4'. Therefore the conclusions for the lemmas below hold for the fixed dimensional regime as well. Those lemmas may be of independent interest, and their proofs are deferred to Section S7.

Lemma 7. *Let* $L_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\boldsymbol{\beta})$ *and* $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$. *Suppose Assumption 2 and 4' in the main paper holds, then there exists a constant* $q_0 > 0$ *such that*

$$\frac{1}{n}\mathrm{E}^*\left[L_n(\boldsymbol{\beta}^0 + \boldsymbol{\delta}) - L_n(\boldsymbol{\delta})\right] \geq \min\left\{\frac{\|G^{1/2}\boldsymbol{\delta}\|^2}{4}, q_0\|G^{1/2}\boldsymbol{\delta}\|\right\}.$$

Lemma 8 (Stochastic Differentiability)**.** *Let* $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$ *and*

$$r_n(\boldsymbol{\delta}) = L_n(\boldsymbol{\beta}^0 + \boldsymbol{\delta}) - L_n(\boldsymbol{\beta}^0) + \boldsymbol{\phi}^T \boldsymbol{X} \boldsymbol{\delta}.$$

Suppose Assumptions 1, 2', 3 and 4' in the paper hold, and $p^2 \log^2 p = o(n)$*. Then we have that*

$$\sup_{\boldsymbol{\delta} \in \mathbb{R}^p} \left| \frac{r_n(\boldsymbol{\delta}) - \mathrm{E}^*[r_n(\boldsymbol{\delta})]}{n\|D^{1/2}\boldsymbol{\delta}\|^2 + 1} \right| \to 0,$$

in pr^**-probability.*

Lemma 9 (Restricted Quadratic Expansion)**.** *Suppose Assumptions 1, 2', 3 and 4' in the paper hold and* $s^4 p^2 \log^2 n = o(n)$*. Furthermore, if* $\lambda = \lambda_n$ *satisfies*

$$\lambda_n \gg \frac{\sqrt{s} p \log p}{\sqrt{n}},$$

then we have

$$L_n(\boldsymbol{\beta}^0 + \boldsymbol{\delta}) - L_n(\boldsymbol{\beta}^0) = \frac{n}{2} \boldsymbol{\delta}^T G \boldsymbol{\delta} - \boldsymbol{\phi}^T \boldsymbol{X} \boldsymbol{\delta} + o_{\mathrm{pr}^*}(1),$$

uniformly on $\boldsymbol{\delta} \in \mathcal{B}_n(K_n)$ *for some sequence* $K_n \to \infty$*, where*

$$\mathcal{B}_n(K_n) = \left\{ \boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}_{\mathcal{S}}\|_2 \leq K_n \sqrt{\frac{s}{n}} \, ; \, \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \leq K_n \frac{s \log p}{n\lambda} \right\}.$$

S4 Proof of Theorem 1

In this section, we focus on the Adaptive Lasso (AL) prior (2.4) in the paper. The analogous result under the Clipped Absolute (CA) prior follows from

Theorem 3, which we shall prove in the next section. To simplify notations, here we shall focus on the posterior distribution for the centered parameter $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$.

We define some additional notations specific for this section with the AL prior. Let $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0$ and $\tilde{\boldsymbol{\delta}}_{\mathcal{S}} = \tilde{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}^0_{\mathcal{S}}$. Let $\tilde{\boldsymbol{w}} = (\tilde{w}_1, \ldots, \tilde{w}_p)$ with $\tilde{w}_j = n^{1/2}\lambda/|\hat{\beta}_j|$. Define $T_n = \exp\{n\tilde{\boldsymbol{\delta}}_{\mathcal{S}}^T G_{\mathcal{S}} \tilde{\boldsymbol{\delta}}_{\mathcal{S}}/2 - \sum_{j\in\mathcal{S}} \tilde{w}_j |\beta_j^0|\}$. We partition $\tilde{\boldsymbol{w}} = (\tilde{\boldsymbol{w}}_{\mathcal{S}}^T, \tilde{\boldsymbol{w}}_{\mathcal{S}^c}^T)^T$, and define $w_{\min} = \min\{\tilde{w}_j : j \notin \mathcal{S}\}$ and $w_{\max} = \max\{\tilde{w}_j : j \notin \mathcal{S}\}$. Let θ_M and θ_m be the maximal/minimum eigenvalue of G, respectively. For $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$, we define

$$
\begin{aligned}
f_n(\boldsymbol{\delta}) &= T_n \times \exp\left\{ -\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}} - \tilde{\boldsymbol{\delta}}_{\mathcal{S}})^T G_{\mathcal{S}} (\boldsymbol{\delta}_{\mathcal{S}} - \tilde{\boldsymbol{\delta}}_{\mathcal{S}}) - \sum_{j\notin\mathcal{S}} \tilde{w}_j |\delta_j| \right\}, \\
p_n(\boldsymbol{\delta}) &= \exp\left\{ -\sum_{j=1}^{p} \tilde{w}_j |\beta_j^0 + \delta_j| \ + \ L_n(\boldsymbol{\beta}^0) - L_n\left(\boldsymbol{\delta} + \boldsymbol{\beta}^0\right) \right\},
\end{aligned}
\tag{S4.3}
$$

where $p_n(\delta)$ is proportional to the posterior density of δ under the AL prior, and $f_n(\delta)$ is proportional to the limiting density in Theorem 1.

S4.1 Part 1 of Theorem 1

We begin from part 1 of Theorem 1. We need the following technical lemma, the proof of which is deferred to Section S8.

Lemma 10. *Under the conditions of Theorem 1, and consider the AL prior.*

We have

$$\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \gtrsim_{\mathrm{pr}^*} \left\{\prod_{j\notin\mathcal{S}}\left(\frac{2}{\tilde{w}_j}\right)\right\}\cdot\left(\frac{2\pi}{n\theta_M}\right)^{s/2}\cdot\exp\left\{-\sum_{j\in\mathcal{S}}\tilde{w}_j|\beta_j^0| + \frac{n}{2}\Delta_s^T G_{\mathcal{S}}\Delta_s\right\},$$

where θ_M is the maximum eigenvalue of G.

Proof of Theorem 1, part 1. Without loss of generality, we assume that the true values of the active coefficients are all positive; the results under other scenarios hold automatically by symmetry. Let

$$\mathcal{B}_n(M_n) = \{\boldsymbol{\delta} : n^{1/2}\|\boldsymbol{\delta}_{\mathcal{S}}\|_2 \le M_n\ ,\ n\lambda\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \le M_n\}, \tag{S4.4}$$

and

$$A = \begin{bmatrix} (n^{1/2})\boldsymbol{I}_s & 0 \\ 0 & (n\lambda)\boldsymbol{I}_{p-s} \end{bmatrix} \in \mathbb{R}^{p\times p}. \tag{S4.5}$$

In the following, we shall prove for any $\alpha = 0, 1, 2$,

$$\frac{\displaystyle\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0, \tag{S4.6}$$

in pr*-probability, for any sequence $M_n \to +\infty$. Setting $\alpha = 0$ in (S4.6) recovers part 1 of Theorem 1, whereas the conclusion for $\alpha = 1, 2$ is useful for later results.

For the constant q_0 in Lemma 7 and some large enough constant $R > 0$

given below, we define the following regions for $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$:

$$\mathcal{A}_n = \left\{ \boldsymbol{\delta} : sgn(\boldsymbol{\beta}^0_{\mathcal{S}} + \boldsymbol{\delta}_{\mathcal{S}}) = sgn(\boldsymbol{\beta}^0_{\mathcal{S}}) \right\},$$

$$\mathcal{B}_n = \mathcal{B}_n(M_n),$$

$$\mathcal{C}_n = \left\{ \boldsymbol{\delta} : \|G^{1/2}\boldsymbol{\delta}\| \leq Rq_0 \right\},$$

where $sgn(\cdot)$ is the sign function and $\mathcal{B}_n(M_n)$ is defined before (S4.6). Note $\|G^{1/2}\boldsymbol{\delta}\|^2 = \boldsymbol{\delta}^T G \boldsymbol{\delta} \geq \theta_m \delta_j^2$; In view of Assumption 4, we can have $\mathcal{B}_n \subset \mathcal{C}_n \subset \mathcal{A}_n$ by choosing R large enough. Therefore, we decompose

$$\mathcal{B}_n^c \subset [\mathcal{B}_n^c \cap \mathcal{C}_n] \cap [\mathcal{C}_n^c].$$

To show (S4.6), it then suffices to upper bound $\int \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\, \mathrm{d}\boldsymbol{\delta}$ on each of the two areas on the right hand side above.

We divide our proof into five parts. In step I, we first give upper bounds for the posterior density $p_n(\boldsymbol{\delta})$; in steps II - II, we obtain upper bounds for the posterior integral $\int \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\, \mathrm{d}\boldsymbol{\delta}$ on each of the two areas $\mathcal{B}_n^c \cap \mathcal{C}_n$ and $\mathcal{C}_n^c$ separately; then in step IV, we show the posterior probabilities of those two areas are both $o_{\mathrm{pr}^*}(1)$; step V contains some auxiliary calculations to supplement the proof.

Step I: Bounding the posterior density

In this step, we give two different upper bounds for $p_n(\boldsymbol{\delta})$, depending on the value of $\boldsymbol{\delta}$.

When $\boldsymbol{\delta} \in \mathcal{C}_n^c$, by the convexity of L_n and Assumption 1:

$$\begin{aligned} L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta} + \boldsymbol{\beta}^0) &\leq \frac{n\|G^{1/2}\boldsymbol{\delta}\|}{Rq_0} \cdot \sup_{\|G^{1/2}\boldsymbol{\delta}\| \geq Rq_0} \left\{ \frac{L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta} + \boldsymbol{\beta}^0)}{n} \right\} \\ &\leq_{\mathrm{pr}^*} -\frac{n\varepsilon_0 \|G^{1/2}\boldsymbol{\delta}\|}{Rq_0}, \end{aligned}$$

uniformly in $\boldsymbol{\delta} \in \mathcal{C}_n^c$, for some constant ε_0 as in Assumption 1. Since $\pi(\boldsymbol{\delta} + \boldsymbol{\beta}^0) \leq 1$ for all $\boldsymbol{\delta}$, we have

$$\begin{aligned} p_n(\boldsymbol{\delta}) &\leq_{\mathrm{pr}^*} \exp\left\{ -\frac{n\varepsilon_0 \|G^{1/2}\boldsymbol{\delta}\|_2}{Rq_0} \right\} \\ &\triangleq \overline{p}_{1n}(\boldsymbol{\delta}), \end{aligned}$$

uniformly when $\boldsymbol{\delta} \in \mathcal{C}_n^c$.

Next we consider when $\boldsymbol{\delta} \in \mathcal{B}_n^c \cap \mathcal{C}_n$. With r_n defined in Lemma 8, we have

$$\begin{aligned} -r_n(\boldsymbol{\delta}) &\leq -\mathrm{E}^*[r_n(\boldsymbol{\delta})] + \left(\sup_{\boldsymbol{\delta} \in \mathcal{C}_n} \frac{|r_n(\boldsymbol{\delta}) - \mathrm{E}^*[r_n(\boldsymbol{\delta})]|}{n\|D^{1/2}\boldsymbol{\delta}\|^2 + 1} \right) \cdot (n\|D^{1/2}\boldsymbol{\delta}\|^2 + 1) \\ &\leq_{\mathrm{pr}^*} -\frac{n}{8}\boldsymbol{\delta}^T G \boldsymbol{\delta} + \frac{1}{8}\underline{f}, \end{aligned} \tag{S4.7}$$

uniformly on $\boldsymbol{\delta} \in \mathcal{C}_n$, where $\underline{f}$ is a constant introduced in Assumption 3; we use Lemma 8 to bound the centered empirical process $r_n(\boldsymbol{\delta}) - \mathrm{E}^*[r_n(\boldsymbol{\delta})]$, Lemma 7 to bound $\mathrm{E}^*[r_n(\boldsymbol{\delta})]$, and the fact that $\underline{f} \cdot D \preceq G$. Therefore, since $\boldsymbol{\delta}^T G \boldsymbol{\delta} \geq \theta_m \boldsymbol{\delta}_{\mathcal{S}}^T \boldsymbol{\delta}_{\mathcal{S}}$ under Assumption 2, we have:

$$\begin{aligned} L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\beta}^0 + \boldsymbol{\delta}) \leq_{\mathrm{pr}^*} & -\frac{n\theta_m}{2R} \left\| \boldsymbol{\delta}_{\mathcal{S}} - \frac{R}{2n\theta_m} \boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{\phi} \right\|^2 + \frac{R}{n\theta_m} \left\| \boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{\phi} \right\|^2 \\ & + \left\| \boldsymbol{X}_{\mathcal{S}^c}^T \boldsymbol{\phi} \right\|_\infty \cdot \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 + \frac{1}{2R}\underline{f}, \end{aligned}$$

which follows from completing the squares in (S4.7).

As for the prior term when $\boldsymbol{\delta} \in \mathcal{B}_n^c \cap \mathcal{C}_n$. Since $\mathcal{C}_n \subset \mathcal{A}_n$, the adaptive lasso prior becomes

$$\pi(\boldsymbol{\delta} + \boldsymbol{\beta}^0) = \exp\left\{-\tilde{\boldsymbol{w}}_{\mathcal{S}}^T(\boldsymbol{\beta}_{\mathcal{S}}^0 + \boldsymbol{\delta}_{\mathcal{S}}) - \sum_{j \notin \mathcal{S}} \tilde{w}_j |\delta_j|\right\}, \quad \boldsymbol{\delta} \in \mathcal{C}_n.$$

Combining the above displayed equations gives the bound for the posterior density

$$\begin{aligned} p_n(\boldsymbol{\delta}) \;\; &\lesssim_{\mathrm{pr}^*} \;\; \exp\left\{\frac{R}{2n\theta_m}\left\|\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{\phi}\right\|^2 - \tilde{\boldsymbol{w}}_{\mathcal{S}}^T\boldsymbol{\beta}_{\mathcal{S}}^0\right\} \cdot \exp\left\{-\frac{n\theta_m}{2R}\left\|\boldsymbol{\delta}_{\mathcal{S}} - \mu_1\right\|^2 - \tilde{\boldsymbol{w}}_{\mathcal{S}}^T\boldsymbol{\delta}_{\mathcal{S}}\right\} \\ &\quad \cdot \exp\left\{-\sum_{j=s+1}^{p}(\tilde{w}_j - \alpha_n)|\delta_j|\right\} \\ &\triangleq \;\; \exp\left\{\frac{R}{2n\theta_m}\left\|\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{\phi}\right\|^2 - \tilde{\boldsymbol{w}}_{\mathcal{S}}^T\boldsymbol{\beta}_{\mathcal{S}}^0\right\} \cdot \overline{p}_{2n}(\boldsymbol{\delta}), \end{aligned}$$

uniformly on $\boldsymbol{\delta} \in \mathcal{B}_n^c \cap \mathcal{C}_n$, where $\alpha_n = \left\|\boldsymbol{X}_{\mathcal{S}^c}^T\boldsymbol{\phi}\right\|_\infty$ and $\mu_1 = R \cdot \boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{\phi}/(n\theta_m)$.

Step II: Bounding the posterior integral on $\boldsymbol{\mathcal{B}_n^c \cap \mathcal{C}_n}$

Here we bound the posterior integral by using the upper bound $\overline{p}_{2n}(\boldsymbol{\delta})$ in step I. Let $\gamma > 0$ be a small enough constant, we define the event $N_1(\gamma) = \{\alpha_n \leq \gamma \cdot w_{\min}\}$, where α_n is defined in the end of step I and $w_{\min} = \min\{\tilde{w}_j : j \notin \mathcal{S}\}$.

Given the data, let $Z \in \mathbb{R}^s$ and $\xi = (\xi_{s+1}, \ldots, \xi_p)$ be distributed as

$$\begin{aligned} \xi_j &\overset{ind.}{\sim} \text{Laplace}\left(\frac{1}{\tilde{w}_j - \alpha_n}\right), \quad j \in \{s+1, \ldots, p\}, \\ Z &\sim \mathrm{N}\left(\mu_1, \frac{R}{n\theta_m} I_s\right), \end{aligned}$$

and Z is independent of ξ. In what follows, we shall write $\mathrm{pr}(\cdot)$ and $\mathrm{E}(\cdot)$ as the probabilistic operators with respect to (Z, ξ) given the data. For any fixed μ_1 and $\alpha_n < w_{\min}$, the function $\overline{p}_{2n}(\boldsymbol{\delta})$ is proportional to the joint density function of the vector (Z, ξ). Therefore, the integration of $\overline{p}_{2n}(\boldsymbol{\delta})$ can be related to the probabilistic statements about (Z, ξ), which gives, on the event $N_1(\gamma)$:

$$
\begin{aligned}
& \int_{\mathcal{C}_n \cap \mathcal{B}_n^c} \|A\boldsymbol{\delta}\|^{\alpha} p_n(\boldsymbol{\delta})\, \mathrm{d}\boldsymbol{\delta} \\
\lesssim_{\mathrm{pr}^*} \quad & \exp\left\{\frac{R}{2n\theta_m}\left\|\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{\phi}\right\|^2 - \tilde{\boldsymbol{w}}_{\mathcal{S}}^T\boldsymbol{\beta}_{\mathcal{S}}^0\right\}\left(\frac{2\pi R}{n\theta_m}\right)^{s/2} \cdot \prod_{j=s+1}^{p}\left(\frac{2}{\tilde{w}_j - \alpha_n}\right) \\
& \cdot \left[\mathrm{E}\left\{\left(\|\sqrt{n}Z\|^{2\alpha} + \|n\lambda\xi\|^{2\alpha}\right)\exp\left(-2\tilde{\boldsymbol{w}}_{\mathcal{S}}^T Z\right)\right\}\right]^{1/2} \qquad \text{(S4.8)} \\
& \cdot \left[\mathrm{pr}\left(\|Z\| \geq \frac{M_n}{\sqrt{n}} \text{ or } \|\xi\|_\infty \geq \frac{M_n}{n\lambda}\right)\right]^{1/2}, \qquad \text{(S4.9)}
\end{aligned}
$$

by the Cauchy-Schwartz inequality and $\|A\boldsymbol{\delta}\|^{2\alpha} \leq 2\|n^{1/2}Z\|^{2\alpha} + 2\|n\lambda\xi\|^{2\alpha}$.

Next we bound (S4.9) and (S4.8) separately. First, by standard tail bounds in Lemma 1 and 2, we have for (S4.9):

$$
\mathrm{pr}\left(\|Z\| \geq \frac{M_n}{\sqrt{n}} \text{ or } \|\xi\|_\infty \geq \frac{M_n}{n\lambda}\right) \lesssim \exp\left\{-\frac{\sqrt{M_n}}{2}\right\},
$$

which holds on the intersection of events $N_2(\gamma) = \{n \cdot \|\mu_1\|^2 \leq \gamma \cdot M_n^2\}$ and $N_3(\gamma) = \left\{\sqrt{M}_n \cdot (w_{\min} - \alpha_n) \geq \gamma \cdot n\lambda\right\}$ for small enough $\gamma > 0$.

Second, by Lemma 1, 2 and 3, we have

$$\begin{aligned}
\mathrm{E}\left(\exp\left\{-2\tilde{\boldsymbol{w}}_{\mathcal{S}}^T Z\right\}\right) &\leq (1+4\gamma), \\
\mathrm{E}\left(\|n\lambda\xi\|^{2\alpha}\right) &\lesssim \left[\frac{p(n\lambda)^2}{(w_{\min}-\alpha_n)^2}\right]^{\alpha} \lesssim M_n^{\alpha},
\end{aligned}$$

and also

$$\begin{aligned}
\mathrm{E}\left(\|\sqrt{n}Z\|^{2\alpha}\exp\left\{-2\tilde{\boldsymbol{w}}_{\mathcal{S}}^T Z\right\}\right) &\leq \left[\mathrm{E}\left(\exp\left\{-4\tilde{\boldsymbol{w}}_{\mathcal{S}}^T Z\right\}\right)\right]^{1/2} \times \left[\mathrm{E}\left(\|\sqrt{n}Z\|^{4\alpha}\right)\right]^{1/2} \\
&\lesssim (1+4\gamma)^{1/2} \times \left(\frac{R}{\theta_m}\right)^{\alpha}\left[s^{2\alpha}+\left(\frac{n\theta_m\|\mu_1\|^2}{R}\right)^{2\alpha}\right]^{1/2} \\
&\lesssim M_n^{2\alpha},
\end{aligned}$$

for $\alpha = 0, 1, 2$, which holds on the event $N_2(\gamma)$, $N_3(\gamma)$, and $N_4(\gamma) = \left\{\|4\tilde{\boldsymbol{w}}_{\mathcal{S}}\| \leq \gamma \cdot \left(\sqrt{2n\theta_m/R} \wedge \|\mu_1\|^{-1}\right)\right\}$. Therefore, (S4.8) is bounded by

$$\left[\mathrm{E}\left\{\left(\|\sqrt{n}Z\|^{2\alpha}+\|n\lambda\xi\|^{2\alpha}\right)\exp\left(-2\tilde{\boldsymbol{w}}_{\mathcal{S}}^T Z\right)\right\}\right]^{1/2} \lesssim M_n^{\alpha}.$$

Finally, note that

$$\prod_{j=s+1}^{p}\left(\frac{2}{\tilde{w}_j-\alpha_n}\right) \lesssim \left(\frac{2}{\tilde{w}_j}\right)^{p-s},$$

on the event $N_1(\gamma)$. Therefore, combining the above displayed equation with the bounds for (S4.8) and (S4.9), we have

$$\begin{aligned}
\int_{\mathcal{C}_n\cap\mathcal{B}_n^c}\|A\boldsymbol{\delta}\|^{\alpha}p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \lesssim_{\mathrm{pr}^*} & \exp\left\{\frac{n\theta_m}{2R}\|\mu_1\|^2-\tilde{\boldsymbol{w}}_{\mathcal{S}}^T\boldsymbol{\beta}_{\mathcal{S}}^0\right\}\left(\frac{2\pi R}{n\theta_m}\right)^{s/2}\left\{\prod_{j=s+1}^{p}\left(\frac{2}{\tilde{w}_j}\right)\right\} \\
& \cdot M_n^{\alpha}\cdot\exp\left\{-\frac{\sqrt{M_n}}{4}\right\},
\end{aligned}$$

on the events $N_1(\gamma)$ through $N_4(\gamma)$, where μ_1 is defined in step I.

Step III: Bound on $\mathcal{C}_n^c$

Here we use the upper bound $\overline{p}_{1n}$ in step I. Note $\|A\boldsymbol{\delta}\|^\alpha \le (n\lambda)^\alpha \|\boldsymbol{\delta}\|^\alpha \le (n\lambda/\theta_m^{1/2})^\alpha \|G^{1/2}\boldsymbol{\delta}\|^\alpha$; For any $\alpha = 0, 1$ and 2, we have

$$\begin{aligned}
\int_{\mathcal{C}_n^c} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} &\lesssim_{\mathrm{pr}^*} \left(\frac{n\lambda}{\sqrt{\theta_m}}\right)^\alpha \int_{\|u\|\ge Rq_0} \|u\|^\alpha \exp\left\{-\frac{n\varepsilon_0\|u\|}{Rq_0}\right\}\mathrm{d}u \\
&\lesssim \left(\frac{n\lambda}{\sqrt{\theta_m}}\right)^\alpha \left[\int_{\mathbb{R}^p} \|u\|^{2\alpha} \exp\left\{-\frac{n\varepsilon_0\|u\|_1}{Rq_0\sqrt{p}}\right\}\mathrm{d}u\right]^{1/2} \\
&\quad \cdot \left[\int_{\|u\|\ge Rq_0} \exp\left\{-\frac{n\varepsilon_0\|u\|_1}{Rq_0\sqrt{p}}\right\}\mathrm{d}u\right]^{1/2} \\
&\lesssim \lambda^\alpha \exp\{-n\varepsilon_0/4\}, \qquad \text{(S4.10)}
\end{aligned}$$

where the second inequality follows from Cauchy-Schwartz inequality and $p^{1/2}\|u\|_2 \ge \|u\|_1$, and the last inequality follows from the Cramér-Chernoff method (Boucheron et al., 2013, Section 2.2) (see also S6.28), and Lemma 2.

Step IV: Final bounds on the posterior probabilities

Here we close the proof by showing the posterior bounds derived in steps II and III are both negligible to $\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}$. From Lemma 10, we have

$$\begin{aligned}
\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} &\gtrsim_{\mathrm{pr}^*} \prod_{j=s+1}^{p} \left(\frac{2}{\tilde{w}_j}\right) \cdot \left(\frac{2\pi}{n\theta_M}\right)^{s/2} \cdot \exp\left(-\tilde{\boldsymbol{w}}_{\mathcal{S}}^T \boldsymbol{\beta}_{\mathcal{S}}^0 + \frac{n}{2}\Delta_s^T G_{\mathcal{S}} \Delta_s\right) \\
&\triangleq \widetilde{P}_n.
\end{aligned}$$

Let $N_5(\gamma) = \{w_{\max} \le \gamma \cdot n\}$. In step V later, we shall show that the events $N_1(\gamma)$ through $N_5(\gamma)$ holds with pr*-probability tending to 1, for

small enough γ. Therefore the bounds derived in steps II and III holds with pr*-probability tending to 1.

For the area $\mathcal{C}_n \cap \mathcal{B}_n^c$, we use the upper bound displayed at the end of step II; comparing it with $\widetilde{P}_n$ gives

$$\begin{aligned}\frac{\int_{\mathcal{C}_n \cap \mathcal{B}_n^c} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{P}_n} &\lesssim_{\mathrm{pr}^*} M_n^\alpha \left(\frac{R\theta_M}{\theta_m}\right)^{s/2} \cdot \exp\left\{\frac{n\theta_m}{2R}\|\mu_1\|^2 - \frac{\sqrt{M_n}}{4}\right\} \\ &\lesssim_{\mathrm{pr}^*} \exp\left\{-\frac{\sqrt{M}_n}{8}\right\},\end{aligned}$$

the last inequality holds on the event $N_3(\gamma)$ for small enough γ.

For the area $\mathcal{C}_n^c$, note $\|\tilde{\boldsymbol{w}}_{\mathcal{S}}\| \le C_1 n^{1/2}$ and $\|\tilde{\boldsymbol{w}}_{\mathcal{S}^c}\|_\infty \le \gamma n$ on the events $N_2(\gamma)$ and $N_5(\gamma)$ for some constant $C_1 > 0$. Therefore, using the bound in step III, we can show

$$\begin{aligned}\frac{\int_{\mathcal{C}_n^c} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{P}_n} &\lesssim_{\mathrm{pr}^*} \lambda^\alpha \exp\left\{p\log n + C_1\sqrt{n} - \frac{n\varepsilon_0}{4}\right\} \\ &\lesssim \exp\left\{-\frac{n\varepsilon_0}{8}\right\},\end{aligned}$$

for large enough n.

Therefore, we have shown (S4.6), and the proof is now complete.

Step V: Auxiliary calculations

Now we show that each of the events

$$N_1(\gamma) = \{\alpha_n \le \gamma \cdot w_{\min}\}, \quad N_4(\gamma) = \left\{ \|\tilde{\boldsymbol{w}}_{\mathcal{S}}\| \le \gamma \cdot \left(\sqrt{2n\theta_m/R} \wedge \frac{1}{\|\mu_1\|} \right) \right\},$$

$$N_2(\gamma) = \left\{ n \cdot \mu_1^T \mu_1 \le \gamma \cdot M_n^2 \right\}, \quad N_3(\gamma) = \left\{ \sqrt{M}_n \cdot (w_{\min} - \alpha_n) \ge \gamma \cdot n\lambda \right\},$$

$$\text{and} \quad N_5(\gamma) = \{w_{\max} \le \gamma \cdot n\},$$

holds with pr*-probability going to 1.

First, $\mathrm{pr}^*(N_2(\gamma)) \to 1$ follows from the Central Limit Theorem. Next we consider the other events. Note $\alpha_n = O_{\mathrm{pr}^*}(\sqrt{n})$ from the Central Limit Theorem, the key is to analyze each $\tilde{w}_j = n^{1/2}\lambda/|\hat{\beta}_j|$. From standard asymptotic results for quantile regression (Koenker, 2005, Section 4.2), we have

$$\begin{aligned} w_{\max} &= O_{\mathrm{pr}^*}(n\lambda), \\ \frac{1}{w_{\min}} &= O_{\mathrm{pr}^*}\left(\frac{1}{n\lambda}\right), \\ \|\tilde{\boldsymbol{w}}_{\mathcal{S}}\|^2 &= O_{\mathrm{pr}^*}\left(n\lambda^2\right), \end{aligned} \tag{S4.11}$$

since s and p are both fixed. Straightforward calculations shows that all the events $N_k(\gamma)$'s have pr*-probability going to 1, for all small enough $\gamma > 0$.

□

S4.2 Part 2 of Theorem 1

Now we prove part 2 of Theorem 1 under the AL prior. To this end, we need the following lemma, the proofs of which is deferred to Section S8.

Lemma 11. *Consider the density function f_n in (S4.3). Under the conditions of Theorem 1, given any diverging sequence $M_n \to +\infty$ we have*

$$\frac{\displaystyle\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\delta\|^\alpha f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0,$$

in pr^**-probability for $\alpha = 0, 1, 2$, where $\mathcal{B}_n$ is defined in (S4.4).*

Proof of Theorem 1, part 2. Similar to part 1 of the proof, we assume $\beta_j^0 \geq b_0 > 0$ for all $j \in \mathcal{S}$. We shall show that $p_n(\boldsymbol{\delta})$ converges to $f_n(\boldsymbol{\delta})$ in total variation, where p_n and f_n are defined in (S4.3). Following the proof of Theorem 1 in Chernozhukov and Hong (2003), it suffices to show

$$\frac{\displaystyle\int_{\mathbb{R}^{p+1}} |p_n(\boldsymbol{\delta}) - f_n(\boldsymbol{\delta})|\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^{p+1}} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} = o_{\mathrm{pr}^*}(1). \tag{S4.12}$$

Recalling $\mathcal{B}_n$ from (S4.4), we can choose a sequence of $K_n \to +\infty$ slow enough such that: (i) $\mathcal{B}_n(K_n)$ satisfies Lemma 9, (ii) $K_n \ll \min\{\lambda^{-1}, n^{1/2}\lambda\}$, and (iii) $K_n R_n = o_{\mathrm{pr}^*}(n^{-1/2})$; here $R_n = \|G_{\mathcal{S}}(\tilde{\boldsymbol{\delta}}_{\mathcal{S}} - \Delta_s)\| = o_{\mathrm{pr}^*}(n^{-1/2})$ owns to the Bahadur representation of the oracle estimator $\tilde{\boldsymbol{\beta}}_{\mathcal{S}}$ (Koenker, 2005, Section 4.2). In the following, we upper bound the integral of $|p_n - f_n|$ on $\mathcal{B}_n(K_n)$ and its complement, separately in steps I and II.

Step I: Bounding the integral on $\mathcal{B}_n(K_n)$

We upper bound $|p_n(\boldsymbol{\delta}) - h_n(\boldsymbol{\delta})|$ when $\boldsymbol{\delta} \in \mathcal{B}_n(K_n)$. To this end, we first simplify the prior and working likelihood in $p_n(\boldsymbol{\delta})$, separately. By

Assumption 4 and the sign of $\boldsymbol{\beta}^0_{\mathcal{S}}$, all entries of $\boldsymbol{\delta}_{\mathcal{S}} + \boldsymbol{\beta}^0_{\mathcal{S}}$ are positive when $\boldsymbol{\delta} \in \mathcal{B}_n(K_n)$; therefore the adaptive lasso prior becomes

$$\pi(\boldsymbol{\delta}) = \exp\left\{-\tilde{w}_{\mathcal{S}}^T(\boldsymbol{\delta}_{\mathcal{S}} + \boldsymbol{\beta}^0_{\mathcal{S}}) - \sum_{j\notin\mathcal{S}} \tilde{w}_j|\delta_j|\right\}, \quad \boldsymbol{\delta} \in \mathcal{B}_n(K_n).$$

Next we show the quantile loss function $L_n(\boldsymbol{\delta}+\boldsymbol{\beta}^0)$ in $p_n(\boldsymbol{\delta})$ can be approximated by a quadratic function. By Lemma 9, we have

$$\begin{aligned}
&\sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \left|L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta}+\boldsymbol{\beta}^0) + \frac{n}{2}\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}} - n\Delta_s^T G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}}\right| \\
\leq\ & o_{\mathrm{pr}^*}(1) + n \sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \left|\Delta_s^T G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}} - \Delta_p^T G\boldsymbol{\delta}\right| \\
\lesssim\ & o_{\mathrm{pr}^*}(1) + \frac{p^{1/2}K_n}{n\lambda}\left\|\sum_{i=1}^n x_{i,\mathcal{S}^c}\phi_\tau(y_i - x_i^T\boldsymbol{\beta}^0)\right\| \\
=\ & o_{\mathrm{pr}^*}(1),
\end{aligned}$$

where Δ_p and Δ_s are defined in Section S3; the last inequality holds by the central limit theorem and that $K_n \ll n^{1/2}\lambda$.

Using the above displayed equations, and comparing p_n to f_n, we have

$$\begin{aligned}
\sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \left|\log\left\{\frac{f_n(\boldsymbol{\delta})}{p_n(\boldsymbol{\delta})}\right\}\right| &= \sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \left|L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta}+\boldsymbol{\beta}^0) + \frac{n}{2}\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}} - n\tilde{\boldsymbol{\delta}}_{\mathcal{S}} G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}} - \tilde{w}_{\mathcal{S}}^T\boldsymbol{\delta}_{\mathcal{S}}\right| \\
&\leq o_{\mathrm{pr}^*}(1) + \sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \left|n(\tilde{\boldsymbol{\delta}}_{\mathcal{S}} - \Delta_s)G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}}\right| + \sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \left|\tilde{w}_{\mathcal{S}}^T\boldsymbol{\delta}_{\mathcal{S}}\right| \\
&= o_{\mathrm{pr}^*}(1);
\end{aligned}$$

in the above, the second term is $o_{\mathrm{pr}^*}(1)$ by our choice of K_n, and the third term is $o_{\mathrm{pr}^*}(1)$ since $K_n \ll \lambda^{-1}\ \tilde{w}_{\mathcal{S}} = O_{\mathrm{pr}^*}(n^{1/2}\lambda)$ from Assumption 4. Therefore, the above displayed equation is $o_{\mathrm{pr}^*}(1)$, which further implies

$|f_n(\boldsymbol{\delta})/p_n(\boldsymbol{\delta}) - 1| = o_{\mathrm{pr}^*}(1)$ uniformly when $\boldsymbol{\delta} \in \mathcal{B}_n(K_n)$.

Now we can bound the integration of $|f_n - p_n|$ on $\mathcal{B}_n(K_n)$ by

$$\begin{aligned}\int_{\mathcal{B}_n(K_n)} |p_n(\boldsymbol{\delta}) - f_n(\boldsymbol{\delta})| \, \mathrm{d}\boldsymbol{\delta} &= \int_{\mathcal{B}_n(K_n)} p_n(\boldsymbol{\delta}) \left| 1 - \left(\frac{f_n(\boldsymbol{\delta})}{p_n(\boldsymbol{\delta})} \right) \right| \mathrm{d}\boldsymbol{\delta} \\ &= o_{\mathrm{pr}^*} \left\{ \int_{\mathbb{R}^{p+1}} p_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta} \right\}.\end{aligned}$$

Step II: Bounding the integral on $\mathcal{B}_n(K_n)^c$

Here we give upper bounds for integrating p_n and f_n separately. For p_n, part 1 of Theorem 1 directly implies that

$$\frac{\displaystyle\int_{\mathcal{B}_n(K_n)^c} p_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^{p+1}} p_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta}} = \Pi\left(\boldsymbol{\delta} \in \mathcal{B}_n(K_n)^c \mid \mathbb{D}_n\right) = o_{\mathrm{pr}^*}(1).$$

For f_n, note its normalizing constant can be explicitly derived from Gaussian and Laplace distributions:

$$\int_{\mathbb{R}^{p+1}} f_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta} = T_n \times |G_{\mathcal{S}}|^{-1/2} \left(\frac{2\pi}{n} \right)^{(s+1)/2} \left\{ \prod_{j \notin \mathcal{S}} \left(\frac{2}{\tilde{w}_j} \right) \right\}.$$

Comparing the normalizing constant for f_n above with $\int p_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta}$ in Lemma 10, we have from Lemma 11 that

$$\frac{\displaystyle\int_{\mathcal{B}_n(K_n)^c} f_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^{p+1}} p_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta}} \lesssim_{\mathrm{pr}^*} \frac{\displaystyle\int_{\mathcal{B}_n(K_n)^c} f_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^{p+1}} f_n(\boldsymbol{\delta}) \, \mathrm{d}\boldsymbol{\delta}} = o_{\mathrm{pr}^*}(1).$$

Combining steps I and II, we have proved (S4.12), which concludes the proof by using Theorem 1 of Chernozhukov and Hong (2003). □

S5 Proof of other results in Section 3

In this Section, we give the proofs to Theorem 2, Proposition 1, and Equation (3.9) in the paper. In line with the paper, we focus on the asymptotic regime where the dimension p is fixed. We give a unified framework to proof Theorem 2 that works for both of the prior choices. We shall use the generic notation $\pi(\boldsymbol{\beta}))$ to denote either the AL or CA prior.

We review some notations first. Let C_{1n}, C_{2n} denote some generic sequence that does not depend on $\boldsymbol{\delta}$ nor $\boldsymbol{\beta}$, but could depend on the choice of prior. For $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$, the posterior density is

$$p_n(\boldsymbol{\delta}) = C_{1n} \times \pi(\boldsymbol{\delta} + \boldsymbol{\beta}^0) \times \exp\left\{L_n(\boldsymbol{\beta}^0) - L_n\left(\boldsymbol{\delta} + \boldsymbol{\beta}^0\right)\right\}. \tag{S5.13}$$

Let $\tilde{\boldsymbol{\delta}}_{\mathcal{S}} = \tilde{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}^0_{\mathcal{S}}$. In accordance with Theorem 1, (S4.3) and (S6.19), we define

$$f_n(\boldsymbol{\delta}) = C_{2n} \times \exp\left\{-\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}} - \tilde{\boldsymbol{\delta}}_{\mathcal{S}})^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \tilde{\boldsymbol{\delta}}_{\mathcal{S}})\right\} \times \left[\prod_{s \notin \mathcal{S}} \pi_j(n\lambda\beta_j)\right], \tag{S5.14}$$

as the density function for the limiting posterior, where $\pi_j(u) = (n^{-1/2}w_j/2)\exp\{-n^{-1/2}w_j|u|\}$ if we use the AL prior and $\pi_j(u) = (1/2)\exp\{-|u|\}$ if we use the CA prior.

S5.1 Theorem 2

To show Theorem 2, we need the following technical lemma, which strengthens the posterior consistency result in part 1 of Theorem 1. The result of the lemma under the AL prior is implied by (S4.6) in Section S4. Recall the definition of $\mathcal{B}_n(M_n)$ from (S4.4) when the dimension is fixed, and the matrix A from (S4.5) such that $A\boldsymbol{\delta} = (n^{1/2}\boldsymbol{\delta}_{\mathcal{S}}^T, n\lambda\boldsymbol{\delta}_{\mathcal{S}^c}^T)^T$.

Lemma 12. *Under the conditions of Theorem 1, and consider either of the AL or CA prior. For the posterior density $p_n(\boldsymbol{\delta})$, we have*

$$\frac{\displaystyle\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\delta\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0, \quad \textit{and} \quad \frac{\displaystyle\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\delta\|^\alpha f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0,$$

in pr**-probability, for any diverging sequence $M_n \to +\infty$ and any $\alpha = 0, 1, 2$.*

Proof of Theorem 2. First, we prove part 1 of Theorem 2, showing that the posterior moments converge at an adaptive rate. Recall $p_n(\boldsymbol{\delta})$ and $f_n(\boldsymbol{\delta})$ from (S5.13) and (S5.14). We first show that

$$\int_{\mathbb{R}^p} \|A\boldsymbol{\delta}\|^\alpha \left| \frac{f_n(\boldsymbol{\delta})}{\displaystyle\int f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} - \frac{p_n(\boldsymbol{\delta})}{\displaystyle\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \right| \mathrm{d}\boldsymbol{\delta} \to 0, \tag{S5.15}$$

in pr*-probability for $\alpha = 1, 2$.

We divide the integral in (S5.15) into two areas. Recalling $\mathcal{B}_n(M_n)$ from (S4.4), by choosing a sequence M_n that diverges slow enough, we have from

Part 2 of Theorem 1 that

$$\int_{\mathcal{B}_n(M_n)} \|A\boldsymbol{\delta}\|^\alpha \left| \frac{f_n(\boldsymbol{\delta})}{\int f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} - \frac{p_n(\boldsymbol{\delta})}{\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \right| \mathrm{d}\boldsymbol{\delta} \;\leq\; 2M_n^\alpha \int_{\mathbb{R}^p} \left| \frac{f_n(\boldsymbol{\delta})}{\int f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} - \frac{p_n(\boldsymbol{\delta})}{\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \right| \mathrm{d}\boldsymbol{\delta}$$
$$= \; o_{\mathrm{pr}^*}(1),$$

since the integral on the right converges towards 0. On $\mathcal{B}_n^c$, from Lemma 12, we obtain

$$\frac{\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\boldsymbol{\delta}\|^\alpha f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\int_{\mathbb{R}^p} f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} = o_{\mathrm{pr}^*}(1), \quad \frac{\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} = o_{\mathrm{pr}^*}(1).$$

Therefore, we have proved (S5.15).

Denote the posterior mean as $\check{\boldsymbol{\delta}}$ under $p_n(\boldsymbol{\delta})$, and recall $\tilde{\boldsymbol{\beta}}_{\mathcal{S}}$ is the oracle quantile regression estimator. Note that

$$\check{\boldsymbol{\delta}}_{\mathcal{S}} = \int_{\mathbb{R}^p} \boldsymbol{\delta}_{\mathcal{S}} \cdot p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}, \quad \text{and} \quad (\tilde{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}}^0) = \int_{\mathbb{R}^p} \boldsymbol{\delta}_{\mathcal{S}} \cdot f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}.$$

Taking $\alpha = 1$ in (S5.15) implies that

$$\sqrt{n}\left[\check{\boldsymbol{\delta}}_{\mathcal{S}} - (\tilde{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}}^0)\right] = \sqrt{n}\left(\check{\boldsymbol{\beta}}_{\mathcal{S}} - \tilde{\boldsymbol{\beta}}_{\mathcal{S}}\right) = o_{\mathrm{pr}^*}(1),$$

and similarly

$$n\lambda\left(\check{\boldsymbol{\delta}}_{\mathcal{S}^c} - 0\right) = n\lambda\left(\check{\boldsymbol{\beta}}_{\mathcal{S}^c} - 0\right) = o_{\mathrm{pr}^*}(1).$$

The desired asymptotic normality in Theorem 2 then follows from the Bahadur representation of $\tilde{\beta}_{\mathcal{S}}$; see e.g. Koenker (2005, Section 4.2).

Next we prove the validity of the sandwich-adjustment. We define some notations specific to the prior we use. Let $v_j = 2\tilde{w}_j^{-2}$ if we use the AL prior, and $v_j = 2(n\lambda)^{-2}$ if we use the CA prior. Straightforward analysis shows that $v_j \asymp_{\mathrm{pr}^*} \{(n\lambda)^{-2}\}$ for either the AL or the CA prior; see also (S4.11). Let $V_0 = diag\{v_{s+1}, \ldots, v_p\}$ be a $p-s$ dimensional diagonal matrix. From standard calculation, the variance-covariance matrix under the limiting density function $f_n(\delta)$ is

$$\begin{bmatrix} \frac{1}{n}G_{\mathcal{S}}^{-1} & 0 \\ 0 & V_0 \end{bmatrix} \in \mathbb{R}^{p\times p}.$$

Taking $\alpha = 2$ in (S5.15), we obtain

$$\left\| A\left(\check{\Sigma} - \begin{bmatrix} \frac{1}{n}G_{\mathcal{S}}^{-1} & 0 \\ 0 & V_0 \end{bmatrix}\right) A^T \right\|_2 = o_{\mathrm{pr}^*}(1), \tag{S5.16}$$

where $\|\cdot\|_2$ is the matrix spectrum norm. Furthermore, let $\hat{D} = \sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T/n$, by the law of large number we have $\|\hat{D} - D\|_2 = o_{\mathrm{pr}^*}(1)$, where $D = \mathrm{E}^*[XX^T]$.

Recall the sandwich form adjustment is $\check{\Sigma}_{adj} = n\tau(1-\tau)\check{\Sigma}\widehat{D}\check{\Sigma}$. Calculating the multiplication by block, and using (S5.16) gives

$$n\,\check{\Sigma}_{adj,\mathcal{S}} = \tau(1-\tau)G_{\mathcal{S}}^{-1}D_{\mathcal{S}}G_{\mathcal{S}}^{-1} + o_{\mathrm{pr}^*}(1).$$

Furthermore, let $\check{\boldsymbol{\sigma}}_j$ be the jth column of $\check{\Sigma}$ for any $j \notin \mathcal{S}$; (S5.16) shows that

$$\check{\boldsymbol{\sigma}}_j = v_j\boldsymbol{e}_j + \left[\frac{\boldsymbol{r}_{\mathcal{S}}^T}{n^{3/2}\lambda}\,,\ \frac{\boldsymbol{r}_{\mathcal{S}^c}^T}{n^2\lambda^2}\right]^T,$$

where $\boldsymbol{e}_j$ is the jth standard unit vector in $\mathbb{R}^p$ and $\boldsymbol{r} \in \mathbb{R}^p = o_{\mathrm{pr}^*}(1)$. Hence

$$\begin{aligned}\check{\Sigma}_{adj}(j,j) &\asymp_{\mathrm{pr}^*} n\check{\sigma}_j^T\check{\sigma}_j \\ &= nv_j^2 + \frac{\|r_{\mathcal{S}}\|^2}{n^2\lambda^2} + \frac{\|r_{\mathcal{S}^c}\|^2}{n^3\lambda^4} + \frac{2r_jv_j}{n\lambda^2},\end{aligned}$$

for any $j \notin \mathcal{S}$, where the first equation follows since $\theta_{\max}(\hat{D})$ and $\theta_{\max}(\hat{D})$ are both bounded in pr*-probability. Noting that $v_j = O_{\mathrm{pr}^*}\{(n\lambda)^{-2}\}$, we obtain

$$\check{\Sigma}_{adj}(j,j) = o_{\mathrm{pr}^*}\left(\frac{1}{n^2\lambda^2}\right),$$

since $n^{-1/2} \ll \lambda \ll 1$. Also, since $v_j^{-1} = O_{\mathrm{pr}^*}\{(n\lambda)^2\}$, we also obtain that

$$\frac{1}{(n^3\lambda^4)\check{\Sigma}_{adj}(j,j)} \leq \frac{1}{n^4\lambda^4v_j^2 + r_jn^2\lambda^2v_j} = O_{\mathrm{pr}^*}(1).$$

Thus, the proof is now complete. □

S5.2 Proposition 1

In this section, we shall continue to give a unified framework regardless of the prior choice. We first introduce some notations for the weighted posterior. Recall $\epsilon = Y - X^T\boldsymbol{\beta}^0$ and $\zeta(x) = f_{\epsilon|X=\boldsymbol{x}}(0)$. For some sequence $C_{1n}^{(w)}$ and $C_{2n}^{(w)}$ that does not depend on $\boldsymbol{\beta}$, the optimally weighted posterior density under either prior is

$$p_n^{(w)}(\boldsymbol{\beta}) = C_{1n}^{(w)} \times \pi(\boldsymbol{\beta}) \times \exp\left\{-\sum_{i=1}^n \zeta(\boldsymbol{x}_i)\rho_\tau(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\right\},$$

and let $\check{\boldsymbol{\beta}}^{(w)}$ be the posterior mean. Similar to the previous subsection, we define

$$f_n^{(w)}(\boldsymbol{\beta}) = C_{2n}^{(w)} \exp\left\{-\frac{n}{2}(\boldsymbol{\beta}_{\mathcal{S}} - \tilde{\boldsymbol{\beta}}_{\mathcal{S}}^{(w)})^T G_{\mathcal{S}}(\boldsymbol{\beta}_{\mathcal{S}} - \tilde{\boldsymbol{\beta}}_{\mathcal{S}}^{(w)})\right\} \times \left[\prod_{j \notin \mathcal{S}} \pi_j(n\lambda\beta_j)\right],$$

where $\pi_j(u) = (n^{-1/2}w_j/2)\exp\{-n^{-1/2}w_j|u|\}$ if we use the AL prior and $\pi_j(u) = (1/2)\exp\{-|u|\}$ if we use the CA prior; furthermore, $\tilde{\boldsymbol{\beta}}_{\mathcal{S}}^{(w)}$ is the infeasible oracle estimator

$$\tilde{\boldsymbol{\beta}}_{\mathcal{S}}^{(w)} = \arg\min_{\boldsymbol{u} \in \mathbb{R}^s} \sum_{i=1}^n \zeta(\boldsymbol{x}_i)\rho_\tau(y_i - \boldsymbol{x}_{i,\mathcal{S}}^T \boldsymbol{u}).$$

From Theorem 5.1 in Koenker (2005), we have

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}}_{\mathcal{S}}^{(w)} - \boldsymbol{\beta}_{\mathcal{S}}^0\right) \to \mathrm{N}\left\{0, \tau(1-\tau)Q_{\mathcal{S}}^{-1}\right\},$$

in distribution, where $Q_{\mathcal{S}} = \mathrm{E}^*[X_{\mathcal{S}} X_{\mathcal{S}}^T f_{\epsilon|X}^2(0)]$.

Proof of Proposition 1. To show Proposition 1, it suffices to show that the posterior mean is asymptotically equivalent to the oracle estimator, i.e., $n^{1/2}(\tilde{\beta}_{\mathcal{S}}^{(w)} - \check{\beta}_{\mathcal{S}}^{(w)}) = o_{\mathrm{pr}^*}(1)$. Noting that the mean of $f_n^{(w)}(\boldsymbol{\beta})$ is $\tilde{\beta}_{\mathcal{S}}^{(w)}$, it suffices to show

$$\int_{\mathbb{R}^p} \|\sqrt{n}\boldsymbol{\delta}_{\mathcal{S}}\| \left| \frac{f_n^{(w)}(\boldsymbol{\delta})}{\int f_n^{(w)}(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} - \frac{p_n^{(w)}(\boldsymbol{\delta})}{\int p_n^{(w)}(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \right| \mathrm{d}\boldsymbol{\delta} = o_{\mathrm{pr}^*}(1), \tag{S5.17}$$

where $f_n^{(w)}(\boldsymbol{\delta}) = f_n^{(w)}(\boldsymbol{\delta} + \boldsymbol{\beta}^0 \mid \mathbb{D}_n)$ and $p_n^{(w)}(\boldsymbol{\delta}) = p_n^{(w)}(\boldsymbol{\delta} + \boldsymbol{\beta}^0 \mid \mathbb{D}_n)$.

Let $\tilde{Y} = \zeta(X)Y$ and $\tilde{X} = \zeta(X)X$. It is straightforward that $(\tilde{X}, \tilde{Y})$ satisfies the linear quantile regression model

$$Q_\tau(\tilde{Y} \mid \tilde{X} = \tilde{\boldsymbol{x}}) = \tilde{\boldsymbol{x}}^T \boldsymbol{\beta}^0,$$

since $\zeta(X)$ is bounded away from 0 and $+\infty$ by Assumption 3. Let $\tilde{\boldsymbol{x}}_i = \zeta(\boldsymbol{x}_i)\boldsymbol{x}_i$, $\tilde{y}_i = \zeta(\boldsymbol{x}_i)y_i$, and $\widetilde{\mathbb{D}}_n = \{(\tilde{\boldsymbol{x}}_i, \tilde{y}_i) : i = 1, \ldots, n\}$. Suppose we observe $\widetilde{\mathbb{D}}_n$, we consider the unweighted posterior for $\boldsymbol{\beta}$:

$$\begin{aligned} p_n(\boldsymbol{\beta} \mid \widetilde{\mathbb{D}}_n) &= \pi(\boldsymbol{\beta}) \times \exp\left\{ -\sum_{i=1}^n \rho_\tau(\tilde{y}_i - \tilde{\boldsymbol{x}}_i^T \boldsymbol{\beta}) \right\} \\ &= p_n^{(w)}(\boldsymbol{\beta} \mid \mathbb{D}_n), \end{aligned}$$

since the quantile loss function $\rho_\tau(u)$ is piece-wise linear. Therefore, to show (S5.17), we instead consider the unweighted posterior $p_n(\beta \mid \widetilde{\mathbb{D}}_n)$ and apply (S5.15) in Theorem 2.

Now we check the conditions of Theorem 2. Assumptions 1 and 4 hold automatically by the generating process of $(\tilde{X}, \tilde{Y})$. Given Assumptions 2-3 hold for $\mathbb{D}_n$, below we check Assumptions 2-3 hold for $\widetilde{\mathbb{D}}_n$ as well.

First note the conditional density of $\tilde{\epsilon} = \tilde{Y} - \tilde{X}^T \beta^0$ given X is

$$f_{\tilde{\epsilon}|X}(t) = \zeta^{-1}(X) \cdot f_{\epsilon|X}\left\{ t\zeta^{-1}(X) \right\},$$

and that $f_{\tilde{\epsilon}|\tilde{X}=x\zeta(x)}(t) = f_{\tilde{\epsilon}|X=x}(t)$. Therefore, Assumption 3 holds for $f_{\tilde{\epsilon}|\tilde{X}}$. Second, note

$$\mathrm{E}^*[\tilde{X}\tilde{X}^T] = \mathrm{E}^*[XX^T f_{\epsilon|X}^2(0)] = Q_{\mathcal{S}}.$$

Hence Assumption 2 follows for $\widetilde{\mathbb{D}}_n$ since $f_{\epsilon|X}(0)$ is bounded away from 0 and $+\infty$.

Let $\check{\beta}_{\mathcal{S}}^{(w)}$ be the posterior mean under $p_n^{(w)}(\beta \mid \mathbb{D}_n)$. Applying Theorem 2 shows (S5.17), which implies

$$\sqrt{n}\left(\check{\beta}_{\mathcal{S}}^{(w)} - \beta_{\mathcal{S}}^0\right) = \sqrt{n}\left(\tilde{\beta}_{\mathcal{S}}^{(w)} - \beta_{\mathcal{S}}^0\right) + o_{\mathrm{pr}^*}(1) \to \mathrm{N}\left\{0, \tau(1-\tau)Q_{\mathcal{S}}^{-1}\right\},$$

in distribution. This concludes the proof.

□

S5.3 Proof of Equation (3.9)

It suffices to show $V_{\mathcal{S}} \preceq Q_{\mathcal{S}}$. First note that $\mathrm{E}^*\{f_{\epsilon|X}(0) \mid X_{\mathcal{S}}\} = f_{\epsilon|X_{\mathcal{S}}}(0)$, which implies

$$\mathrm{var}^*\{f_{\epsilon|X}(0) \mid X_{\mathcal{S}}\} = \mathrm{E}^*\{f_{\epsilon|X}^2(0) \mid X_{\mathcal{S}}\} - f_{\epsilon|X_{\mathcal{S}}}^2(0) \geq 0.$$

Therefore, by the law of total expectation

$$\begin{aligned} Q_{\mathcal{S}} - V_{\mathcal{S}} &= \mathrm{E}^*[X_{\mathcal{S}}X_{\mathcal{S}}^T\{f_{\epsilon|X}^2(0) - f_{\epsilon|X_{\mathcal{S}}}^2(0)\}] \\ &= \mathrm{E}^*[X_{\mathcal{S}}X_{\mathcal{S}}^T \,\mathrm{var}^*\{f_{\epsilon|X}(0) \mid X_{\mathcal{S}}\}], \end{aligned}$$

which is positive semi-definite since the conditional variance is non-negative.

S6 Proof of Theorem 3

In this section, we shall prove Theorem 3 with the CA prior in (2.5) of the paper, under the asymptotic regime where the dimension p may grow with the sample size n. We also provide the proof for Proposition S1 under the new properized CA prior. We define some additional notations for this section. We consider the centered posterior for $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$, which is

$$p_n(\delta) = \pi_{CA}(\delta + \boldsymbol{\beta}^0) \cdot \exp\left\{L_n(\boldsymbol{\beta}^0) - L_n\left(\delta + \boldsymbol{\beta}^0\right)\right\}.$$

For any sequence $M_n \to \infty$, we define

$$\mathcal{B}_n(M_n) = \left\{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}_{\mathcal{S}}\|_2 \leq M_n\sqrt{\frac{s}{n}}\ ,\ \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \leq M_n\frac{s\log p}{n\lambda}\right\}. \quad \text{(S6.18)}$$

Let $T_n = \exp\{n\Delta_s^T G_{\mathcal{S}}\Delta_s/2\}$ and $Q_n = \exp\{n\Delta_p^T G\Delta_p/2\}$; we define the following functions:

$$\begin{aligned} h_n(\boldsymbol{\delta}) &= Q_n \times \exp\left\{-\frac{n}{2}(\boldsymbol{\delta} - \Delta_p)^T G(\boldsymbol{\delta} - \Delta_p) - sn\lambda^2 - n\lambda\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\} \\ f_n(\boldsymbol{\delta}) &= T_n \times \exp\left\{-\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}} - \widetilde{\boldsymbol{\delta}}_{\mathcal{S}})^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \widetilde{\boldsymbol{\delta}}_{\mathcal{S}}) - sn\lambda^2 - n\lambda\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\}, \end{aligned} \quad \text{(S6.19)}$$

where $\widetilde{\boldsymbol{\delta}}_{\mathcal{S}} = \widetilde{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}^0_{\mathcal{S}}$. Note that $f_n(\boldsymbol{\delta})$ is proportional to the density function of the limiting distribution in Theorem 3, and $h_n(\boldsymbol{\delta})$ is an intermediate function; they both correspond the joint distribution of (i) a s-dimensional Gaussian distribution and (ii) a $p - s$ dimensional independent Laplace distribution with scale parameter $n\lambda$.

S6.1 Part 1 of Theorem 3

We start from part 1 of the theorem. Parallel to the proofs in Section S4, we need the following lemmas, the proof of which are deferred to Sections S7 and S8.

Lemma 13. *Suppose Assumptions 2', 3 and 4' hold. In addition, suppose the matrix $A \in \mathbb{R}^{q\times p}$ satisfies*

$$\boldsymbol{e}_j^T ADA\boldsymbol{e}_j \le C_0, \quad j \in \{1, \dots q\},$$

for some constant C_0, where $\boldsymbol{e}_j$ is the jth standard unit vector in $\mathbb{R}^q$. Then we have

$$\left\| \sum_{i=1}^n \phi_\tau(y_i - x_i^T\beta^0) Ax_i \right\|_\infty = O_{\mathrm{pr}^*}\left(\sqrt{n\log q}\right).$$

Lemma 14. *Let $\Delta_s = G_{\mathcal{S}}^{-1}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{\phi}$ and $\Delta_p = G^{-1}\boldsymbol{X}^T\boldsymbol{\phi}$. Suppose Assumption 2 holds, then we have*

$$\Delta_s^T G_{\mathcal{S}}\Delta_s = O_{\mathrm{pr}^*}(s/n), \quad \Delta_p^T G\Delta_p = O_{\mathrm{pr}^*}(p/n).$$

Lemma 15. *Under the conditions of Theorem 3, we have*

$$\int_{\mathbb{R}^p} p_n(\delta)\,\mathrm{d}\boldsymbol{\delta} \gtrsim_{\mathrm{pr}^*} \left(\frac{2\pi}{n}\right)^{s/2}\left(\frac{2}{n\lambda}\right)^{p-s} \cdot \frac{\exp\left(-sn\lambda^2 + n\Delta_s^T G_{\mathcal{S}}\Delta_s/2\right)}{\sqrt{|G_{\mathcal{S}}|}}.$$

Lemma 16. *Let*

$$\mathcal{A}_n = \left\{\boldsymbol{\delta} \in \mathbb{R}^p : \min_{j\in\mathcal{S}} |\beta_j| < \lambda, \quad or \quad \max_{j\notin\mathcal{S}} |\beta_j| > \lambda\right\}.$$

Under the conditions of Theorem 3, we have

$$\Pi\left(\boldsymbol{\delta} \in \mathcal{A}_n \ \mid \ \mathbb{D}_n\right) \to 0,$$

in pr**-probability.*

Proof of Theorem 3, part 1. Define $\mathcal{B}_n = \mathcal{B}_n(M_n)$ from (S6.18) and $\mathcal{A}_n$ in Lemma 16; furthermore, we define

$$\mathcal{C}_n = \left\{\boldsymbol{\delta} : \|G^{1/2}\boldsymbol{\delta}\| \leq 4q_0\right\},$$

where q_0 is the constant in Lemma 7. Note

$$\Pi\left(\boldsymbol{\delta} \in \mathcal{B}_n^c \mid \mathbb{D}_n\right) \leq \Pi\left(\boldsymbol{\delta} \in \mathcal{A}_n \mid \mathbb{D}_n\right) + \Pi\left(\boldsymbol{\delta} \in \mathcal{B}_n^c \cap \mathcal{C}_n \cap \mathcal{A}_n^c \mid \mathbb{D}_n\right) + \Pi\left(\boldsymbol{\delta} \in \mathcal{C}_n^c \cap \mathcal{A}_n^c \ \mid \ \mathbb{D}_n\right),$$

In view of Lemma 16, it suffices to show the latter two terms on the right hand side are both $o_{\mathrm{pr}^*}(1)$. Since those two areas are subsets of $\mathcal{A}_n^c$, the CA prior becomes

$$\pi_{CA}(\boldsymbol{\delta} + \boldsymbol{\beta}^0) = \exp\left\{-sn\lambda^2 - n\lambda \cdot \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\}, \quad \boldsymbol{\delta} \in \mathcal{A}_n^c, \tag{S6.20}$$

throughout the proof.

We divide our proof into five parts. In step I, we first give upper bounds for the posterior density $p_n(\boldsymbol{\delta})$; in steps II - III, we obtain upper bounds for the posterior integral $\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}$ on the two areas $\mathcal{B}_n^c \cap \mathcal{C}_n \cap \mathcal{A}_n^c$ and $\mathcal{C}_n^c \cap \mathcal{A}_n^c$ separately; then in step IV, we show the posterior probabilities of those

two areas are both $o_{\mathrm{pr}^*}(1)$; step V contains some auxiliary calculations to supplement the proof.

Step I: Bounding the posterior density

In this section, we give two different upper bounds for $p_n(\boldsymbol{\delta})$: the bound (S6.21) for $\boldsymbol{\delta} \in \mathcal{C}_n \cap \mathcal{A}_n^c$, and the bound (S6.23) for $\boldsymbol{\delta} \in \mathcal{C}_n^c \cap \mathcal{A}_n^c$.

We first consider $\boldsymbol{\delta} \in \mathcal{C}_n^c \cap \mathcal{A}_n^c$. By the convexity of L_n, we have

$$\begin{aligned} L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta} + \boldsymbol{\beta}^0) \quad &\leq \quad \frac{n\|G^{1/2}\boldsymbol{\delta}\|}{4q_0} \cdot \sup_{\|G^{1/2}\boldsymbol{\delta}\| \geq 4q_0} \left\{ \frac{L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta} + \boldsymbol{\beta}^0)}{n} \right\} \\ &\leq_{\mathrm{pr}^*} \quad -\frac{n\varepsilon_0\|G^{1/2}\boldsymbol{\delta}\|}{4q_0}, \end{aligned}$$

uniformly in $\boldsymbol{\delta} \in \mathcal{C}_n^c$ by Assumption 1. Therefore, combining with (S6.20), we have

$$\begin{aligned} p_n(\boldsymbol{\delta}) \quad &\leq_{\mathrm{pr}^*} \quad \exp\left\{ -\frac{n\varepsilon_0\|G^{1/2}\boldsymbol{\delta}\|_2}{q_0} - sn\lambda^2 \right\} \\ &\triangleq \quad \exp\left\{ -sn\lambda^2 \right\} \cdot \overline{p}_{1n}(\boldsymbol{\delta}), \end{aligned} \tag{S6.21}$$

uniformly when $\boldsymbol{\delta} \in \mathcal{C}_n^c \cap \mathcal{A}_n^c$.

Next we consider when $\boldsymbol{\delta} \in \mathcal{C}_n \cap \mathcal{A}_n$. Using Lemma 8, Lemma 7 and

(S4.7), we have

$$
\begin{aligned}
L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta}+\boldsymbol{\beta}^0) \;&\leq_{\mathrm{pr}^*}\; \sum_{i=1}^n \phi_i x_i^T \boldsymbol{\delta} - \frac{n}{8}(\boldsymbol{\delta}_{\mathcal{S}} + A_2\boldsymbol{\delta}_{\mathcal{S}^c})^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} + A_2\boldsymbol{\delta}_{\mathcal{S}^c}) + \frac{1}{8}\underline{f} \\
&\leq\; -\frac{n}{8}(\boldsymbol{\delta}_{\mathcal{S}} + A_2\boldsymbol{\delta}_{\mathcal{S}^c} - 4\Delta_s)^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} + A_2\boldsymbol{\delta}_{\mathcal{S}^c} - 4\Delta_s) \\
&\quad + \left\| \sum_{i=1}^n \phi_i \left(x_{i,\mathcal{S}^c} - A_2^T x_{i,\mathcal{S}} \right) \right\|_\infty \cdot \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 \\
&\quad + 2n\Delta_s^T G_{\mathcal{S}} \Delta_s + \frac{1}{8}\underline{f}, \qquad \text{(S6.22)}
\end{aligned}
$$

where $A_2 = G_{\mathcal{S}}^{-1} G_{\mathcal{S},\mathcal{S}^c}$; the first inequality relies on the Schur-decomposition of G; and (S6.22) follows by completing the squares with respect to $\boldsymbol{\delta}_{\mathcal{S}}$ and Hölder's inequality. Finally, letting $\alpha_n = \left\| \sum_{i=1}^n \phi_i \left(x_{i,\mathcal{S}^c} - A_2^T x_{i,\mathcal{S}} \right) \right\|_\infty$ and $\widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c}) = 4\Delta_s - A_2\boldsymbol{\delta}_{\mathcal{S}^c}$. Combining (S6.22) with (S6.20), we have

$$
\begin{aligned}
p_n(\boldsymbol{\delta}) \;&\lesssim_{\mathrm{pr}^*}\; \exp\left\{-sn\lambda^2 + 2n\Delta_s^T G_{\mathcal{S}}\Delta_s\right\} \cdot \exp\left\{-\frac{n}{8}(\boldsymbol{\delta}_{\mathcal{S}} - \widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c}))^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c}))\right\} \\
&\qquad \cdot \exp\left\{-(n\lambda - \alpha_n)\cdot \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\} \\
&\triangleq\; \exp\left\{-sn\lambda^2 + 2n\Delta_s^T G_{\mathcal{S}}\Delta_s\right\} \cdot \overline{p}_{2n}(\boldsymbol{\delta}_{\mathcal{S}}, \boldsymbol{\delta}_{\mathcal{S}^c}), \qquad \text{(S6.23)}
\end{aligned}
$$

uniformly when $\boldsymbol{\delta} \in \mathcal{C}_n \cap \mathcal{A}_n^c$.

Step II: Bounding the posterior integral on $\mathcal{C}_n \cap \mathcal{B}_n^c \cap \mathcal{A}_n^c$.

Here we bound the posterior integral of $p_n(\boldsymbol{\delta})$ on $\mathcal{C}_n \cap \mathcal{B}_n^c \cap \mathcal{A}_n^c$. Using the upper bound (S6.23) in step I, we relate the integration to standard probabilistic tail bounds. Given the data, let $Z \in \mathbb{R}^s$ and $\xi = (\xi_1, \ldots, \xi_{p-s})$

be distributed as

$$\xi_1, \ldots, \xi_{p-s} \overset{i.i.d.}{\sim} \text{Laplace}\left(\frac{1}{n\lambda - \alpha_n}\right),$$
$$Z \mid \xi \sim \text{N}\left(\widetilde{\mu}(\xi),\ \frac{4}{n} G_{\mathcal{S}}^{-1}\right),$$

where α_n and $\widetilde{\mu}$ are defined at the end of Step I. In what follows, we shall write $\text{pr}(\cdot)$ as the probability with respect to (Z, ξ) given Δ_s and α_n. For any fixed Δ_s and $\alpha_n < n\lambda$, the function $\overline{p}_{2n}(\boldsymbol{\delta}_{\mathcal{S}}, \boldsymbol{\delta}_{\mathcal{S}^c})$ is proportional to the joint density function of the vector (Z, ξ). Therefore, the integration of $\overline{p}_{2n}(\boldsymbol{\delta}_{\mathcal{S}}, \boldsymbol{\delta}_{\mathcal{S}^c})$ can be related to the probabilistic statements about (Z, ξ), which gives

$$\int_{\mathcal{C}_n \cap \mathcal{B}_n^c \cap \mathcal{A}_n^c} p_n(\boldsymbol{\delta})\, \mathrm{d}\boldsymbol{\delta} \lesssim_{\text{pr}^*} \exp\{-sn\lambda^2 + 2n\Delta_s^T G_{\mathcal{S}} \Delta_s\} \cdot \left(\frac{2}{n\lambda - \alpha_n}\right)^{p-s} \cdot \frac{1}{\sqrt{|G_{\mathcal{S}}|}} \cdot \left(\frac{8\pi}{n}\right)^{s/2}$$
$$\cdot\, \text{pr}\left(\|Z\|_2 \geq M_n \sqrt{\frac{s}{n}} \ \text{ or } \ \|\xi\|_\infty \geq \frac{M_n s \log p}{n\lambda}\right), \tag{S6.24}$$

where we insert the normalizing constants of Laplace and Gaussian distributions in the second equality; note the displayed equation holds on the event $N_1(\gamma) = \{p\alpha_n < \gamma \cdot n\lambda\}$ for small enough constant $\gamma > 0$.

To compute the tail probability in (S6.24), we break it into two parts. First, we have

$$\text{pr}\left(\|\xi\|_\infty \geq \frac{M_n s \log p}{n\lambda}\right) \leq p \cdot \exp\{-M_n s \log p/2\} \leq \exp\{-M_n s \log p/4\},$$

on the event $N_1(\gamma)$, which follows from Lemma 2. Second, using standard

conditional probability formula, we can show

$$\begin{aligned}
& \operatorname{pr}\left(\|Z\|_2 \geq M_n\sqrt{\frac{s}{n}}\ ,\ \|\xi\|_\infty \leq \frac{M_n s\log p}{n\lambda}\right) \\
\leq & \sup_{\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \leq M_n s\log p/(n\lambda)} \operatorname{pr}\left(\|Z\|_2 \geq M_n\sqrt{\frac{s}{n}}\ \Big|\ \xi = \boldsymbol{\delta}_{\mathcal{S}^c}\right) \\
\leq & \exp\left\{-\frac{\theta_0 M_n^2 s}{16}\right\},
\end{aligned}$$

where the last inequality holds on the event

$$N_2 = \left\{8n \cdot \sup_{\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \leq M_n s\log p/(n\lambda)} \left[\widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})^T G_{\mathcal{S}}\widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})\right] \leq M_n^2\theta_0 s\right\},$$

by Lemma 1; since $Z \mid \xi = \boldsymbol{\delta}_{\mathcal{S}^c}$ follows a Gaussian distribution with mean $\widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})$. Thus, combining the two tail bounds above, we have

$$\begin{aligned}
& \operatorname{pr}\left(\|Z\|_2 \geq M_n\sqrt{\frac{s}{n}}\ \text{ or }\ \|\xi\|_\infty \geq \frac{M_n s\log p}{n\lambda}\right) \\
\leq & \ 2\exp\left\{-c_0 M_n s\right\}, \qquad \text{(S6.25)}
\end{aligned}$$

for some constant $c_0 > 0$.

Substituting (S6.25) into (S6.24), we have

$$\begin{aligned}
& \int_{\mathcal{C}_n\cap\mathcal{B}_n^c\cap\mathcal{A}_n^c} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \\
\lesssim_{\mathrm{pr}^*} & \ \exp\{-sn\lambda^2 + 2n\Delta_s^T G_{\mathcal{S}}\Delta_s\}\cdot|G_{\mathcal{S}}|^{-1/2}\cdot\left(\frac{8\pi}{n}\right)^{s/2}\left(\frac{2}{n\lambda}\right)^{p-s} \\
& \cdot\exp\left\{-c_0 M_n s\right\}, \qquad \text{(S6.26)}
\end{aligned}$$

on the events $N_1(\gamma)$ and N_2, since we have

$$\left(\frac{2}{n\lambda-\alpha_n}\right)^{p-s} \ \leq\ \left(\frac{2}{n\lambda}\right)^{p-s}\cdot\frac{1}{1-\gamma}\ , \qquad \text{(S6.27)}$$

on the event $N_1(\gamma)$.

Step III: Bounding the posterior integral on $\mathcal{C}_n^c \cap \mathcal{A}_n^c$

When $\boldsymbol{\delta} \in \mathcal{C}_n^c \cap \mathcal{A}_n^c$, we first bound the posterior density $p_n(\boldsymbol{\delta})$ with (S6.21) in step I. Then, by letting $u = G^{1/2}\boldsymbol{\delta}$, the posterior integral is bounded by

$$\begin{aligned}
\int_{\mathcal{C}_n^c \cap \mathcal{A}_n^c} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} &\lesssim_{\mathrm{pr}^*} \exp\left\{-sn\lambda^2\right\} \cdot \int_{\mathcal{C}_n^c \cap \mathcal{A}_n^c} \overline{p}_{1n}(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \\
&\leq \exp\{-sn\lambda^2\} \cdot \frac{1}{\sqrt{|G|}} \cdot \int_{\|u\|_2 \geq 4q_0} \exp\left\{-\frac{n\varepsilon_0\|u\|_2}{4q_0}\right\}\mathrm{d}u \\
&= \exp\{-sn\lambda^2 + p - n\varepsilon_0\} \cdot \frac{1}{\sqrt{|G|}} \cdot \left(\frac{4q_0}{\sqrt{p}}\right)^p, \qquad \text{(S6.28)}
\end{aligned}$$

by the Cramér-Chernoff method (Boucheron et al., 2013, Section 2.2).

Step IV: Final bounds on the posterior probability

Recall the events $N_1(\gamma)$ and N_2 in Step II; and we further define a event $N_3(\gamma) = \left\{n\Delta_s^T G_{\mathcal{S}} \Delta_s \leq \gamma \cdot M_n s\right\}$. In step V later, we shall verify that all three events have pr*-probability going to 1, which implies that the upper bounds (S6.26) and (S6.28) hold with pr*-probability going to , in accordance with Lemma 6.

We close the proof by showing the posterior probability of $\mathcal{C}_n \cap \mathcal{B}_n^c \cap \mathcal{A}_n^c$ and $\mathcal{C}_n^c \cap \mathcal{A}_n^c$ are both $o_{\mathrm{pr}^*}(1)$. Since Lemma 15 implies that

$$\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \gtrsim_{\mathrm{pr}^*} \left(\frac{2\pi}{n}\right)^{s/2} \left(\frac{2}{n\lambda}\right)^{p-s} \cdot \frac{\exp\left(n \cdot \Delta_s^T G_{\mathcal{S}} \Delta_s/2 - sn\lambda^2\right)}{\sqrt{|G_{\mathcal{S}}|}} \triangleq \widetilde{P}_n,$$

it suffices to verify that

$$\int_{\mathcal{C}_n \cap \mathcal{B}_n^c \cap \mathcal{A}_n^c} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} + \int_{\mathcal{C}_n^c \cap \mathcal{A}_n^c} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} = o_{\mathrm{pr}^*}\left(\widetilde{P}_n\right).$$

For the first area $\mathcal{C}_n \cap \mathcal{B}_n^c \cap \mathcal{A}_n^c$, we compare $\widetilde{P}_n$ with (S6.26) in step II. After cancellation, we have

$$\begin{aligned}\frac{\int_{\mathcal{C}_n \cap \mathcal{B}_n^c \cap \mathcal{A}_n^c} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{P}_n} &\lesssim_{\mathrm{pr}^*} 2^s \cdot \exp\left\{\frac{3n}{2}\Delta_s^T G_{\mathcal{S}} \Delta_s - c_0 \cdot M_n s\right\} \\ &= o_{\mathrm{pr}^*}(1),\end{aligned}$$

which follows since the event $N_3(\gamma) = \left\{n\Delta_s^T G_{\mathcal{S}} \Delta_s \le \gamma M_n s\right\}$ has pr*-probability tending to 1.

For the second area $\mathcal{C}_n^c \cap \mathcal{A}_n^c$, first note

$$\widetilde{P}_n \ge \exp\left(-sn\lambda^2\right) \cdot \left(\frac{1}{n}\right)^p \cdot \frac{1}{\sqrt{|G_{\mathcal{S}}|}},$$

for all sufficiently large n since $\lambda \ll 1$. Therefore, we obtain from (S6.28) in step III that:

$$\begin{aligned}\frac{\int_{\mathcal{C}_n^c \cap \mathcal{A}_n^c} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{P}_n} &\lesssim_{\mathrm{pr}^*} \left(\frac{4nq_0}{\sqrt{p}}\right)^p \cdot \frac{1}{\sqrt{|\widetilde{G}_{\mathcal{S}^c}|}} \cdot \exp\{p - n\varepsilon_0\} \\ &= o_{\mathrm{pr}^*}(1), \qquad \text{(S6.29)}\end{aligned}$$

where $\widetilde{G}_{\mathcal{S}^c} = G_{\mathcal{S}^c} - G_{\mathcal{S}^c,\mathcal{S}} G_{\mathcal{S}}^{-1} G_{\mathcal{S},\mathcal{S}^c}$ is the Schur-complement of $G_{\mathcal{S}}$; we have used that $|\widetilde{G}_{\mathcal{S}^c}| \ge [\theta_{\min}(G)]^p \gtrsim (c_1 p)^{-p}$ for some constant $c_1 > 0$ by Assumption 2', and that $p\log(n \vee p) \ll n$.

Step V: Auxiliary calculations

Now we show that each of the events

$$N_1(\gamma) = \{p\alpha_n \le \gamma \cdot n\lambda\}, \quad N_3(\gamma) = \left\{n\Delta_s^T G_{\mathcal{S}} \Delta_s \le \gamma M_n s\right\},$$
$$\text{and} \quad N_2 = \left\{8n \cdot \sup_{\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \le M_n s \log p/(n\lambda)} \left[\widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})^T G_{\mathcal{S}} \widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})\right] \le M_n^2 \theta_0 s\right\},$$

holds with pr*-probability tending to 1, for all small enough $\gamma > 0$. Note α_n and $\widetilde{\mu}$ are given before (S6.23).

For $N_1(\gamma)$, let $A_2 = G_{\mathcal{S}}^{-1} G_{\mathcal{S},\mathcal{S}^c}$, $A = [-A_2^T, I_{p-s}]$ and $v_i = Ax_i \in \mathbb{R}^{p-s}$; then $\alpha_n = \|\sum_{i=1}^n \phi_i v_i\|_\infty$. Letting e_j be the jth standard unit vector in $\mathbb{R}^{p-s}$, we have

$$\begin{aligned} e_j^T ADA^T e_j &\lesssim e_j^T G_{\mathcal{S}^c} e_j \\ &\le c_0, \end{aligned}$$

uniformly for all $j = 1, \ldots, p-s$ from Assumption 2' and 3. Then, Lemma 13 implies that $\text{pr}^*\left(N_1^c(\gamma)\right) \to 0$ as $n \to \infty$, since $\lambda \gg p \log p/\sqrt{n}$.

For the event $N_3(\gamma)$, Lemma 14 directly implies that $\text{pr}^*(N_3(\gamma)) \to 1$ as $n \to \infty$ for all small enough $\gamma > 0$.

Finally, we consider the event N_2. Note that $\|x-y\|^2 \le 2(\|x\|^2 + \|y\|^2)$,

we have

$$\begin{aligned}
\sup_{\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \le M_n s \log p/(n\lambda)} \left[n \cdot \widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})^T G_{\mathcal{S}} \widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})\right] &\lesssim n \cdot \Delta_s^T G_{\mathcal{S}} \Delta_s + \sup_{\boldsymbol{\delta}_{\mathcal{S}^c}} \left[n \cdot \boldsymbol{\delta}_{\mathcal{S}^c}^T A_2^T G_{\mathcal{S}} A_2 \boldsymbol{\delta}_{\mathcal{S}^c}\right] \\
&\le O_{\mathrm{pr}^*}(s) + n \cdot \theta_{\max}(G) p \left(\frac{s M_n \log p}{n\lambda}\right)^2 \\
&= O_{\mathrm{pr}^*}(s) + o_{\mathrm{pr}^*}(M_n^2 s),
\end{aligned}$$

where the first inequality owns to Lemma 14 and the Schur-complements of G; the last inequality inequality holds as $\theta_{\max}(G) \lesssim p$ in Assumption 2'; the last equation holds since $\lambda \gg s^{1/2} p \log p / n^{1/2}$. Therefore, we conclude that $\mathrm{pr}^*(N_2) \to 1$. □

S6.2 Part 2 of Theorem 3

Now we prove part 2 of Theorem 3 with the CA prior. To this end, we need the following lemma, the proof of which is deferred to Section S8.

Lemma 17. *Under the conditions of Theorem 3, we have the following:*

$$\left\| \frac{f_n(\boldsymbol{\delta})}{\int_{\mathbb{R}^p} f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} - \frac{h_n(\boldsymbol{\delta})}{\int_{\mathbb{R}^p} h_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \right\|_{TV} \to 0,$$

in pr^**-probability, where* f_n *and* h_n *are defined in (S6.19).*

Proof of Theorem 3, part 2. In what follows, we shall write $\int f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}$ for integrating a function f_n on $\mathbb{R}^p$. In view of Lemma 17, we only need to

show that p_n converges to h_n,

$$\left\| \frac{p_n(\boldsymbol{\delta})}{\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} - \frac{h_n(\boldsymbol{\delta})}{\int h_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \right\|_{TV} \to 0,$$

in pr*-probability. Note Lemma 15 implies that

$$\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \gtrsim_{\mathrm{pr}^*} \left(\frac{2\pi}{n}\right)^{s/2} \left(\frac{2}{n\lambda}\right)^{p-s} \cdot \frac{T_n \cdot \exp\left(-sn\lambda^2\right)}{\sqrt{|G_{\mathcal{S}}|}} \triangleq \widetilde{P}_n.$$

Therefore, following the proof of Theorem 1 of Chernozhukov and Hong (2003), it suffices to show

$$\frac{\int_{\mathbb{R}^p} |p_n(\boldsymbol{\delta}) - h_n(\boldsymbol{\delta})|\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{P}_n \vee \int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0,$$

in pr*-probability.

For the diverging sequence $K_n \to +\infty$ that satisfies the condition in Lemma 9, we define

$$\mathcal{B}_n(K_n) = \left\{ \boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}_{\mathcal{S}}\|_2 \le K_n \sqrt{\frac{s}{n}}\ ,\ \text{and}\ \ \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \le K_n \frac{s\log p}{n\lambda} \right\},$$

in accordance with (S6.18). In the following, we upper bound the integral of $|p_n - h_n|$ on $\mathcal{B}_n(K_n)$ and its complement, separately in steps I and II.

Step I: Bounding the integral on $\boldsymbol{\mathcal{B}_n(K_n)}$

First, note that $\mathcal{B}_n(K_n) \subset \mathcal{A}_n^c$, the CA prior in p_n and h_n cancels as in

(S6.20); Lemma 9 further implies that

$$\begin{aligned}\sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)}\left|\log\left(\frac{h_n(\boldsymbol{\delta})}{p_n(\boldsymbol{\delta})}\right)\right| &= \sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)}\left|L_n(\boldsymbol{\beta}^0)-L_n(\boldsymbol{\delta}+\boldsymbol{\beta}^0)+\frac{n}{2}\boldsymbol{\delta}^TG\boldsymbol{\delta}-n\cdot\boldsymbol{\delta}_p^TG\boldsymbol{\delta}\right| \\ &= o_{\mathrm{pr}^*}(1),\end{aligned}$$

which further implies $|h_n(\boldsymbol{\delta})/p_n(\boldsymbol{\delta})-1| = o_{\mathrm{pr}^*}(1)$ uniformly when $\boldsymbol{\delta}\in\mathcal{B}_n(K_n)$. Therefore, we have:

$$\begin{aligned}\int_{\mathcal{B}_n(K_n)}|p_n(\boldsymbol{\delta})-h_n(\boldsymbol{\delta})|\,\mathrm{d}\boldsymbol{\delta} &= \int_{\mathcal{B}_n(K_n)}p_n(\boldsymbol{\delta})\left|1-\left(\frac{h_n(\boldsymbol{\delta})}{p_n(\boldsymbol{\delta})}\right)\right|\mathrm{d}\boldsymbol{\delta} \\ &= o_{\mathrm{pr}^*}\left(\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}\right). \qquad \text{(S6.30)}\end{aligned}$$

Step II: Bounding the integral on $[\mathcal{B}_n(K_n)]^c$

Here we analyze the integration of p_n and h_n separately on $[\mathcal{B}_n(K_n)]^c$. For p_n, part 1 of Theorem 3 directly implies that

$$\int_{[\mathcal{B}_n(K_n)]^c}p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}=o_{\mathrm{pr}^*}\left(\widetilde{P}_n\right).$$

Recall A_2 and α_n defined before (S6.23), and we further define $\tilde{\nu}(\boldsymbol{\delta}_{\mathcal{S}^c}) = \boldsymbol{\delta}_s - A_2\boldsymbol{\delta}_{\mathcal{S}^c}$. For $h_n(\boldsymbol{\delta})$, we first upper bound it by a similar argument as (S6.23):

$$\begin{aligned}h_n(\boldsymbol{\delta}) &= \exp\left\{-\frac{n}{2}\boldsymbol{\delta}^TG\boldsymbol{\delta}+n\boldsymbol{\delta}_s^TG\boldsymbol{\delta}-sn\lambda^2-n\lambda\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\} \\ &\le T_n\cdot\exp\left\{-sn\lambda^2-\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}}-\tilde{\nu}_1(\boldsymbol{\delta}_{\mathcal{S}^c}))^TG_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}}-\tilde{\nu}_1(\boldsymbol{\delta}_{\mathcal{S}^c}))-(n\lambda-\alpha_n)\cdot\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\} \\ &\triangleq T_n\cdot\exp\{-sn\lambda^2\}\cdot\overline{h}_n(\boldsymbol{\delta}_{\mathcal{S}},\boldsymbol{\delta}_{\mathcal{S}^c}),\end{aligned}$$

where T_n is defined before (S6.19).

Similar to (S6.24) in Section S6.1, we can relate the integration of $\overline{h}_n(\boldsymbol{\delta}_{\mathcal{S}}, \boldsymbol{\delta}_{\mathcal{S}^c})$ to Gaussian and Laplace tail bounds. Let $Z \in \mathbb{R}^s$ and $\xi = (\xi_1, \ldots, \xi_{p-s})$ be distributed as

$$\begin{aligned} \xi_1, \ldots, \xi_{p-s} &\overset{i.i.d.}{\sim} \text{Laplace}\left(\frac{1}{n\lambda - \alpha_n}\right), \\ Z \mid \xi &\sim \quad \mathrm{N}\left(\widetilde{\nu}(\xi),\ \frac{1}{n} G_{\mathcal{S}}^{-1}\right). \end{aligned}$$

Following (S6.24), we have

$$\begin{aligned} \int_{[\mathcal{B}_n(K_n)]^c} h_n(\boldsymbol{\delta})\, \mathrm{d}\boldsymbol{\delta} &\leq T_n \cdot \exp\{-sn\lambda^2\} \cdot \left(\frac{2}{n\lambda - \alpha_n}\right)^{p-s} \cdot \frac{1}{\sqrt{|G_{\mathcal{S}}|}} \cdot \left(\frac{2\pi}{n}\right)^{s/2} \\ &\quad \cdot \mathrm{pr}\left(\|Z\|_2 \geq K_n \sqrt{\frac{s}{n}} \ \text{ or } \ \|\xi\|_\infty \geq \frac{K_n s \log p}{n\lambda}\right) \\ &\lesssim \left(\frac{n\lambda}{n\lambda - \alpha_n}\right)^{p-s} \cdot \widetilde{P}_n \cdot \exp\left\{-c_0 K_n s\right\} \\ &\lesssim \widetilde{P}_n \cdot \exp\left\{-c_0 K_n s\right\}, \end{aligned} \tag{S6.31}$$

for some constant $c_0 > 0$; in the second inequality we bound the tail probability with (S6.25), which holds on the events

$$\begin{aligned} N_1(\gamma) &= \{p \cdot \alpha_n \leq \gamma \cdot n\lambda\}, \\ N_2 &= \left\{8n \cdot \sup_{\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \leq K_n^2 s \log p/(n\lambda)} \left[\tilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})^T G_{\mathcal{S}} \tilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c})\right] \leq K_n s \theta_{\min}(G_{\mathcal{S}})\right\}, \end{aligned}$$

for a small enough constant $\gamma > 0$; the last inequality follows from (S6.27).

Similar to Step V in Section S6.1, we can verify that both the events

$N_1(\gamma)$ and N_2 have pr*-probability tending to 1; therefore, we conclude that

$$\int_{[\mathcal{B}_n(K_n)]^c} h_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} = o_{\mathrm{pr}^*}\left(\widetilde{P}_n\right),$$

as $K_n \to \infty$.

Combining steps I and II, we obtain

$$\frac{\int_{\mathbb{R}^p} |p_n(\boldsymbol{\delta}) - h_n(\boldsymbol{\delta})|\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{P}_n \vee \int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0,$$

in pr*-probability, which concludes the proof by Theorem 1 of Chernozhukov and Hong (2003).

□

S7 Some technical lemmas in the regime of increasing dimensions

S7.1 Proof of Lemma 8 and Lemma 9

In this subsection, we prove two key results, Lemma 8 and 9, which controls the uniform variation of the quantile-loss function. When the dimension p grows with the sample size n, our results are new; they are not implied by the results in Chao et al. (2017) and Belloni et al. (2019) under our conditions on the design and dimensions. When p is fixed, those lemmas are standard from the empirical process literature; see e.g., Knight (1998),

Pollard (1985) and Andrews (1994).

Proof of Lemma 8. First, it is easy to see $|r_n(\boldsymbol{\delta})| \leq 2\sum_{i=1}^n |\boldsymbol{x}_i^T\boldsymbol{\delta}|$. Therefore, for large enough $a > 0$, we have that

$$\begin{aligned}
\sup_{\|D^{1/2}\boldsymbol{\delta}\|\geq n^a}\left|\frac{r_n(\boldsymbol{\delta})-\mathrm{E}^*[r_n(\boldsymbol{\delta})]}{n\|D^{1/2}\boldsymbol{\delta}\|^2+1}\right| &\leq \sup_{\|D^{1/2}\boldsymbol{\delta}\|\geq n^a}\frac{\sum_{i=1}^n|\boldsymbol{x}_i^T\boldsymbol{\delta}|+\sum_{i=1}^n\mathrm{E}^*\left[|\boldsymbol{x}_i^T\boldsymbol{\delta}|\right]}{n\cdot\|D^{1/2}\boldsymbol{\delta}\|^2}\\
&\leq \frac{\sum_{i=1}^n\|D^{-1/2}\boldsymbol{x}_i\|}{n^{1+a}}+\frac{\mathrm{E}^*\left[\|D^{-1/2}\boldsymbol{x}_i\|\right]}{n^a}\\
&\to 0,
\end{aligned}$$

in pr*-probability, where the last inequality follows since $\mathrm{Cov}^*(D^{-1/2}\boldsymbol{x}_i) = I_p$. Therefore, it suffices to show

$$\sup_{\boldsymbol{\delta}:\|D^{1/2}\boldsymbol{\delta}\|\leq n^a}\left|\frac{r_n(\boldsymbol{\delta})-\mathrm{E}^*[r_n(\boldsymbol{\delta})]}{n\|D^{1/2}\boldsymbol{\delta}\|+1}\right|\to 0,$$

in pr*-probability, for any constant $a > 0$.

Let $\gamma_n^4 = (p^2\log^2 n)/n \to 0$. In the following steps, we apply a generic chaining argument to show that the above display is of order $O_{\mathrm{pr}^*}(\gamma_n)$. To simplify notations, we define $f_n(\boldsymbol{\delta}) = r_n(\boldsymbol{\delta})/(n\|D^{1/2}\boldsymbol{\delta}\|^2+1)$.

Step I: Main chaining

First, we define the following concentric 'cubes':

$$\mathcal{C}_k = \{\boldsymbol{\delta}\in\mathbb{R}^p : \|D^{1/2}\boldsymbol{\delta}\|_\infty \leq d_k\},\quad k = 0,\ldots,K_n,$$

where d_k is the edge length of each cube; we take those lengths to be

$$d_k = (k+1)\varepsilon_n,\quad \varepsilon_n = \frac{\gamma_n}{np},\quad k = 0,\ldots,K_n.$$

Letting $K_n = \lceil pn^{a+1}/\gamma_n \rceil - 1$, it is easy to check $\{\|D^{1/2}\boldsymbol{\delta}\| \le n^a\} \subset \mathcal{C}_{K_n}$. It then suffices to show the desired uniform convergence in $\boldsymbol{\delta} \in \mathcal{C}_{K_n}$.

For each of $\mathcal{C}_k \backslash \mathcal{C}_{k-1}$, we further partition it into smaller cubes of length most ε_n. That is, $\mathcal{C}_k$ builds upon $\mathcal{C}_{k-1}$ by one layer of such small cubes with edge length ε_n. For each $k \ge 1$, there are at most $B_k = (2k)^p - [2(k-1)]^p$ such small cubes, which are denoted as $\mathcal{C}_k^j$, $j = 1, \ldots, B_k$. For any $\boldsymbol{\delta}, \boldsymbol{\delta}' \in \mathcal{C}_k^j$, we have $\|D^{1/2}(\boldsymbol{\delta} - \boldsymbol{\delta}')\|_\infty \le \varepsilon_n$. Letting $\boldsymbol{\delta}_k^j$ be the center of $\mathcal{C}_k^j$, we have

$$
\begin{aligned}
\sup_{\boldsymbol{\delta}:\|D^{1/2}\boldsymbol{\delta}\|\le n^a} |f_n(\boldsymbol{\delta}) - \mathrm{E}^*(f_n(\boldsymbol{\delta}))| &\le \sup_{\boldsymbol{\delta}\in\mathcal{C}_0} |f_n(\boldsymbol{\delta}) - \mathrm{E}^*(f_n(\boldsymbol{\delta}))| + \sup_{\substack{k=1,\ldots,K_n \\ j=1,\ldots,B_k}} \sup_{\boldsymbol{\delta}\in\mathcal{C}_k^j} |f_n(\boldsymbol{\delta}) - \mathrm{E}^*(f_n(\boldsymbol{\delta}))| \\
&\le \left(\sup_{\boldsymbol{\delta}\in\mathcal{C}_0} |f_n(\boldsymbol{\delta}) - \mathrm{E}^*(f_n(\boldsymbol{\delta}))|\right) \\
&\quad + \left(\sup_{\substack{k=1,\ldots,K_n \\ j=1,\ldots,B_k}} \sup_{\boldsymbol{\delta}\in\mathcal{C}_k^j} \left[|f_n(\boldsymbol{\delta}) - f_n(\boldsymbol{\delta}_k^j)| + \mathrm{E}^*|f_n(\boldsymbol{\delta}) - f_n(\boldsymbol{\delta}_k^j)|\right]\right) \\
&\quad + \left(\sup_{\substack{k=1,\ldots,K_n \\ j=1,\ldots,B_k}} |f_n(\boldsymbol{\delta}_k^j) - \mathrm{E}^*[f_n(\boldsymbol{\delta}_k^j)]|\right) \\
&\triangleq R_1 + R_2 + R_3.
\end{aligned}
$$

Step II: Auxiliary chaining

In this step, we compute the stochastic order of R_1, R_2, and R_3 separately.

Let $v(\boldsymbol{\delta}) = \|D^{1/2}\boldsymbol{\delta}\|^2$; for any $\boldsymbol{\delta}, \boldsymbol{\delta}' \neq \mathbf{0}$, define

$$\begin{aligned}
\Delta_0(\boldsymbol{\delta}) &= |f_n(\boldsymbol{\delta})| = \frac{|r_n(\boldsymbol{\delta})|}{n\|D^{1/2}\boldsymbol{\delta}\|^2 + 1}, \\
\Delta_1(\boldsymbol{\delta}, \boldsymbol{\delta}') &= \frac{|r_n(\boldsymbol{\delta}') - r_n(\boldsymbol{\delta})|}{n\|D^{1/2}\boldsymbol{\delta}'\|^2 + 1}, \\
\Delta_2(\boldsymbol{\delta}, \boldsymbol{\delta}') &= \left| r_n(\boldsymbol{\delta}) \cdot \frac{1}{n\|D^{1/2}\boldsymbol{\delta}\|^2 + 1} - \frac{1}{n\|D^{1/2}\boldsymbol{\delta}'\|^2 + 1} \right|.
\end{aligned} \tag{S7.32}$$

In step IV below, we show that for any $d < 1/\sqrt{n}$,

$$\mathrm{E}^* \left[\sup_{\substack{\boldsymbol{\delta}, \boldsymbol{\delta}' \subset \mathcal{C}_{K_n} \\ \|D^{1/2}(\boldsymbol{\delta}' - \boldsymbol{\delta})\| \leq d}} \Delta_k(\boldsymbol{\delta}, \boldsymbol{\delta}') \right] \lesssim n\sqrt{p}d, \quad k = 0, 1, 2.$$

We bound R_1 and R_2 using Δ_0 through Δ_2 defined above.

Recall $d_0 = \varepsilon_n = \gamma_n/(np)$. For R_1, note when $\boldsymbol{\delta} \in \mathcal{C}_0$ we have $\|D^{1/2}\boldsymbol{\delta}\| \leq \sqrt{p}\|D^{1/2}\boldsymbol{\delta}\|_\infty \leq \sqrt{p}d_0 \leq 1/\sqrt{n}$, therefore

$$\mathrm{E}^*[R_1] \leq 2\mathrm{E}^* \left[\sup_{\|D^{1/2}\boldsymbol{\delta}\| \leq \sqrt{p}\varepsilon_n} \Delta_0(\boldsymbol{\delta}) \right] \lesssim np\varepsilon_n \leq \gamma_n.$$

For R_2, note for any $\boldsymbol{\delta}, \boldsymbol{\delta}' \neq 0$, we have

$$\begin{aligned}
|f_n(\boldsymbol{\delta}') - f_n(\boldsymbol{\delta})| &\leq \frac{|r_n(\boldsymbol{\delta}') - r_n(\boldsymbol{\delta})|}{n\|D^{1/2}\boldsymbol{\delta}'\|^2 + 1} + \left| r_n(\boldsymbol{\delta}) \cdot \frac{1}{n\|D^{1/2}\boldsymbol{\delta}\|^2 + 1} - \frac{1}{n\|D^{1/2}\boldsymbol{\delta}'\|^2 + 1} \right| \\
&= \Delta_1(\boldsymbol{\delta}, \boldsymbol{\delta}') + \Delta_2(\boldsymbol{\delta}, \boldsymbol{\delta}');
\end{aligned}$$

furthermore, whenver $\boldsymbol{\delta} \in \mathcal{C}_k^j$, we have $\|D^{1/2}(\boldsymbol{\delta} - \boldsymbol{\delta}_k^j)\| \leq \sqrt{p}\varepsilon_n \leq 1/\sqrt{n}$. Therefore

$$\mathrm{E}^*[R_2] \leq 2\sum_{u=1}^{2} \left(\mathrm{E}^* \left[\sup_{\substack{\boldsymbol{\delta}, \boldsymbol{\delta}' \subset \mathcal{C}_{K_n} \\ \|D^{1/2}(\boldsymbol{\delta}' - \boldsymbol{\delta})\| \leq \sqrt{p}\varepsilon_n}} \Delta_u(\boldsymbol{\delta}, \boldsymbol{\delta}') \right] \right) \lesssim np\varepsilon_n \leq \gamma_n.$$

Chebyshev's inequality then implies

$$R_1 = O_{\mathrm{pr}^*}(\gamma_n), \qquad R_2 = O_{\mathrm{pr}^*}(\gamma_n).$$

Next we bound R_3. In step III below, we show that for any fixed $\boldsymbol{\delta}$, the following inequality holds for all $t_n > 0$:

$$\mathrm{pr}^*\left(|r_n(\boldsymbol{\delta}) - \mathrm{E}^*[r_n(\boldsymbol{\delta})]| \geq t_n\right) \;\leq\; 2\exp\left\{-\frac{t_n^2}{2\left(c_1 n[v(\boldsymbol{\delta})]^{3/2} + c_2\sqrt{v(\boldsymbol{\delta})}t_n\right)}\right\}, \tag{S7.33}$$

where $v(\boldsymbol{\delta}) = \|D^{1/2}\boldsymbol{\delta}\|^2$ and $c_1, c_2 > 0$ are two constants. Recall there are at most $(2K_n)^p \leq (4pn^{a+1}/\gamma_n)^p$ small cubes with edge-length ε_n; therefore, for large enough $M > 0$,

$$\begin{aligned}
\mathrm{pr}^*(R_3 \geq M\gamma_n) &\leq \sum_{k=1}^{K_n}\sum_{j=1}^{B_k}\mathrm{pr}^*\left(\frac{r_n(\boldsymbol{\delta}_k^j) - \mathrm{E}^*\left[r_n(\boldsymbol{\delta}_k^j)\right]}{n\|D^{1/2}\boldsymbol{\delta}\|^2 + 1} \geq M\gamma_n\right) \\
&\leq \left(\frac{4pn^{a+1}}{\gamma_n}\right)^p \cdot \exp\left\{-\inf_{\boldsymbol{\delta}\in\mathcal{C}_{K_n}} \frac{M^2\gamma_n^2(nv(\boldsymbol{\delta})+1)^2}{2\left(c_1 n[v(\boldsymbol{\delta})]^{3/2} + c_2\sqrt{v(\boldsymbol{\delta})}\cdot M\gamma_n(nv(\boldsymbol{\delta})+1)\right)}\right\} \\
&\leq \exp\left\{(a+2)p\log n - \frac{M^2\sqrt{n}\gamma_n^2}{c_1}\right\} \\
&\to 0,
\end{aligned}$$

when M is large enough, since $n^{1/2}\gamma_n^2 = p\log n$; to compute the infimum in the penultimate inequality, we define $z = n\{v(\boldsymbol{\delta})\}^{1/2} + 1/\{v(\boldsymbol{\delta})\}^{1/2}$, which gives

$$\frac{(nv(\boldsymbol{\delta})+1)^2}{c_1 n[v(\boldsymbol{\delta})]^{3/2} + c_2\sqrt{v(\boldsymbol{\delta})}\cdot M\gamma_n(nv(\boldsymbol{\delta})+1)} \geq \frac{z^2}{(c_1 + M\gamma_n\cdot c_2)z} \geq \frac{\sqrt{n}}{c_1}.$$

Collecting the results for R_1, R_2 and R_3 and recalling that $\gamma_n^4 = (p^2 \log^2 n)/n$, we have

$$\sup_{\boldsymbol{\delta}:\|D^{1/2}\boldsymbol{\delta}\|\le n^a} |f_n(\boldsymbol{\delta}) - \mathrm{E}^*(f_n(\boldsymbol{\delta}))| = O_{\mathrm{pr}^*}\left(\sqrt[4]{\frac{p^2\log^2 n}{n}}\right) = o_{\mathrm{pr}^*}(1),$$

since $p^2 \log^2 n \ll n$. Thus, we have shown the asserted claim of the Lemma.

Step III: Exponential inequality

Here we show the exponential inequality (S7.33) holds. Without loss of generality, we assume the scale-parameter $\sigma_0 = 1$ in Assumption 2' of the paper; therefore, standard calculation leads to

$$\mathrm{E}^*\left[|\boldsymbol{x}_i^T D^{-1/2} u|^q\right] \lesssim q! \cdot \|u\|^q. \tag{S7.34}$$

Note for fixed $\boldsymbol{\delta}$, $r_n(\boldsymbol{\delta})$ can be written as

$$r_n(\boldsymbol{\delta}) = \sum_{i=1}^n \int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} \left(1[y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}^0 \le s] - 1[y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}^0 \le 0]\right) \mathrm{d}s \triangleq \sum_{i=1}^n \int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} h_i(s)\,\mathrm{d}s, \tag{S7.35}$$

which follows directly from Knight's identity (Knight, 1998); note the above summands $\int h_i(s)\mathrm{d}s$ are non-negative. Below we apply Lemma 4 to bound the tail probability of $r_n(\boldsymbol{\delta})$.

We check the conditions for Lemma 4. Let $F_i(y)$ and $f_i(y)$ be the conditional distribution function and conditional density function of $(Y -$

$X^T\boldsymbol{\beta}^0) \mid X = \boldsymbol{x}_i$, respectively. Letting $A_n = n\|D^{1/2}\boldsymbol{\delta}\|^3$, we first have

$$\begin{aligned}
\sum_{i=1}^n \mathrm{E}^*\left[\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} h_i(s)\,\mathrm{d}s\right]^2 &= \sum_{i=1}^n \mathrm{E}_X^*\left[\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} \mathrm{E}_{Y|X=x}^*\left\{h_i(u)h_i(s)\right\}\,\mathrm{d}s\,\mathrm{d}u\right] \\
&= n\cdot \mathrm{E}_X^*\left[\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} F_i(u\wedge s)+F_i(0)-F_i(u\wedge 0)-F_i(s\wedge 0)\,\mathrm{d}s\,\mathrm{d}u\right] \\
&\le 2n\cdot \mathrm{E}_X^*\left[\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\mathrm{d}s\int_0^s |u|f_i(\tilde{u})\,\mathrm{d}u\right] \\
&\lesssim A_n,
\end{aligned}$$

where the first inequality owns to the mean value theorem; the last inequality follows from Assumption 3 and the moment calculation (S7.34). Next, it is easy to see from induction that for all integers $q \ge 3$,

$$\begin{aligned}
\sum_{i=1}^n \mathrm{E}^*\left[\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} h_i(s)\,\mathrm{d}s\right]^q &\le \sum_{i=1}^n \mathrm{E}^*\left(|\boldsymbol{x}_i^T\boldsymbol{\delta}|^q\cdot \mathbf{1}_{[0\le|y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}^0|\le|\boldsymbol{x}_i^T\boldsymbol{\delta}|]}\right) \\
&\lesssim n\cdot \mathrm{E}_X^*\left(|\boldsymbol{x}_i^T\boldsymbol{\delta}|^{q+1}\right) \\
&\le q!\cdot A_n\cdot B_n^{q-2},
\end{aligned}$$

where $B_n = 2\|D^{1/2}\boldsymbol{\delta}\|$, and the last inequality follows from (S7.34).

Recall that $v(\boldsymbol{\delta}) = \|D^{1/2}\boldsymbol{\delta}1\|$. Applying Lemma 4 gives

$$\begin{aligned}
\mathrm{pr}^*\left(\sum_{i=1}^n\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}}(h_i(s)-\mathrm{E}^*[h_i(s)])\,\mathrm{d}s \ge t_n\right) &\le \exp\left\{-\frac{t_n^2}{2(c_1A_n+c_2B_nt_n)}\right\} \\
&\le \exp\left\{-\frac{t_n^2}{2\left(nc_1\{v(\boldsymbol{\delta})\}^{3/2}+2c_2\{v(\boldsymbol{\delta})\}^{1/2}t_n\right)}\right\},
\end{aligned}$$

which is precisely the one-sided version of (S7.33). The inequality for the opposite direction follows in a similar manner since $\int h_i(s)\mathrm{d}s \ge 0$.

Step IV: Control of the supremum Here we compute the expectation of the supremum of Δ_0, Δ_1 and Δ_2 defined in (S7.32).

For Δ_1, from the proof of Theorem 1 in Pollard (1991), we deduce,

$$\begin{aligned}
\mathrm{E}^* \sup_{\substack{\boldsymbol{\delta},\boldsymbol{\delta}'\in\mathcal{C}_{K_n}\\ \|D^{1/2}(\boldsymbol{\delta}'-\boldsymbol{\delta})\|\le d}} |r_n(\boldsymbol{\delta})-r_n(\boldsymbol{\delta}')| &\le \mathrm{E}^* \sup_{\boldsymbol{\delta},\boldsymbol{\delta}'} \left(\sum_n |\boldsymbol{x}_i^T(\boldsymbol{\delta}'-\boldsymbol{\delta})|\cdot 1[|y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}^0|\le|\boldsymbol{x}_i^T\boldsymbol{\delta}|\vee|\boldsymbol{x}_i^T\boldsymbol{\delta}'|]\right)\\
&\le \mathrm{E}^* \sup_{\boldsymbol{\delta},\boldsymbol{\delta}'} \left(\sum_{i=1}^n \|\boldsymbol{x}_i^T D^{-1/2}\|\cdot\|D^{1/2}(\boldsymbol{\delta}'-\boldsymbol{\delta})\|\right)\\
&\le nd\sqrt{p},
\end{aligned}$$

since $\mathrm{E}^*[\|D^{-1/2}\boldsymbol{x}_i\|]\le\{\mathrm{E}^*[\|D^{-1/2}\boldsymbol{x}_i\|^2]\}^{1/2}=p^{1/2}$. Observing the denominator of $f_n(\boldsymbol{\delta})$ is no less than 1, we obtain for $d<1$,

$$\mathrm{E}^* \sup_{\substack{\boldsymbol{\delta},\boldsymbol{\delta}'\in\mathcal{C}_{K_n}\\ \|D^{1/2}(\boldsymbol{\delta}'-\boldsymbol{\delta})\|\le d}} \Delta_1(\boldsymbol{\delta}',\boldsymbol{\delta})\lesssim nd\sqrt{p}.$$

For Δ_2, observe that for any $\|D^{1/2}(\boldsymbol{\delta}-\boldsymbol{\delta}')\|\le d$, we have $|v(\boldsymbol{\delta}')-v(\boldsymbol{\delta})|\le d^2+2d\{v(\boldsymbol{\delta})\}^{1/2}$, which implies

$$\left|\frac{1}{n\|D^{1/2}\boldsymbol{\delta}\|^2+1}-\frac{1}{n\|D^{1/2}\boldsymbol{\delta}'\|^2+1}\right|\le\frac{n|v(\boldsymbol{\delta})-v(\boldsymbol{\delta}')|}{(nv(\boldsymbol{\delta})+1)(nv(\boldsymbol{\delta}')+1)}\le\frac{nd^2+2nd\sqrt{v(\boldsymbol{\delta})}}{(nv(\boldsymbol{\delta})+1)}.$$

Furthermore, note $r_n(\boldsymbol{\delta})\le\sum_{i=1}^n|\boldsymbol{x}_i^T\boldsymbol{\delta}|$ from (S7.35), we obtain:

$$\begin{aligned}
\mathrm{E}^* \sup_{\substack{\boldsymbol{\delta},\boldsymbol{\delta}'\in\mathcal{C}_{K_n}\\ \|D^{1/2}(\boldsymbol{\delta}'-\boldsymbol{\delta})\|\le d}} \Delta_2(\boldsymbol{\delta},\boldsymbol{\delta}') &\le \mathrm{E}^* \sup_{\boldsymbol{\delta}\in\mathcal{C}_{K_n}} \left[\sum_{i=1}^n|\boldsymbol{x}_i^T\boldsymbol{\delta}|\cdot\frac{nd^2+2nd\sqrt{v(\boldsymbol{\delta})}}{(nv(\boldsymbol{\delta})+1)}\right]\\
&\le \mathrm{E}^*\left[\sum_{i=1}^n\|D^{-1/2}\boldsymbol{x}_i\|\right]\cdot\sup_{\boldsymbol{\delta}\in\mathcal{C}_{K_n}}\left(\sqrt{v(\boldsymbol{\delta})}\cdot\frac{nd^2+2nd\sqrt{v(\boldsymbol{\delta})}}{(nv(\boldsymbol{\delta})+1)}\right)\\
&\lesssim n\sqrt{p}(\sqrt{n}d^2+2d).
\end{aligned}$$

where the supremum in the penultimate inequality is bounded by $2(n^{1/2}d^2 + 2d)$. Therefore, when $d \leq 1/\sqrt{n}$, the above display is bounded by $n\sqrt{p}d$.

For Δ_0, from (S7.35) we have

$$\begin{aligned}
\mathrm{E}^* \sup_{\boldsymbol{\delta}:\|D^{1/2}\boldsymbol{\delta}\|\leq d} f_n(\boldsymbol{\delta}') &\leq \mathrm{E}^* \sup_{\boldsymbol{\delta}:\|D^{1/2}\boldsymbol{\delta}\|\leq d} |r_n(\boldsymbol{\delta})| \\
&\leq \mathrm{E}^* \sup_{\boldsymbol{\delta}} \left(\sum_n |\boldsymbol{x}_i^T\boldsymbol{\delta}| \cdot 1\big[y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}^0| \leq |\boldsymbol{x}_i^T\boldsymbol{\delta}|\big] \right) \\
&\leq \mathrm{E}^* \left(\sum_{i=1}^n d \cdot \|\boldsymbol{x}_i^T D^{-1/2}\| \cdot 1\big[|y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}^0| \leq d \cdot \|\boldsymbol{x}_i^T D^{-1/2}\|\big] \right) \\
&\lesssim nd^2p.
\end{aligned}$$

□

Proof of Lemma 9. We shall specify K_n later. Recall the definition of $r_n(\boldsymbol{\delta})$ in Lemma 8, it suffices to show

$$\sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \left| r_n(\boldsymbol{\delta}) - \frac{n}{2}\boldsymbol{\delta}^T G\boldsymbol{\delta} \right| = o_{\mathrm{pr}^*}(1).$$

First, we have

$$\begin{aligned}
\sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \|D^{1/2}\boldsymbol{\delta}\|^2 &\leq 2 \sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)} \left(\boldsymbol{\delta}_{\mathcal{S}}^T D_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}} + \boldsymbol{\delta}_{\mathcal{S}^c}^T D_{\mathcal{S}^c}\boldsymbol{\delta}_{\mathcal{S}^c}\right) \\
&\lesssim \frac{sK_n^2}{n} + p^2\left(\frac{sK_n\log p}{n\lambda}\right)^2 \\
&\lesssim \frac{sK_n^2}{n}, \qquad \text{(S7.36)}
\end{aligned}$$

since $\theta_{\max}(D) \leq p$ from Assumption 2', and that $\lambda \gg s^{1/2}p\log p/n^{1/2}$.

Lemma 8 then implies that

$$\sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)}\left|\frac{r_n(\boldsymbol{\delta})-\mathrm{E}^*[r_n(\boldsymbol{\delta})]}{K_n^2 s+1}\right| \lesssim \sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)}\left|\frac{r_n(\boldsymbol{\delta})-\mathrm{E}^*[r_n(\boldsymbol{\delta})]}{n\|D^{1/2}\boldsymbol{\delta}\|^2+1}\right| = O_{\mathrm{pr}^*}\left(\sqrt[4]{\frac{p^2\log^2 p}{n}}\right).$$

Therefore, for $K_n \to \infty$ such that $K_n^8 \ll n/(s^4p^2\log^2 p)$, we have

$$\sup_{\boldsymbol{\delta}\in\mathcal{B}_n(K_n)}|r_n(\boldsymbol{\delta})-\mathrm{E}^*[r_n(\boldsymbol{\delta})]| = o_{\mathrm{pr}^*}(1),$$

if $s^4p^2\log^2 p = o(n)$.

Next we compute $\mathrm{E}^*[r_n(\boldsymbol{\delta})]$. For simplicity, let F_i and f_i be the conditional distribution/density function for $(Y - X^T\boldsymbol{\beta}^0) \mid X = \boldsymbol{x}_i$. Note

$$\frac{1}{2}\boldsymbol{\delta}^T G\boldsymbol{\delta} = \mathrm{E}^*\left(\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} [sf_i(0)]\ \mathrm{d}s\right).$$

By Knight's identity in (S7.35), we have

$$\begin{aligned}
\left|\mathrm{E}^*[r_n(\boldsymbol{\delta})] - \frac{n}{2}\boldsymbol{\delta}^T G\boldsymbol{\delta}\right| &= n\left|\mathrm{E}_X^*\left(\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} [F_i(s)-F_i(0)-sf_i(0)]\ \mathrm{d}s\right)\right| \\
&\le n\left|\mathrm{E}_X^*\left(\int_0^{\boldsymbol{x}_i^T\boldsymbol{\delta}} \left[L_1 s^2\right]\ \mathrm{d}s\right)\right| \\
&\lesssim n\cdot\|D^{1/2}\boldsymbol{\delta}\|^3 \\
&= O_{\mathrm{pr}^*}\left(\left\{\frac{s^3K_n^6}{n}\right\}^{1/2}\right),
\end{aligned}$$

uniformly on $\boldsymbol{\delta} \in \mathcal{B}_n(K_n)$, where L_1 is the uniform Lipshitz constant in Assumption 3, the penultimate inequality owns to the moment calculation (S7.34), and the last inequality uses (S7.36). Therefore, the proof is now complete by choosing $K_n^6 \ll n/s^3$. □

The following result is a useful corollary from the above lemmas.

Corollary 1. *Define*

$$\mathcal{E}_n(K_n) = \left\{ \|\boldsymbol{\delta}_{\mathcal{S}}\|_2 \le K_n \sqrt{\frac{s}{n}}\,;\ \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \le K_n \frac{\log p}{n\lambda} \right\}.$$

Under the condition of Lemma 9, there exists a sequence $K_n \to +\infty$ such that

$$\sup_{\boldsymbol{\delta}\in\mathcal{E}_n(K_n)} |r_n(\boldsymbol{\delta}) - \frac{n}{2}\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}}| = o_{\mathrm{pr}^*}(1),$$

where $r_n(\boldsymbol{\delta})$ is defined in Lemma 8.

Proof. We only need to verify $n\boldsymbol{\delta}^T G\boldsymbol{\delta} - n\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}} = o_{\mathrm{pr}^*}(1)$ on $\boldsymbol{\delta} \in \mathcal{E}_n(K_n)$. Observe that

$$\begin{aligned}
\left|n\boldsymbol{\delta}^T G\boldsymbol{\delta} - n\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}}\right| &\le n\boldsymbol{\delta}_{\mathcal{S}^c}^T G_{\mathcal{S}^c}\boldsymbol{\delta}_{\mathcal{S}^c} + 2n\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S},\mathcal{S}^c}\boldsymbol{\delta}_{\mathcal{S}^c} \\
&\le n\theta_{\max}(G)p\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty^2 + 2n\|\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S}}^{1/2}\| \cdot \|G_{\mathcal{S}}^{-1/2} G_{\mathcal{S},\mathcal{S}^c}\boldsymbol{\delta}_{\mathcal{S}^c}\| \\
&\lesssim np^2\left(\frac{K_n \log p}{n\lambda}\right)^2 + 2nK_n^2\sqrt{\frac{s}{n}} \cdot \sqrt{\theta_{\max}(G)p} \cdot \frac{\log p}{n\lambda} \\
&\lesssim K_n^2 \frac{p^2 \log^2 p}{n\lambda^2} + K_n^2 \frac{\sqrt{s}p\log p}{\sqrt{n}\lambda} \\
&\to 0,
\end{aligned}$$

provided that K_n diverges slow enough; the second to the last inequality holds due to $\|G_{\mathcal{S}}^{-1/2} G_{\mathcal{S},\mathcal{S}^c}\|^2 \le \theta_{\max}(G) \le p$ as in Assumption 2'. □

S7.2 Proof of Lemma 7, 13 and 14

In this subsection, we give the proof for three other auxiliary lemmas in the asymptotic regime of increasing dimensions. We first prove Lemma 7.

Proof of Lemma 7. From Lemma 4 of Belloni et al. (2011), it suffices to verify the Restricted Non-linearity Condition hold, i.e., z_n is uniformly bounded away from 0 under Assumptions 2' and 3, where

$$z_n = \inf_{\boldsymbol{\delta}\in\mathbb{R}^p, \boldsymbol{\delta}\neq 0} \frac{[\mathrm{E}^*(|X^T\boldsymbol{\delta}|^2)]^{3/2}}{\mathrm{E}^*(|X^T\boldsymbol{\delta}|^3)}.$$

First, from the moment calculation in (S7.34), we have, under Assumption 2', that $\mathrm{E}^*(|X^T\boldsymbol{\delta}|^3) \lesssim \|D^{1/2}\boldsymbol{\delta}\|^3$. Therefore

$$z_0 \gtrsim \inf_{\boldsymbol{\delta}\neq 0} \frac{(\boldsymbol{\delta}^T D\boldsymbol{\delta})^{3/2}}{\|D^{1/2}\boldsymbol{\delta}\|^3} = 1,$$

which concludes the proof by following the proof of Lemma 4 of Belloni et al. (2011). □

Proof of Lemma 13. For each $k = 1, \ldots, q$, let e_k be the kth standard unit vector in $\mathbb{R}^q$ and $v_k = D^{1/2}A^T e_k$. For each $k \in \{1, \ldots, q\}$, the kth element of AX is $v_k^T D^{-1/2} Z$; they are uniformly sub-exponential in the sense of

$$\sup_{k\in\{1,\ldots,q\}} \mathrm{pr}^*\left(|(AX)_k| \geq z\right) \quad \leq \quad 2\exp\left\{-\frac{z}{C_0\sigma_0}\right\},$$

for some constant $C_0 > 0$ by Assumption 2'.

Furthermore, note that ϕ_i and $\boldsymbol{x}_i$ are independent for each i. For sufficiently large M, we have:

$$\begin{aligned}\mathrm{pr}^*\left(\left\|\sum_{i=1}^n \phi_i A\boldsymbol{x}_i\right\|_\infty \geq M\sqrt{n\log q}\right) &\leq q\cdot \mathrm{E}^*_\phi\left[\exp\left\{-C_2\frac{M^2 n\log q}{\|\phi\|_2^2}\right\}\right]\\ &\leq \exp\left\{-(C_2/2)\cdot M^2\log(q)\right\}.\end{aligned}$$

where the first inequality holds for some constant $C_2 > 0$ by the Bernstein inequality (Lemma 5); the last inequality follows since $\|\phi\|_\infty \leq 1$. The proof is now complete.

□

Proof of Lemma 14. We only prove the result for $\Delta_s^T G_{\mathcal{S}}\Delta_s$, the result for $\Delta_p^T G\Delta_p$ follows in a similar fashion. Let $\phi_i = \phi_\tau(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}^0)$, $\phi = [\phi_1,\ldots,\phi_n]$ and $\widetilde{X}_{\mathcal{S}} = [x_{1,\mathcal{S}}^T,\ldots,x_{n,\mathcal{S}}^T]^T$. First note

$$\mathrm{E}^*\left[\phi\phi^T \mid x_1,\ldots,x_n\right] = \tau(1-\tau)I_n, \quad \mathrm{E}^*\left[\widetilde{X}_{\mathcal{S}}^T\widetilde{X}_{\mathcal{S}}\right] = nD_{\mathcal{S}}.$$

By assumption 3, we have $\Delta_s^T G_{\mathcal{S}}\Delta_s \lesssim \Delta_s^T G_{\mathcal{S}} D_{\mathcal{S}}^{-1} G_{\mathcal{S}}\Delta_s \leq \|D^{-1/2}\sum_{i=1}^n \phi_i x_{i,\mathcal{S}}\|^2$. By Chebyshev's inequality and switching the expectation with trace, we have

$$\begin{aligned}\mathrm{pr}^*\left(\Delta_s^T G_{\mathcal{S}}\Delta_s \geq M\frac{s}{n}\right) &\leq \frac{\mathrm{tr}\left(D_{\mathcal{S}}^{-1}\mathrm{E}^*[\widetilde{X}_{\mathcal{S}}^T\phi\phi^T\widetilde{X}_{\mathcal{S}}]\right)}{M\cdot sn}\\ &\lesssim \frac{n\tau(1-\tau)\cdot s}{M\cdot sn}\\ &\lesssim \frac{1}{M},\end{aligned}$$

where the second equality holds by conditioning on $X_{\mathcal{S}}$ first. The proof is now complete. □

S8 Proof for some auxiliary lemmas

S8.1 Proof of Lemma 10 and 11 with the AL prior

In this subsection, we prove some auxiliary lemmas involved in Section S4. In particular, we consider the AL prior and operate in the asymptotic regime where the dimension p is fixed. We shall continue to use the notations in Section S3.1 and S4. In particular, recall $p_n(\boldsymbol{\delta})$ and $f_n(\boldsymbol{\delta})$ from (S4.3). For simplicity, we shall write $\mathcal{B}_n = \mathcal{B}_n(K_n)$ for some sequence K_n as in (S4.4), and we define

$$A = \begin{bmatrix} (n^{1/2})\boldsymbol{I}_s & 0 \\ 0 & (n\lambda)\boldsymbol{I}_{p-s} \end{bmatrix}.$$

Proof of Lemma 10. Recalling θ_M is the maximum eigenvalue of G, let

$$\widetilde{P}_n = \left\{ \prod_{j \notin \mathcal{S}} \left(\frac{2}{\tilde{w}_j} \right) \right\} \left(\frac{2\pi}{n\theta_M} \right)^{s/2} \times \exp\left\{ -\tilde{\boldsymbol{w}}_{\mathcal{S}}^T \boldsymbol{\beta}_{\mathcal{S}}^0 + \frac{n}{2} \Delta_s^T G_{\mathcal{S}} \Delta_s \right\}.$$

In step I, we first show that $p_n(\boldsymbol{\delta}) \gtrsim_{\mathrm{pr}^*} \underline{p}_n(\boldsymbol{\delta})$ on the are $\mathcal{B}_n$, for a sequence K_n such that Corollary 1 holds; and in step II we lower bound the integration of $\underline{p}_n(\boldsymbol{\delta})$ on $\mathcal{B}_n$ by $\widetilde{P}_n$ to conclude the proof.

Step I: Lower bounding the posterior density $p_n(\boldsymbol{\delta})$

We first analyze the adaptive Lasso prior when $\boldsymbol{\delta} \in \mathcal{B}_n$. Note that

$sgn(\boldsymbol{\delta}_{\mathcal{S}} + \beta^0_{\mathcal{S}}) = sgn(\beta^0_{\mathcal{S}})$ when $\boldsymbol{\delta} \in \mathcal{B}_n$; the adaptive Lasso prior then becomes

$$\pi(\boldsymbol{\delta} + \boldsymbol{\beta}^0) = \exp\left\{-w_{\mathcal{S}}^T(\boldsymbol{\delta}_{\mathcal{S}} + \boldsymbol{\beta}^0_{\mathcal{S}}) - \sum_{j=s+1}^{p} \tilde{w}_j|\delta_j|\right\}, \quad \boldsymbol{\delta} \in \mathcal{B}_n.$$

Next we consider the working likelihood on $\boldsymbol{\delta} \in \mathcal{B}_n$. Corollary 1 gives

$$\begin{aligned} L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta} + \boldsymbol{\beta}^0) &= \sum_{i=1}^{n} \phi_i \boldsymbol{x}_i^T \boldsymbol{\delta} - \frac{n}{2}\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S}} \boldsymbol{\delta}_{\mathcal{S}} + o_{\mathrm{pr}^*}(1) \\ &\geq -\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}} - \Delta_s)^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \Delta_s) + \frac{n}{2}\Delta_s^T G_{\mathcal{S}} \Delta_s \\ &\quad - \left\| \sum_{i=1}^{n} \phi_i x_{i,\mathcal{S}^c} \right\|_\infty \cdot \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 + o_{\mathrm{pr}^*}(1), \quad \text{(S8.37)} \end{aligned}$$

uniformly on $\boldsymbol{\delta} \in \mathcal{B}_n$. Therefore, we have the following lower bound of the posterior density:

$$\begin{aligned} p_n(\boldsymbol{\delta}) \gtrsim_{\mathrm{pr}^*} & \exp\left\{\frac{n}{2}\Delta_s^T G_{\mathcal{S}} \Delta_s - \tilde{\boldsymbol{w}}_{\mathcal{S}}^T \boldsymbol{\beta}^0_{\mathcal{S}}\right\} \cdot \exp\left\{-\frac{n\theta_M}{2}\|\boldsymbol{\delta}_{\mathcal{S}} - \Delta_s\|^2 - \tilde{\boldsymbol{w}}_{\mathcal{S}}^T \boldsymbol{\delta}_{\mathcal{S}}\right\} \\ & \cdot \exp\left\{-\sum_{j=s+1}^{p} (\alpha_n + \tilde{w}_j)|\delta_j|\right\} \\ \triangleq & \exp\left\{\frac{n}{2}\Delta_s^T G_{\mathcal{S}} \Delta_s - \tilde{\boldsymbol{w}}_{\mathcal{S}}^T \boldsymbol{\beta}^0_{\mathcal{S}}\right\} \cdot \underline{p}_n(\boldsymbol{\delta}_{\mathcal{S}}, \boldsymbol{\delta}_{\mathcal{S}^c}), \end{aligned}$$

where $\theta_M = \theta_{\max}(G)$, $\alpha_n = \|\sum_{i=1}^n \phi_i x_{i,\mathcal{S}^c}\|_\infty$ and $\Delta_s = G_{\mathcal{S}}^{-1} \sum_{i=1}^n \phi_i x_{i,\mathcal{S}}$.

Step II: Lower bounding the integral

Now we relate the integral of $\underline{p}_n(\boldsymbol{\delta})$ to probabilistic calculations. Let

$Z \in \mathbb{R}^s$ and $\xi = (\xi_{s+1}, \ldots, \xi_p)$ be distributed as

$$\begin{aligned} \xi_j &\overset{ind.}{\sim} \text{Laplace}\left(\frac{1}{\tilde{w}_j + \alpha_n}\right), \quad j = s+1, \ldots, p, \\ Z &\sim \text{N}\left(\Delta_s, \frac{1}{n\theta_M}\boldsymbol{I}_s\right), \end{aligned}$$

where Z and ξ are independent. Similar to (S4.8) in Section S4, we have

$$\begin{aligned} \frac{\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\int_{\mathbb{R}^p} \underline{p}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} &\geq \frac{\int_{\mathcal{B}_n} \underline{p}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\int_{\mathbb{R}^p} \underline{p}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \\ &\gtrsim_{\mathrm{pr}^*} \mathrm{E}\left(\exp\left\{-\tilde{\boldsymbol{w}}_{\mathcal{S}}^T Z\right\} \cdot \mathbf{1}[\|Z\| \leq K_n/\sqrt{n}]\right) \cdot \mathrm{pr}\left(\|\xi\|_\infty \leq \frac{K_n}{n\lambda}\right) \\ &\gtrsim_{\mathrm{pr}^*} 1, \end{aligned}$$

the last inequality is similar to (S4.9) and (S4.8)

Finally, note the normalizing constant for $\underline{p}_n$ can be explicitly derived from Gaussian and Laplace distributions:

$$\begin{aligned} \int_{\mathbb{R}^p} \underline{p}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} &= \exp\left\{\frac{n}{2}\Delta_s^T G_{\mathcal{S}} \Delta_s - \tilde{\boldsymbol{w}}_{\mathcal{S}}^T \boldsymbol{\beta}_{\mathcal{S}}^0\right\} \cdot \left(\frac{2\pi}{n\theta_M}\right)^{s/2} \cdot \left\{\prod_{j=s+1}^{p} \left(\frac{2}{\tilde{w}_j + \alpha_n}\right)\right\} \\ &\gtrsim \widetilde{P}_n \times \left(\frac{1}{1 + \alpha_n/w_{\min}}\right)^{p-s} \\ &\gtrsim_{\mathrm{pr}^*} \widetilde{P}_n, \end{aligned}$$

where $w_{\min} = \min\{\tilde{w}_j : j \notin \mathcal{S}\}$; the last inequality follows since $(1+x)^{-1} \geq \exp\{-x\}$ and $p\alpha_n \leq_{\mathrm{pr}^*} \gamma \cdot w_{\min}$ as in the proof for part 1 of Theorem 1 (Step V). Therefore, the proof is now complete. □

Proof of Lemma 11. Let $\widetilde{\boldsymbol{\delta}}_{\mathcal{S}} = \widetilde{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}^0_{\mathcal{S}}$, where $\widetilde{\boldsymbol{\beta}}_{\mathcal{S}}$ is the oracle estimator. Let $Z \in \mathbb{R}^s$ and $\xi = (\xi_1, \ldots, \xi_{p-s})$ be distributed as

$$\begin{aligned} \xi_1, \ldots, \xi_{p-s} &\overset{i.i.d.}{\sim} \text{Laplace}\left(\frac{1}{\tilde{w}_j}\right), \\ Z &\sim \text{N}\left(\widetilde{\boldsymbol{\delta}}_{\mathcal{S}},\ \frac{1}{n} G_{\mathcal{S}}^{-1}\right), \end{aligned}$$

and ξ is independent of Z. Similar to (S4.8), the integration of $\|A\boldsymbol{\delta}\| f_n(\boldsymbol{\delta})$ can be transformed to Gaussian and Laplace moments:

$$\begin{aligned} \frac{\displaystyle\int_{\mathcal{B}_n^c} \|A\boldsymbol{\delta}\|^\alpha f_n(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta}}{\displaystyle\int_{\mathbb{R}^p} f_n(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta}} &\leq \left[\mathrm{E}\left\{\left(\|\sqrt{n}Z\|^{2\alpha} + \|n\lambda\xi\|^{2\alpha}\right)\right\}\right]^{1/2} \\ &\quad \cdot \left[\mathrm{pr}\left(\|Z\| \geq \frac{M_n}{\sqrt{n}} \text{ or } \|\xi\|_\infty \geq \frac{M_n}{n\lambda}\right)\right]^{1/2}, \\ &= o_{\mathrm{pr}^*}(1), \end{aligned}$$

where we bound the two terms similar to how we bound (S4.8) and (S4.9), since $\widetilde{\boldsymbol{\delta}}_{\mathcal{S}} = O_{\mathrm{pr}^*}(n^{-1/2})$. The proof is now complete. □

S8.2 Proof of Lemma 15, 16 and 17 with the CA prior

In this subsection, we prove some auxiliary lemmas involved in Section S6. In particular, we consider the CA prior and operate in the asymptotic regime where the dimension p is increasing with the sample size. We shall continue to use the notations in Section S3.1 and S6. In particular, recall $p_n(\boldsymbol{\delta})$, $h_n(\boldsymbol{\delta})$ and $f_n(\boldsymbol{\delta})$ from (S6.19). For simplicity, we shall write $\mathcal{B}_n =$

$\mathcal{B}_n(K_n)$ for some sequence K_n as in (S6.18), and we define

$$A = \begin{bmatrix} (n^{1/2})\boldsymbol{I}_s & 0 \\ 0 & (n\lambda)\boldsymbol{I}_{p-s} \end{bmatrix}.$$

Proof of Lemma 15. We provide a lower bound of the integral by restricting to the area

$$\mathcal{E}_n = \mathcal{E}_n(K_n) = \{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}_{\mathcal{S}}\|_2 \leq K_n\sqrt{s/n};\ \ \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \leq K_n \log p/(n\lambda)\},$$

where the sequence $K_n \to \infty$ satisfies the requirement in Corollary 1. We define

$$\widetilde{P}_n = \left(\frac{2\pi}{n}\right)^{s/2}\left(\frac{2}{n\lambda}\right)^{p-s} \cdot \frac{\exp\left(-sn\lambda^2 + n\Delta_s^T G_{\mathcal{S}}\Delta_s/2\right)}{\sqrt{|G_{\mathcal{S}}|}}.$$

In step I below, we provide a lower bound for the posterior density $p_n(\boldsymbol{\delta})$ on the area $\boldsymbol{\delta} \in \mathcal{E}_n$, which is denoted by $\underline{p}_n$; and in step II we show the integration of $\underline{p}_n$ is lower bounded by $\widetilde{P}_n$.

Step I: Lower bounding the posterior density

We bound the likelihood and the prior separately. First, similar to (S8.37), Corollary 1 gives a lower bound for the working likelihood as

$$\begin{aligned} L_n(\boldsymbol{\beta}^0) - L_n(\boldsymbol{\delta} + \boldsymbol{\beta}^0) \ \ \geq \ \ & -\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}} - \Delta_s)^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \Delta_s) \\ & + \frac{n}{2}\Delta_s^T G_{\mathcal{S}}\Delta_s - \left\|\sum_{i=1}^n \phi_i x_{i,\mathcal{S}^c}\right\|_\infty \cdot \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1 + o_{\mathrm{pr}^*}(1), \end{aligned}$$

uniformly on $\boldsymbol{\delta} \in \mathcal{E}_n$. Second, as $\mathcal{E}_n \subset \mathcal{A}_n$ in Lemma 16, we have from

(S6.20),

$$\pi_{CA}(\boldsymbol{\delta}+\boldsymbol{\beta}^0) = \exp\{-sn\lambda^2 - n\lambda\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\}, \quad \boldsymbol{\delta}\in\mathcal{E}_n.$$

Combining the above displayed equations, the posterior density on $\mathcal{E}_n$ is bounded from below by

$$\begin{aligned} p_n(\boldsymbol{\delta}) \;\gtrsim_{\mathrm{pr}^*}\; & \exp\left\{-sn\lambda^2 + \frac{n}{2}\Delta_s^T G_{\mathcal{S}}\Delta_s\right\}\cdot\exp\left\{-\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}}-\Delta_s)^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}}-\Delta_s)\right\} \\ & \cdot\exp\left\{-\left(n\lambda+\alpha_n\right)\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\} \\ \triangleq\; & \exp\left\{-sn\lambda^2 + \frac{n}{2}\Delta_s^T G_{\mathcal{S}}\Delta_s\right\}\cdot \underline{p}_n(\boldsymbol{\delta}), \end{aligned}$$

where $\alpha_n = \|\sum_{i=1}^n \phi_i x_{i,\mathcal{S}^c}\|_\infty$.

Step II: Bounding the posterior integration

Now, we relate the integration of $\underline{p}_n(\boldsymbol{\delta})$ to probabilistic calculations. Let $Z\in\mathbb{R}^s$ and $\xi=(\xi_1,\ldots,\xi_{p-s})$ be distributed as

$$\begin{aligned} \xi_1,\ldots,\xi_{p-s} \;&\overset{i.i.d.}{\sim}\; \text{Laplace}\left(\frac{1}{n\lambda+\alpha_n}\right), \\ Z \;&\sim\; \mathrm{N}\left(\Delta_s,\, \frac{1}{n}G_{\mathcal{S}}^{-1}\right), \end{aligned}$$

and Z is independent of ξ. Similar to the arguments in (S6.24), we have

$$\begin{aligned} \int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \;\gtrsim_{\mathrm{pr}^*}\; & \exp\left\{-sn\lambda^2 + \frac{n}{2}\Delta_s^T G_{\mathcal{S}}\Delta_s\right\}\cdot\left(\frac{2}{n\lambda+\alpha_n}\right)^{p-s}\cdot\frac{1}{\sqrt{|G_{\mathcal{S}}|}}\cdot\left(\frac{2\pi}{n}\right)^{s/2} \\ & \cdot\,\mathrm{pr}\left(\|Z\|_2\le K_n\sqrt{\frac{s}{n}}\,,\; \|\xi\|_\infty\le\frac{K_n\log p}{n\lambda}\right) \\ \gtrsim\; & \widetilde{P}_n, \end{aligned}$$

which holds on the events

$$N_1(\gamma) = \{p\alpha_n \le \gamma \cdot n\lambda\}, \quad N_2(\gamma) = \{n\Delta_s^T G_{\mathcal{S}} \Delta_s \le \gamma \cdot K_n^2 s\}.$$

for small enough $\gamma > 0$, as in (S6.25).

In the proof of Theorem 3 (Step V), we show the events $N_1(\gamma)$ and $N_2(\gamma)$ have pr*-probability going to 1. Therefore, the proof is now complete.

□

Proof of Lemma 16. Recall $p_n(\boldsymbol{\delta})$ is the posterior density under the CA prior (2.5) in the paper, and

$$\mathcal{A}_n = \left\{ \boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0 : \min_{j=1,\ldots,s} |\beta_j| < \lambda, \text{ or } \max_{j=s+1,\ldots,p} |\beta_j| > \lambda \right\},$$

$$\mathcal{C}_n = \{\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0 : \|G^{1/2}\boldsymbol{\delta}\| \le 4q_0\}.$$

First we provide a decomposition of $\mathcal{A}_n$. Let $Q(\boldsymbol{\delta}) = \{j = 1, \ldots, s : |\delta_j + \beta_j^0| < \lambda\}$ and $R(\boldsymbol{\delta}) = \{j = s+1, \ldots, p : |\delta_j| > \lambda\}$ be two index sets; then $\mathcal{A}_n$ can be decomposed into

$$\begin{aligned} \mathcal{A}_n &\subseteq \bigcup_{\substack{0 \le q \le s;\ 0 \le r \le p-s \\ q+r>0}} \{\boldsymbol{\delta} : |Q(\boldsymbol{\delta})| = q, \ \ |R(\boldsymbol{\delta})| = r\} \\ &\triangleq \bigcup_{\substack{0 \le q \le s;\ 0 \le r \le p-s \\ q+r>0}} \mathcal{E}_{q,r}. \end{aligned} \tag{S8.38}$$

Therefore, it suffices to show the posterior probabilities of $\mathcal{E}_{q,r}$ adds up to be $o_{\text{pr}^*}(1)$.

In the following steps I - II, we upper bound the posterior integral on $\mathcal{A}_n \cap \mathcal{C}_n$ by the decomposition (S8.38). In step III, we compute the posterior integral of $p_n(\boldsymbol{\delta})$ on $\mathcal{C}_n^c$; in step IV we verify the posterior probabilities are $o_{\mathrm{pr}^*}(1)$.

Step I: Bounding the posterior integral on $\mathcal{E}_{q,r} \cap \mathcal{C}_n$.

We first give upper bounds for $p_n(\boldsymbol{\delta})$ on each of $\mathcal{E}_{q,r} \cap \mathcal{C}_n$. For the CA prior, we have

$$\pi_{CA}(\boldsymbol{\delta} + \boldsymbol{\beta}^0) \leq \exp\{-(s-q+r)n\lambda^2\}, \quad \boldsymbol{\delta} \in \mathcal{E}_{q,r}.$$

Using (S4.7) to upper bound the quantile loss function $-L_n(\boldsymbol{\delta} + \boldsymbol{\beta}^0)$, we have

$$\begin{aligned} p_n(\boldsymbol{\delta}) \quad &\lesssim_{\mathrm{pr}^*} \quad \exp\left\{2n\Delta_p^T G \Delta_p - \frac{n}{8}(\boldsymbol{\delta} - 4\Delta_p)^T G(\boldsymbol{\delta} - 4\Delta_p) - (s-q+r)n\lambda^2\right\} \\ &\triangleq \quad \exp\left\{2n\Delta_p^T G \Delta_p - (s-q+r)n\lambda^2\right\} \cdot \overline{p}_n(\boldsymbol{\delta}), \end{aligned}$$

uniformly on $\boldsymbol{\delta} \in \mathcal{E}_{q,r} \cap \mathcal{C}_n$, where $\Delta_p = G^{-1}\sum_{i=1}^n \boldsymbol{x}_i \phi_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}^0)$.

Next, note when $\boldsymbol{\delta} \in \mathcal{E}_{q,r}$, we have

$$\|\boldsymbol{\delta}_{\mathcal{S}}\|^2 \geq \sum_{j \in Q(\boldsymbol{\delta})} (|\beta_j^0| - \lambda)^2 \geq \frac{q \cdot \underline{\mathrm{b}}^2}{4} \geq \frac{32q \cdot \lambda^2}{\theta_{\min}(G)},$$

$$\|\boldsymbol{\delta}_{\mathcal{S}^c}\|^2 \geq \sum_{j \in R(\boldsymbol{\delta})} (\lambda)^2 \geq r \cdot \lambda^2,$$

since $\lambda \ll \underline{b} \cdot (1 \wedge \theta_{\min}(G))$ as stated in the lemma; hence, we have $\|\boldsymbol{\delta}\|^2 \geq \lambda^2(r + 32q/\theta_{\min}(G))$.

Finally, we use Gaussian tail bounds to bound the integral of $p_n(\boldsymbol{\delta})$. Let $\widetilde{Z} \in \mathbb{R}^p$ follow

$$\widetilde{Z} \sim \mathrm{N}\left(4\Delta_p, \frac{4}{n}G^{-1}\right);$$

as in (S6.24), we have for $r + q > 0$:

$$\begin{aligned}
\int_{\mathcal{E}_{q,r}\cap\mathcal{C}_n} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \quad \lesssim_{\mathrm{pr}^*} \quad & \exp\{2n\Delta_p^T G\Delta_p - sn\lambda^2\} \cdot \left(\frac{8\pi}{n}\right)^{p/2} \cdot \frac{1}{\sqrt{|G|}} \\
& \cdot \exp\{(q-r)n\lambda^2\} \cdot \mathrm{pr}\left(\left\|\widetilde{Z}\right\|^2 \geq \lambda^2 \cdot \left[r + \frac{32q}{\theta_{\min}(G)}\right]\right) \\
\leq \quad & \exp\{2n\Delta_p^T G\Delta_p - sn\lambda^2\} \cdot \left(\frac{8\pi}{n}\right)^{p/2} \cdot \frac{1}{\sqrt{|G|}} \\
& \cdot \exp\left\{-(r+q)n\lambda^2\right\}, \qquad \text{(S8.39)}
\end{aligned}$$

where the last inequality holds on the events

$$\begin{aligned}
N_1(\gamma) &= \left\{n \cdot \Delta_p^T G\Delta_p \leq \gamma \cdot n\lambda^2 \cdot (\theta_{\min}(G) \wedge 1)\right\}, \\
N_2 &= \left\{n\lambda^2(\theta_{\min}(G) \wedge 1) \gg p\right\}
\end{aligned}$$

for small enough $\gamma > 0$ by Lemma 1.

Step II: Bounding the posterior integral on $\mathcal{A}_n \cap \mathcal{C}_n$ Motivated the decomposition (S8.38), the posterior integral on $\mathcal{A}_n \cap \mathcal{C}_n$ is bounded by

the following:

$$\begin{aligned}
\int_{\mathcal{A}_n\cap\mathcal{C}_n} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \quad &\leq \quad \sum_{q=0}^{s}\sum_{\substack{r=0\\ r+q>0}}^{p-s}\int_{\mathcal{E}_{q,r}\cap\mathcal{C}_n} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \\
&\lesssim_{\mathrm{pr}^*} \quad \exp\{2n\Delta_p^T G\Delta_p - sn\lambda^2\}\cdot\left(\frac{8\pi}{n}\right)^{p/2}\cdot\frac{1}{\sqrt{|G|}} \\
&\qquad \cdot\left[\sum_{q=0}^{s}\sum_{r=0}^{p-s}\exp\left\{-(r+q)n\lambda^2\right\}-1\right] \\
&= \quad \exp\{2n\Delta_p^T G\Delta_p - sn\lambda^2\}\cdot\left(\frac{8\pi}{n}\right)^{p/2}\cdot\frac{1}{\sqrt{|G|}} \\
&\qquad \cdot\left[\left(\frac{1}{1-\exp\{-n\lambda^2\}}\right)^2-1\right], \qquad \text{(S8.40)}
\end{aligned}$$

where the second inequality holds on the event $N_1(\gamma)$ and N_2, given by step I; and the last inequality uses the property of geometric series.

Since $\theta_{\min}(G)\gtrsim 1/p$ and $\lambda\gg p/\sqrt{n}$, the deterministic event N_2 always holds; furthermore, Lemma 14 implies that $N_1(\gamma)$ has pr*-probability tending to 1. Thus, the previous displayed equation holds with pr*-probability tending to 1.

Step III: Bounding the posterior integral on $\mathcal{C}_n^c$. On $\boldsymbol{\delta}\in\mathcal{C}_n^c$, note that $\pi_{CA}(\boldsymbol{\delta}+\boldsymbol{\beta}^0)\leq 1$. We use the same argument in (S6.28), which gives

$$\int_{\mathcal{C}_n^c} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \quad \lesssim_{\mathrm{pr}^*} \quad \frac{1}{\sqrt{|G|}}\cdot\left(\frac{4q_0}{\sqrt{p}}\right)^p\exp\{p-n\varepsilon_0\}.$$

Step IV: Final bound for posterior probability Finally we verify the posterior probability for $\mathcal{A}_n\cap\mathcal{C}_n$ and $\mathcal{C}_n^c$ are both $o_{\mathrm{pr}^*}(1)$.

First, the normalizing constant for p_n can be bounded by Lemma 15

$$\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \quad \gtrsim_{\mathrm{pr}^*} \quad \left(\frac{1}{n}\right)^p \cdot \exp\left(-sn\lambda^2\right) \cdot \frac{1}{\sqrt{|G_{\mathcal{S}}|}},$$

since $\lambda \ll 1$.

Following the proof of Theorem 3 (step IV), we now consider the posterior probability. On $\mathcal{A}_n \cap \mathcal{C}_n$, we use the bound in step II:

$$\begin{aligned}
\frac{\displaystyle\int_{\mathcal{A}_n \cap \mathcal{C}_n} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \quad &\lesssim_{\mathrm{pr}^*} \quad \exp\left\{c_1 \cdot p\log n + c_2 p\log p + 2n\Delta_p^T G \Delta_p\right\} \cdot 8\exp\left\{-n\lambda^2\right\} \\
&\lesssim_{\mathrm{pr}^*} \quad \exp\left\{-n\lambda^2/4\right\},
\end{aligned}$$

where c_1, c_2 are some positive constants; the first inequality follows from (S6.29) and that $(1-x)^{-2} - 1 \leq 8x$ for all $0 < x < 1/2$; and the last inequality follows from Lemma 14, since $p\log(p \vee n) \ll n\lambda^2$.

On $\mathcal{C}_n^c$, from (S6.29) and the bound in step III we have:

$$\begin{aligned}
\frac{\displaystyle\int_{\mathcal{C}_n^c} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{P}_n} \quad &\lesssim_{\mathrm{pr}^*} \quad \exp\left\{c_3 \cdot p\log n + c_4 \cdot p\log p + sn\lambda^2 - n\varepsilon_0\right\} \\
&\leq \quad \exp\left\{-n\varepsilon_0/4\right\},
\end{aligned}$$

for some constant c_3, $c_4 > 0$; since $p\log(p \vee n) \ll n$ and $\lambda^2 \ll 1/s$. The proof is now complete. □

Proof of Lemma 17. We introduce some notations first. Recall Δ_p, Δ_s from

Section S3, and $h_n(\boldsymbol{\delta})$ and $f_n(\boldsymbol{\delta})$ from (S6.19); we further define

$$\widetilde{f}_n(\boldsymbol{\delta}) = T_n \cdot \exp\left\{-\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}} - \Delta_s)^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \Delta_s) - sn\lambda^2 - n\lambda\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\},$$

where T_n are defined before (S6.19). Let $\widetilde{F}_n = \int \widetilde{f}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}$. Fix a diverging sequence $K_n \to +\infty$ to be specified later, we define

$$\mathcal{E}_n = \left\{\boldsymbol{\delta} \in \mathbb{R}^p : \|\boldsymbol{\delta}_{\mathcal{S}}\|_2 \le K_n\sqrt{\frac{s}{n}}\ ,\ \text{and}\ \ \|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \le K_n\frac{\log p}{n\lambda}\right\}.$$

Similar to the proof of Theorem 3, we first show that

$$\frac{\displaystyle\int_{\mathbb{R}^p}\left|h_n(\boldsymbol{\delta}) - \widetilde{f}_n(\boldsymbol{\delta})\right|\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{F}_n} = o_{\mathrm{pr}^*}(1),$$

which implies h_n converges to $\widetilde{f}_n$ in total variation; to achieve this, in the following steps I - II we bound the integral of $|\widetilde{f}_n - h_n|$ on $\mathcal{B}_n$ and its complement separately; finally in step III, we show that $\widetilde{f}_n$ converges to f_n in total variation, which concludes the proof.

First, the normalizing constant of $\widetilde{f}_n(\boldsymbol{\delta})$, i.e., $\widetilde{F}_n$, can be explicitly computed; using the normalizing constant of Gaussian and Laplace distributions:

$$\widetilde{F}_n \;=\; T_n \cdot \exp\{-sn\lambda^2\} \cdot \frac{1}{\sqrt{|G_{\mathcal{S}}|}} \cdot \left(\frac{2\pi}{n}\right)^{s/2} \cdot \left(\frac{2}{n\lambda}\right)^{p-s}.$$

Step I: Bounding $\int |\widetilde{f}_n - h_n|\,\mathrm{d}\boldsymbol{\delta}$ on $\mathcal{E}_n$

Similar to the proof for part 2 of Theorem 1 in Section S4, we have the

bound

$$\begin{aligned}\sup_{\boldsymbol{\delta}\in\mathcal{E}_n}\left|\log\left(\frac{h_n(\boldsymbol{\delta})}{\widetilde{f}_n(\boldsymbol{\delta})}\right)\right| &\leq \frac{n}{2}\sup_{\boldsymbol{\delta}\in\mathcal{E}_n}\left|\boldsymbol{\delta}^T G\boldsymbol{\delta}-\boldsymbol{\delta}_{\mathcal{S}}^T G_{\mathcal{S}}\boldsymbol{\delta}_{\mathcal{S}}\right| + \left\|\sum_{i=1}^n \phi_i \boldsymbol{x}_i\right\|_\infty \sup_{\boldsymbol{\delta}\in\mathcal{E}_n}\|\boldsymbol{\delta}_{\mathcal{S}}^c\|_1 \\ &= o_{\mathrm{pr}^*}(1),\end{aligned}$$

provided that $\lambda \gg (p\log^{3/2} p)/\sqrt{n}$ and K_n increases with n slow enough; we bound the first term above by Corollary 1, and the second term by Lemma 13. Therefore, as in (S6.30),

$$\int_{\mathcal{E}_n}\left|h_n(\boldsymbol{\delta})-\widetilde{f}_n(\boldsymbol{\delta})\right|\,\mathrm{d}\boldsymbol{\delta} \;=\; o_{\mathrm{pr}^*}\left(\widetilde{F}_n\right).$$

Step II: Bounding $\int|\widetilde{f}_n - h_n|\,\mathrm{d}\boldsymbol{\delta}$ on $\mathcal{E}_n^c$

Here we upper bound in integral of $h_n(\boldsymbol{\delta})$, and $\widetilde{f}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}$ separately.

For $h_n(\boldsymbol{\delta})$, similar to (S6.31), we have

$$\frac{\displaystyle\int_{\mathcal{E}_n^c} h_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{F}_n} \;\lesssim_{\mathrm{pr}^*}\; \exp\left\{-c_0 K_n\right\},$$

for some constant $c_0 > 0$.

For $\widetilde{f}_n$, let $Z \in \mathbb{R}^s$ and $\xi = (\xi_1, \ldots, \xi_{p-s})$ be distributed as

$$\begin{aligned}\xi_1,\ldots,\xi_{p-s} &\overset{i.i.d.}{\sim} \text{Laplace}\left(\frac{1}{n\lambda}\right),\\ Z &\sim \mathrm{N}\left(\Delta_s,\ \frac{1}{n}G_{\mathcal{S}}^{-1}\right),\end{aligned}$$

and Z is independent of ξ. Then on the event $N_1 = \{8n\Delta_s^T G_{\mathcal{S}}\Delta_s \leq$

$\theta_{11}K_n^2 s\}$.

$$\begin{aligned}\frac{\int_{\mathcal{E}_n^c} \widetilde{f}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\widetilde{F}_n} &= \mathrm{pr}\left(\|Z\|_2 \geq K_n\sqrt{\frac{s}{n}} \text{ or } \|\xi\|_\infty \geq \frac{K_n \log p}{n\lambda}\right) \\ &\lesssim \left(p\cdot\exp\{-K_n\log p\} + \exp\left\{-\frac{\theta_{11}K_n^2 s}{4}\right\}\right),\end{aligned}$$

by Lemma 1 and 2, similar to (S6.24); θ_{11} is the minimal eigenvalue of $G_{\mathcal{S}}$.

In the proof of Theorem 3 (step V), we show the event N_1 holds with pr*-probability tending to 1, as $n \to \infty$. Therefore, combining steps I and II, we conclude

$$\left\|\frac{\widetilde{f}_n(\boldsymbol{\beta})}{\int \widetilde{f}_n(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta}} - \frac{h_n(\boldsymbol{\beta})}{\int h_n(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta}}\right\|_{TV} = o_{\mathrm{pr}^*}(1).$$

by Theorem 1 in Chernozhukov and Hong (2003).

Step III: Convergence of $\widetilde{f}_n$ to f_n

Here we show that $\widetilde{f}_n$ converges to f_n in total variation by bounding their KL divergence.

First note that $\int_{\mathbb{R}^p} f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} = \int_{\mathbb{R}^p} \widetilde{f}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}$. Furthermore,

$$\log\left(\frac{\widetilde{f}_n(\boldsymbol{\delta})}{f_n(\boldsymbol{\delta})}\right) = \frac{n}{2}(\Delta_s - \widetilde{\boldsymbol{\delta}}_{\mathcal{S}})^T G_{\mathcal{S}}(\Delta_s - \widetilde{\boldsymbol{\delta}}_{\mathcal{S}}) - n(\Delta_s - \widetilde{\boldsymbol{\delta}}_{\mathcal{S}})^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \Delta_s).$$

Recall that $\widetilde{f}_n(\boldsymbol{\delta}_{\mathcal{S}}, \boldsymbol{\delta}_{\mathcal{S}^c})/\widetilde{F}_n$ coincides with the probability density of (Z, ξ), as defined in step II. Now, by Pinsker's inequality (Tsybakov, 2008, Lemma 2.5), we can bound the total variation distance by their KL diver-

gence,

$$\begin{aligned}
\left\| \frac{\widetilde{f}_n(\boldsymbol{\delta})}{\int \widetilde{f}_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} - \frac{f_n(\boldsymbol{\delta})}{\int f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \right\|_{TV} &\lesssim \left(\mathrm{E}_{(Z,\xi)} \left[\log \frac{f_n(Z,\xi)}{\widetilde{f}_n(Z,\xi)} \right] \right)^{1/2} \\
&\lesssim \sqrt{n} \|\Delta_s - \widetilde{\boldsymbol{\delta}}_1\| \\
&= o_{\mathrm{pr}^*}(1),
\end{aligned}$$

where the second equality holds since $\mathrm{E}_Z(Z - \Delta_s) = 0$; and the last inequality follows from the Bahadur representation of the quantile regression estimators (He and Shao, 2000). Thus, the proof is now complete. □

S8.3 Proof of Lemma 12

In this subsection, we prove Lemma 12 in Section S5; in particular, we show separately that the lemma holds under either the AL or the CA prior. In line with the setting in Section S5, we assume that the covariate dimension p is fixed throughout this subsection.

For the Adaptive Lasso prior

We shall start with the adaptive lasso prior. To match the notations in Section S4, we define for $\boldsymbol{\delta} = \boldsymbol{\beta} - \boldsymbol{\beta}^0$:

$$f_n(\boldsymbol{\delta}) = T_n \times \exp\left\{ -\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}} - \tilde{\boldsymbol{\delta}}_{\mathcal{S}})^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \tilde{\boldsymbol{\delta}}_{\mathcal{S}}) - \sum_{j \notin \mathcal{S}} \tilde{w}_j |\delta_j| \right\},$$

$$p_n(\boldsymbol{\delta}) = \exp\left\{ -\sum_{j=1}^{p} \tilde{w}_j |\beta_j^0 + \delta_j| \ + \ L_n(\boldsymbol{\beta}^0) - L_n\left(\boldsymbol{\delta} + \boldsymbol{\beta}^0\right) \right\},$$

in accordance with (S4.3). Recall the set $\mathcal{B}_n(M_n)$ from (S4.4) and the matrix A from (S4.5). To prove Lemma 12 under the AL prior, it suffices to show

$$\frac{\displaystyle\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\delta\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0, \quad \text{and} \quad \frac{\displaystyle\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\delta\|^\alpha f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0,$$

in pr*-probability, for any diverging sequence $M_n \to +\infty$ and any $\alpha = 0, 1, 2$. Note the first equation above is implied by Equation (S4.6), which we showed in the proof of Theorem 1; the second equation is precisely the conclusion of Lemma 11. Therefore, Lemma 12 holds under the AL prior.

For the Clipped Absolute prior

Now we prove Lemma 12 under the CA prior. We shall continue to use the notations defined in S3.1, and we switch to use the notations in Section S6 for the CA prior. In particular, the posterior density is (up to a

normalization constant)

$$p_n(\delta) = \pi_{CA}(\delta + \boldsymbol{\beta}^0) \cdot \exp\left\{L_n(\boldsymbol{\beta}^0) - L_n\left(\delta + \boldsymbol{\beta}^0\right)\right\};$$

the limiting posterior density is:

$$f_n(\boldsymbol{\delta}) = T_n \times \exp\left\{-\frac{n}{2}(\boldsymbol{\delta}_{\mathcal{S}} - \widetilde{\boldsymbol{\delta}}_{\mathcal{S}})^T G_{\mathcal{S}}(\boldsymbol{\delta}_{\mathcal{S}} - \widetilde{\boldsymbol{\delta}}_{\mathcal{S}}) - sn\lambda^2 - n\lambda\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_1\right\},$$

where $T_n = \exp\{n\Delta_s^T G_{\mathcal{S}}\Delta_s/2\}$. Recall the matrix A from (S4.5), and we define the set $\mathcal{B}_n(M_n)$ as in (S4.4) when the dimension is fixed.

Proof of Lemma 12 under the CA prior. We first show that

$$\frac{\displaystyle\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\delta\|^\alpha f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} f_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0.$$

Let $Z \in \mathbb{R}^s$ and $\xi = (\xi_1, \ldots, \xi_{p-s})$ be distributed as

$$\begin{aligned} \xi_1, \ldots, \xi_{p-s} &\overset{i.i.d.}{\sim} \text{Laplace}\left(\frac{1}{n\lambda}\right), \\ Z &\sim \mathrm{N}\left(\widetilde{\boldsymbol{\delta}}_{\mathcal{S}},\ \frac{1}{n}G_{\mathcal{S}}^{-1}\right), \end{aligned}$$

and ξ is independent of Z. Similar to the proof of Lemma 11, the integration of $\|A\boldsymbol{\delta}\|f_n(\boldsymbol{\delta})$ can be transformed to Gaussian and Laplace moments:

$$\begin{aligned} \frac{\displaystyle\int_{\mathcal{B}_n^c} \|A\boldsymbol{\delta}\|^\alpha f_n(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta}}{\displaystyle\int_{\mathbb{R}^p} f_n(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta}} &\leq \left[\mathrm{E}\left\{\left(\|\sqrt{n}Z\|^{2\alpha} + \|n\lambda\xi\|^{2\alpha}\right)\right\}\right]^{1/2} \\ &\quad \cdot \left[\mathrm{pr}\left(\|Z\| \geq \frac{M_n}{\sqrt{n}} \text{ or } \|\xi\|_\infty \geq \frac{M_n}{n\lambda}\right)\right]^{1/2}, \\ &= o_{\mathrm{pr}^*}(1), \end{aligned}$$

where we can bound the moment by (S4.8) and the tail probability by (S6.25), respectively. Therefore the conclusion for f_n in Lemma 12 holds.

Next we shall show

$$\frac{\displaystyle\int_{\{\mathcal{B}_n(M_n)\}^c} \|A\delta\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}}{\displaystyle\int_{\mathbb{R}^p} p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}} \to 0.$$

Following the proof of Theorem 3 in Section S6, we define the following regions:

$$\mathcal{A}_n = \left\{\boldsymbol{\delta} \in \mathbb{R}^p : \min_{j\in\mathcal{S}} |\beta_j| < \lambda, \quad \text{or} \quad \max_{j\notin\mathcal{S}} |\beta_j| > \lambda\right\},$$

$$\mathcal{B}_n(M_n) = \{\boldsymbol{\delta} : n^{1/2}\|\boldsymbol{\delta}_{\mathcal{S}}\|_2 \le M_n \ , \ n\lambda\|\boldsymbol{\delta}_{\mathcal{S}^c}\|_\infty \le M_n\},$$

$$\mathcal{C}_n = \left\{\boldsymbol{\delta} : \ \|G^{1/2}\boldsymbol{\delta}\| \le 4q_0\right\}.$$

Note that

$$\{\mathcal{B}_n(M_n)\}^c \subset (\mathcal{A}_n \cap \mathcal{C}_n) \cup (\{\mathcal{B}_n(M_n)\}^c \cap \mathcal{C}_n \cap \mathcal{A}_n^c) \cup (\mathcal{C}_n^c \cap \mathcal{A}_n^c)$$

And we shall show that the posterior integral is $o_{\mathrm{pr}^*}(1)$ in each of the region on the right hand side above.

Step I: Bound on the region $\mathcal{A}_n \cap \mathcal{C}_n$

We only need a slight modification for the proof of Lemma 16 in Section S8.2. Following the same vein as the proof therein, we divide $\mathcal{A}_n$ into the union of $\mathcal{E}_{q,r}$'s, where $0 \le q \le s, 0 \le r \le p-s$ and $q+r > 0$. Note on each $\mathcal{E}_{q,r} \cap \mathcal{C}_n$, we can use Gaussian moment/tail bounds to bound the posterior

integral. Let $\widetilde{Z} \in \mathbb{R}^p$ follow

$$\widetilde{Z} \sim \mathrm{N}\left(4\Delta_p, \frac{4}{n}G^{-1}\right);$$

and we have from (S8.39):

$$\begin{aligned}
\int_{\mathcal{E}_{q,r}\cap\mathcal{C}_n} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \quad &\lesssim_{\mathrm{pr}^*} \quad \exp\{2n\Delta_p^T G\Delta_p - sn\lambda^2\}\cdot\left(\frac{8\pi}{n}\right)^{p/2}\cdot\frac{1}{\sqrt{|G|}} \\
&\qquad \cdot\exp\{(q-r)n\lambda^2\}\left[\mathrm{E}\left(\left\|n\lambda\widetilde{Z}\right\|^{2\alpha}\right)\right]^{1/2}\cdot\left[\mathrm{pr}\left(\left\|\widetilde{Z}\right\|^2 \geq \lambda^2\cdot\left[r+\frac{64q}{\theta_m}\right]\right)\right]^{1/} \\
&\lesssim_{\mathrm{pr}^*} \quad \exp\{2n\Delta_p^T G\Delta_p - sn\lambda^2\}\cdot\left(\frac{8\pi}{n}\right)^{p/2}\cdot\frac{1}{\sqrt{|G|}} \\
&\qquad \cdot\exp\{-(q+r)n\lambda^2\}\cdot(\sqrt{n}\lambda)^\alpha,
\end{aligned}$$

where we bound the moment by Lemma 1 and the tail probability by (S8.39). Therefore, we can obtain from (S8.40)

$$\begin{aligned}
\int_{\mathcal{A}_n\cap\mathcal{C}_n} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \quad &\lesssim_{\mathrm{pr}^*} \quad \exp\{2n\Delta_p^T G\Delta_p - sn\lambda^2\}\cdot\left(\frac{8\pi}{n}\right)^{p/2}\cdot\frac{1}{\sqrt{|G|}} \\
&\qquad \cdot(\sqrt{n}\lambda)^\alpha\exp\{-n\lambda^2\},
\end{aligned}$$

Step II: Bound on the region $\{\mathcal{B}_n(M_n)\}^c \cap \mathcal{C}_n \cap \mathcal{A}_n^c$

In this area, we use the upper bound on the posterior density $p_n(\boldsymbol{\delta})$ in (S6.23), and then relate the integration of p_n to probabilistic calculations. Similar to (S6.24), let $Z \in \mathbb{R}^s$ and $\xi = (\xi_1, \ldots, \xi_{p-s})$ be distributed as

$$\begin{aligned}
\xi_1, \ldots, \xi_{p-s} \quad &\overset{i.i.d.}{\sim} \quad \mathrm{Laplace}\left(\frac{1}{n\lambda - \alpha_n}\right), \\
Z \mid \xi \quad &\sim \quad \mathrm{N}\left(\widetilde{\mu}(\xi),\ \frac{4}{n}G_{\mathcal{S}}^{-1}\right),
\end{aligned}$$

where $\alpha_n = \left\|\sum_{i=1}^n \phi_i \left(x_{i,\mathcal{S}^c} - A_2^T x_{i,\mathcal{S}}\right)\right\|_\infty$ and $\widetilde{\mu}(\boldsymbol{\delta}_{\mathcal{S}^c}) = 4\Delta_s - A_2\boldsymbol{\delta}_{\mathcal{S}^c}$ and $A_2 = G_{\mathcal{S}}^{-1}G_{\mathcal{S},\mathcal{S}^c}$. Therefore,

$$\begin{aligned}\int_{\mathcal{C}_n\cap\mathcal{B}_n^c\cap\mathcal{A}_n^c} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \quad \lesssim_{\mathrm{pr}^*} \quad & \exp\{-sn\lambda^2 + 2n\Delta_s^T G_{\mathcal{S}}\Delta_s\}\cdot\left(\frac{2}{n\lambda-\alpha_n}\right)^{p-s}\cdot\frac{1}{\sqrt{|G_{\mathcal{S}}|}}\cdot\left(\frac{8\pi}{n}\right)^{s/2} \\ & \cdot\left[\mathrm{E}\left\{\left(\|\sqrt{n}Z\|^{2\alpha} + \|n\lambda\xi\|^{2\alpha}\right)\right\}\right]^{1/2} \\ & \cdot\left[\mathrm{pr}\left(\|Z\|_2 \geq M_n\sqrt{\frac{s}{n}} \ \text{ or } \ \|\xi\|_\infty \geq \frac{M_n s\log p}{n\lambda}\right)\right]^{1/2},\end{aligned}$$

where the moment term can be bounded from (S4.8), and the tail probability can be bounded by (S6.25).

Step III: Bound on the region $\mathcal{C}_n^c \cap \mathcal{A}_n^c$

For this area, we first bound the posterior density $p_n(\boldsymbol{\delta})$ with (S6.21). Noting that $\|A\boldsymbol{\delta}\|^\alpha \leq (n\lambda)^\alpha\|\boldsymbol{\delta}\|^\alpha \leq (n\lambda/\theta_m^{1/2})^\alpha\|G^{1/2}\boldsymbol{\delta}\|^\alpha$, then the posterior integral can be bounded similar to (S6.28) and (S4.10)

$$\begin{aligned}\int_{\mathcal{C}_n^c\cap\mathcal{A}_n^c} \|A\boldsymbol{\delta}\|^\alpha p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta} \quad \lesssim_{\mathrm{pr}^*} \quad & \exp\{-sn\lambda^2\}\cdot\left(\frac{n\lambda}{\sqrt{\theta_m}}\right)^\alpha\cdot\frac{1}{\sqrt{|G|}} \\ & \cdot\int_{\|u\|_2\geq 4q_0}\|u\|^\alpha\exp\left\{-\frac{n\varepsilon_0\|u\|_2}{4q_0}\right\}\mathrm{d}u \\ \lesssim \quad & \exp\{-sn\lambda^2\}\cdot\left(\frac{n\lambda}{\sqrt{\theta_m}}\right)^\alpha\cdot\frac{1}{\sqrt{|G|}} \\ & \left[\int_{\mathbb{R}^p}\|u\|^{2\alpha}\exp\left\{-\frac{n\varepsilon_0\|u\|_1}{4q_0\sqrt{p}}\right\}\mathrm{d}u\right]^{1/2} \\ & \cdot\left[\int_{\|u\|\geq 4q_0}\exp\left\{-\frac{n\varepsilon_0\|u\|_1}{4q_0\sqrt{p}}\right\}\mathrm{d}u\right]^{1/2},\end{aligned}$$

which can be bounded by Equation (S6.28) and Lemma 2 by the properties of Laplace distribution.

Finally, comparing the upper bounds in the three steps above with the normalization constant $\int p_n(\boldsymbol{\delta})\,\mathrm{d}\boldsymbol{\delta}$ in Lemma 15, we've shown the conclusion for p_n in Lemma 12 holds as well. The proof is now complete.

□

Bibliography

Alhamzawi, R. (2018). *Brq: An R package for Bayesian Quantile Regression.* R package version 5.38.

Alhamzawi, R., K. Yu, and D. F. Benoit (2012). Bayesian adaptive lasso quantile regression. *Statistical Modelling 12*(3), 279–297.

Andrews, D. W. (1994). Empirical process methods in econometrics. *Handbook of econometrics 4*, 2247–2294.

Belloni, A., V. Chernozhukov, et al. (2011). $\ell 1$-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics 39*(1), 82–130.

Belloni, A., V. Chernozhukov, D. Chetverikov, and I. Fernández-Val (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics 213*(1), 4–29.

Birgé, L. et al. (2001). An alternative point of view on lepski's method. *Lecture Notes-Monograph Series 36*, 113–133.

Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

Chao, S.-K., S. Volgushev, G. Cheng, et al. (2017). Quantile processes for semi and nonparametric regression. *Electronic Journal of Statistics 11*(2), 3272–3331.

Chen, C. W., D. B. Dunson, C. Reed, and K. Yu (2013). Bayesian variable selection in quantile regression. *Statistics and its Interface 6*(2), 261–274.

Chernozhukov, V. and H. Hong (2003). An mcmc approach to classical estimation. *Journal of Econometrics 115*(2), 293–346.

He, X. and Q.-M. Shao (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis 73*(1), 120–135.

Knight, K. (1998). Limiting distributions for l1 regression estimators under general conditions. *Annals of Statistics 26*(2), 755–770.

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. (2018). *quantreg: Quantile Regression*. R package version 5.38.

Kohns, D. and T. Szendrei (2020). Horseshoe prior bayesian quantile regression. *arXiv preprint arXiv:2006.07655*.

Martin, A. D., K. M. Quinn, and J. H. Park (2011). Mcmcpack: Markov chain monte carlo in r.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory 1*(3), 295–313.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*(2), 186–199.

Sherwood, B. and A. Maidman (2022). *rqPen: Penalized Quantile Regression*. R package version 2.3.

Tsybakov, A. B. (2008). *Introduction to nonparametric estimation*. Springer Science & Business Media.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge university press.

Wang, L., I. Van Keilegom, and A. Maidman (2018). Wild residual bootstrap inference for penalized quantile regression with heteroscedastic errors. *Biometrika 105*(4), 859–872.

Wu, Y. and Y. Liu (2009). Variable selection in quantile regression. *Statistica Sinica 19*(2), 801–817.

Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters 54*(4), 437–447.