The Aeronautical Journal (2026), **130**, pp. 2080–2107
doi:10.1017/aer.2026.10166

RESEARCH ARTICLE

Transfer learning in hard landing risk prediction

Li Huang[1], Liming Wang[2] and Peng Lai[1]

[1]NUIST: Nanjing University of Information Science and Technology, China
[2]Xi'an University of Finance and Economics, China
Corresponding author: Peng Lai; Email: laipeng@nuist.edu.cn

Received: 2 December 2025; **Revised:** 24 March 2026; **Accepted:** 24 March 2026

Keywords: composite quantile regression; data fusion; hard landing risk; transfer learning

Abstract

In flight safety monitoring, risk prediction demands high accuracy for rare but critical events such as engine failures and weather-related hazards. However, the limited sample size of such extreme cases within a single dataset poses a significant challenge to achieving precise predictions. Aggregating data across sources is a potential solution, but datasets are often not independently and identically distributed. Transfer learning can improve target performance by leveraging related sources, yet classical methods struggle with heavy-tailed data due to the lack of robust loss functions. This paper proposes a transfer learning algorithm based on composite quantile regression (CQR). It identifies and transfers information from similar sources and constructs pseudo response variables to linearise the CQR loss, enabling robust computation under heavy-tailed noise. Extensive simulation studies demonstrate the superior performance of the proposed method. The practical utility of the algorithm is further illustrated through its application to quick access recorder (QAR) flight data for hard landing risk prediction. This work aims to identify and quantify key controllable factors influencing landing safety through a data-driven approach. Our research provides quantitative analytical tools for multiple core safety aspects, including flight operation standardisation, aircraft condition monitoring, operational environment assessment and support resource allocation, thereby enabling more precise and proactive risk mitigation strategies.

Nomenclature

x	covariates
y	response variable
n	sample size
S	number of quantiles
d	similarity between target and sources
h	bandwidth
K	number of sources
G	composite quantile loss function
g	the subgradient of the loss function
H	overall Hessian matrix

Greek symbol

τ	composite quantile level
ϵ	random error
β	regression coefficient
λ	penalty term
δ	bias term
$\mathcal{A}$	informative set

© The Author(s), 2026. Published by Cambridge University Press on behalf of Royal Aeronautical Society.

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

1.0 Introduction

With the continuous growth of civil aviation passenger traffic, aviation management has become a global concern. A core component of modern aviation management is flight safety monitoring, which relies on analysing large-scale data recorded by the quick access recorder (QAR). This process not only involves real-time recording and anomaly identification of flight status, but also is an important basis for aviation safety management, accident prevention and policy optimisation. QAR data records aircraft attitude parameters and engine-related data at different stages, which provides effective data support for evaluating pilot operation level, improving aircraft performance and optimising flight technology. Currently, researchers are committed to analysing and predicting flight risks by building data-driven models to reduce the occurrence of aviation accidents. QAR data has its own unique analysis challenges. The data are high-dimensional, with complex and non-normally distributed relationships among variables. It is more critical that there are significant differences in data distribution across aircraft types and airlines. These differences lead to the problem that when constructing traditional prediction models based on the single source data, new aircraft types and some routes encounter a smaller sample size, and the generalisation ability and prediction accuracy do not meet the demand.

In order to study the relationship between landing distance and flight parameters, Wang et al. [1] established logistic regression and linear regression models, respectively. The traditional mean regression can only capture the central trend of the distribution, while the flight safety problem often involves the extreme situations, which generally exist at the tail of the conditional distribution of the response variable. Therefore, Chen et al. [2] employed quantile regression to identify the key factors affecting the maximum vertical acceleration during the landing phase and used the dimension reduction method to verify the results. Finally, the form of the action of key factors on hard landing was systematically expounded. More recently, Zhao J et al. [3] proposed the use of generative adversarial networks to fill missing values in QAR data and achieved a significant improvement in data quality by introducing a self-attention mechanism. However, the problem of insufficient sample size of new aircraft models has not been solved. We consider transfer learning to transfer from existing tasks, so as to achieve better performance on limited datasets [4, 5].

A central issue in transfer learning is how to effectively transfer the information from the sources to the target. Using only target data will ignore the information from the sources that may be useful. If all sources are used, it may ignore the potential heterogeneity between the sources and the target, resulting in 'negative transfer'. The heterogeneity between domains is mainly manifested as the following three aspects. (1) Data distribution differences: the distribution of covariate $f(\boldsymbol{x}_i)$ may be different, which is called covariate shift in some related literature. (2) Model differences: there may also be differences in the regression coefficients of the conditional distribution $f(y_i \mid \boldsymbol{x}_i)$. (3) Error distribution differences: the error distribution of different models may also be significantly different.

The current transfer learning framework mainly solves the first two heterogeneous problems [6]. Li et al. [7] considered the complex situation of multiple sources, proposed a two-step transfer learning algorithm under the high-dimensional linear model, and further discussed how to detect the informative set when the informative set is unknown. Tian and Feng [8] extended the two-step transfer learning algorithm to generalised linear model and proposed a more simple informative set detection method, which made multi-source transfer learning more efficient and convenient. Chen et al. [2] studied the high-dimensional linear regression model with heavy-tailed noise and established a new relationship between quantile regression and linear regression by transforming the response variables. Jin et al. [9] developed a transfer learning method for high-dimensional quantile regression, which achieves smaller errors based on sample size, signal-to-noise ratio, etc.

In the practical application of QAR data, we may not be able to determine a specific quantile point. Through residual analysis, we can find that the data does not obey the normal distribution and has heteroscedasticity, and the variables are scattered near the adjacent quantile points. This situation will be unstable due to the quantile point disturbance. Quantile regression shows its superiority in dealing with data heterogeneity and heavy-tailed error [10–13]. Zou and Yuan [14] proposed the composite quantile

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

regression (CQR) model, which considers multiple quantile points and captures the distribution information of different quantile points of the data. The theoretical results show that this not only improves the performance of the model, but also enhances the adaptability to complex data distribution. In view of the above challenges, we propose a transfer learning based on the composite quantile regression model, which uses the source data to assist the target modeling and synthesises multiple quantile information, so as to make the estimation more robust.

Compared with the existing work under QAR data [15, 16], the composite quantile regression model based on transfer learning proposed in this paper has several advantages. (1) A transfer learning framework for hard landing prediction of QAR data is constructed. The composite quantile loss is transformed into a linear form by Newton-Raphson iteration, so that the transfer learning can realise the composite quantile regression with non-differentiable loss. (2) This paper considers the information of multiple quantile points at the same time, and the model is more robust when the quantile points are unknown. (3) When some routes encounter a small amount of data, other source information can be introduced. The application on the QAR dataset shows that the algorithm significantly improves the performance of the model.

The rest of this article is organised as follows. In Section 2, the construction of the composite quantile regression model in transfer learning is introduced. In Section 3, the pseudo response is first constructed based on the Newton-Raphson iteration, and the transfer learning algorithm based on the composite quantile regression model is described. When the informative set is unknown, a cross-validation algorithm is proposed to detect the informative set. Section 4 presents the theoretical results of the proposed method. Section 5 provides numerical simulation results to verify the performance of the proposed algorithm. Section 6 shows the application of the algorithm in the aircraft hard landing problem. Finally, Section 7 summarises the full text and discusses the possible direction of the future work. The proofs of theorem and numerical analysis discussions are listed in the Appendix.

2.0 Transfer learning settings of composite quantile regression

2.1 Composite quantile regression model

This paper mainly studies the problem of data insufficiency and quantile perturbation in QAR data. Consider a dataset consisting of n independent observations $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$, where the response variable y_i and the covariate vector $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ are collected independently from the random variables y and $\boldsymbol{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$. Consider the following quantile regression model

$$y_i = \boldsymbol{x_i}^T \boldsymbol{\beta} + \varepsilon_i, P(\varepsilon_i \leq 0 | \boldsymbol{x_i}) = \tau, i = 1, 2, \ldots, n, \tag{1}$$

where $\boldsymbol{\beta}$ is the true regression coefficient at the quantile level τ, and ϵ is a random error with a density function $f(\cdot)$. The covariate $\boldsymbol{x}$ is independent of noise ϵ. The quantile loss of the fixed quantile point is defined as $\rho_\tau(u) = u[\tau \cdot I_{(u \geq 0)} - (1 - \tau) \cdot I_{(u \leq 0)}]$, where $I(\cdot)$ is an indicator function. Given the observed data $(\boldsymbol{x}_i, y_i)_{i=1}^n$, the quantile regression estimator is obtained by solving

$$\hat{\boldsymbol{\beta}}^{QR} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} \right). \tag{2}$$

In practical applications, a specific quantile level may not be predetermined, and the estimation may suffer from instability due to perturbations in the quantile level. Therefore, composite quantile regression, as an extension of quantile regression, can consider the information of multiple quantile points at the same time. Suppose that there are S quantiles satisfying $0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_S \leq 1$, the parameter estimation of composite quantile regression is the solution of the following optimisation problem

$$\hat{\boldsymbol{\beta}}^{CQR} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{nS} \sum_{s=1}^{S} \sum_{i=1}^{n} \rho_{\tau_s} \left(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} \right). \tag{3}$$

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

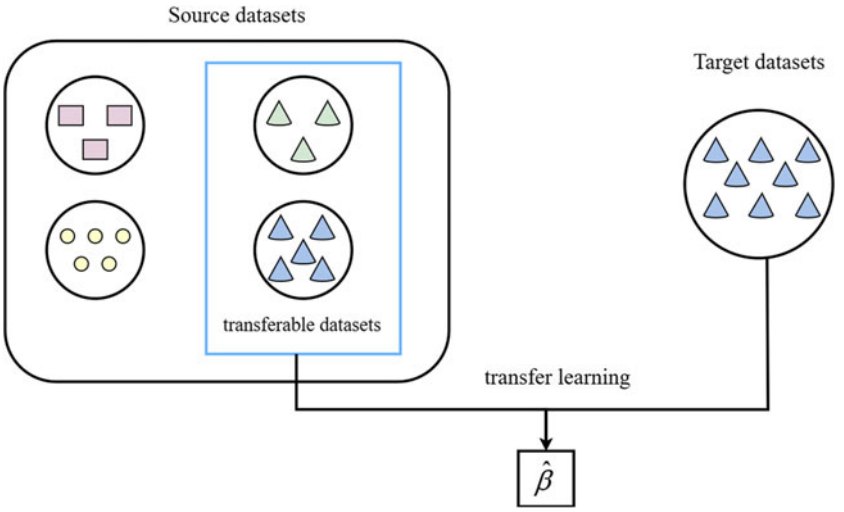


Figure 1. *Transfer learning graphical abstract.*

2.2 Multi-source transfer learning setting

This paper extends the CQR framework to a transfer learning setting with multiple sources. It is assumed that the target dataset of the study is expressed as $(\boldsymbol{X}^{(0)}, \boldsymbol{y}^{(0)}) \in \mathbb{R}^{n_0 \times p} \times \mathbb{R}^{n_0}$, where n_0 is the sample size of the target dataset. There are K source datasets, where the kth source data is represented as $(\boldsymbol{X}^{(k)}, \boldsymbol{y}^{(k)}) \in \mathbb{R}^{n_k \times p} \times \mathbb{R}^{n_k}$, where n_k is the sample size of the kth source dataset. The parameters of the composite quantile regression model established on the kth dataset are denoted as $\boldsymbol{\beta}^{(k)}$, $k = 0, 1, \ldots, \mathrm{K}$.

As shown in Fig. 1, the core principle of transfer learning is that if the kth source model has a certain degree of similarity to the target model, the information of the kth source dataset can be appropriately transferred and utilised. The sparsity of the difference between $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\beta}^{(k)}$ is defined as the criterion for measuring the difference of the model, i.e. $\|\boldsymbol{\delta}^{(k)}\|_1 = \|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^{(k)}\|_1$. If the similarity satisfies $\|\boldsymbol{\delta}^{(k)}\|_1 \leq d$, it is considered that the kth source is a transferable source, where d is a non-negative value to be specified. The informative set is a set of transferable source ordinal numbers, defined as $\mathcal{A} = \left\{1 \leq k \leq K \ \|\boldsymbol{\delta}^{(k)}\|_1 \leq d\right\}$. It is necessary to adjust the similarity according to different needs. Excessive selection of d value may lead to negative transfer, that is the source information will damage the generalisation performance of the target model. If the value of d is too small, some potential source information may be lost. Therefore, it is necessary to select the appropriate similarity by regularisation path analysis or cross-validation.

3.0 Parameter estimation based on transfer learning

This section will introduce the main content of the algorithm. Firstly, the relationship between composite quantile regression and linear regression is established in the context of transfer learning, and then it is integrated into the two-step transfer learning process. Since the informative set of the target task is often unknown in practical applications, this section presents an informative set detection algorithm based on sample splitting.

3.1 Construct pseudo responses

Consider the following composite regression optimisation problem

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} E(G(\boldsymbol{\beta}; \boldsymbol{X}, \boldsymbol{Y})), \tag{4}$$

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

where the composite quantile loss is $G(\boldsymbol{\beta}; \boldsymbol{X}, \boldsymbol{Y}) = \sum_{s=1}^{S} \rho_{\tau_s}(\boldsymbol{y} - \boldsymbol{x}^T\boldsymbol{\beta})$. The Newton-Raphson iterative algorithm is based on the local quadratic approximation principle of the target function, and the roots of the equation are solved iteratively. The iterative formula of the estimation equation is as follows

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 - H(\boldsymbol{\beta}_0)^{-1} E[g(\boldsymbol{\beta}_0; \boldsymbol{X}, \boldsymbol{Y})], \tag{5}$$

where $\boldsymbol{\beta}_0$ is the initial value, $g(\boldsymbol{\beta}; \boldsymbol{X}, \boldsymbol{Y})$ is the subgradient of the loss function $G(\boldsymbol{\beta}; \boldsymbol{X}, \boldsymbol{Y})$ and $H(\boldsymbol{\beta}) := \frac{\partial E(g(\boldsymbol{\beta}; \boldsymbol{X}, \boldsymbol{Y}))}{\partial \boldsymbol{\beta}}$ is the overall Hessian matrix. The corresponding subgradient is given by $g(\boldsymbol{\beta}; \boldsymbol{X}, \boldsymbol{Y}) = \sum_{s=1}^{S} \boldsymbol{X}[I(\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{\beta} \le 0) - \tau_s]$, and the Hessian matrix is $H(\boldsymbol{\beta}) = \sum_{s=1}^{S} E(\boldsymbol{X}\boldsymbol{X}^T f(\boldsymbol{X}^T(\boldsymbol{\beta} - \boldsymbol{\beta}^*)))$,where $f(\cdot)$ is the density function of the noise ϵ. When the initial estimator $\boldsymbol{\beta}_0$ is close enough to the true parameter $\boldsymbol{\beta}^*$, $H(\boldsymbol{\beta}_0)$ is close to $H(\boldsymbol{\beta}^*) = S\boldsymbol{\Sigma}f(0)$, where $\boldsymbol{\Sigma} = E(\boldsymbol{X}\boldsymbol{X}^T)$ is the overall covariance matrix. Substituting $H(\boldsymbol{\beta}^*)$ into (5) yields

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 - H(\boldsymbol{\beta}^*)^{-1} E[g(\boldsymbol{\beta}_0; \boldsymbol{X}, \boldsymbol{Y})]. \tag{6}$$

The first-order Taylor expansion of $E(g(\boldsymbol{\beta}_0; \boldsymbol{X}, \boldsymbol{Y}))$ at point $\boldsymbol{\beta}^*$ is

$$E(g(\boldsymbol{\beta}_0; \boldsymbol{X}, \boldsymbol{Y})) = H(\boldsymbol{\beta}^*)(\boldsymbol{\beta}_0 - \boldsymbol{\beta}^*) + O\big(\|\boldsymbol{\beta}_0 - \boldsymbol{\beta}^*\|_2^2\big). \tag{7}$$

Substituting this approximation into (6) yields

$$\begin{aligned} \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}^*\|_2 &= \|\boldsymbol{\beta}_0 - \boldsymbol{H}(\boldsymbol{\beta}^*)^{-1}\boldsymbol{E}(\boldsymbol{g}(\boldsymbol{\beta}_0; \boldsymbol{X}, \boldsymbol{Y})) - \boldsymbol{\beta}^*\|_2 \\ &= O\big(\|\boldsymbol{\beta}_0 - \boldsymbol{\beta}^*\|_2^2\big). \end{aligned} \tag{8}$$

That is to say, given a consistent initial estimator $\boldsymbol{\beta}_0$, we can improve it by the (6) iterative method. The initial estimator $\boldsymbol{\beta}_0$ can be obtained using the standard Lasso procedure. For specific accuracy-level requirements of the initial estimator, refer to Chen et al. [2]. In order to solve the problem of how to transform composite quantile regression into linear regression optimisation, Equation (6) is first rewritten as

$$\begin{aligned} \boldsymbol{\beta}_1 &= \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}\boldsymbol{\beta}_0 - \frac{1}{S}f^{-1}(0)E(g(\boldsymbol{\beta}_0; \boldsymbol{X}, \boldsymbol{Y}))\right) \\ &= \boldsymbol{\Sigma}^{-1}\left(E\left[\boldsymbol{X}\left\{\boldsymbol{X}^T\boldsymbol{\beta}_0 - \frac{1}{S}f^{-1}(0)\sum_{s=1}^{S}\left[I\left(\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{\beta} \le 0\right) - \tau_s\right]\right\}\right]\right). \end{aligned} \tag{9}$$

Motivated by the form of 9, we define a new response variable $\tilde{\boldsymbol{Y}}$ as

$$\tilde{\boldsymbol{Y}} = \boldsymbol{X}^T\boldsymbol{\beta}_0 - \frac{1}{S}f^{-1}(0)\sum_{s=1}^{S}[I(\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{\beta} \le 0) - \tau_s]. \tag{10}$$

Then Equation (9) can be written in the form of linear regression estimation, and the parameter estimation becomes the solution of the following optimisation problem $\boldsymbol{\beta}_1 = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min}\, E(\tilde{\boldsymbol{Y}} - \boldsymbol{X}^T\boldsymbol{\beta})^2$.

3.2 Oracle transfer learning

The design of Algorithm 1 in this paper follows, on a theoretical level, the analytical framework proposed by Chen et al. [2]. This framework demonstrates that for high-dimensional sparse models, if the initial estimator achieves a certain convergence rate and recovers the true support set with high probability, then a one-step Newton-type approximate update can attain nearly optimal estimation efficiency. Building on this theoretical guarantee, this section first introduces an oracle transfer learning method in Algorithm 1, assuming that the informative set has been given, that is, we know that it is most advantageous to transfer information from certain specific sources.

The first step performs composite quantile regression on different sources in a distributed manner, rather than concentrating the data together. This process can be achieved in parallel, so that the calculation time is mainly determined by the maximum local sample size $\max_{k \in \mathcal{A} \cup \{0\}} n_k$ in all sources. Therefore,

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

Algorithm 1. Oracle Trans-Lasso CQR

Input: $\{\boldsymbol{X}^{(0)}, \boldsymbol{y}^{(0)}\}, \{\boldsymbol{X}^{(k)}, \boldsymbol{y}^{(k)}\}_{k=1}^{K}$;

Output: $\hat{\boldsymbol{\beta}}$;

Step 1: For each $k \in \mathcal{A} \cup \{0\}$, compute

$$\hat{\boldsymbol{\beta}}^{(k)} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{n_k S} \sum_{s=1}^{S} \sum_{i=1}^{n_k} \left\{\tau_s - I\left(y_i^{(k)} - \boldsymbol{x}_i^{(k)T}\boldsymbol{\beta} \le 0\right)\right\}\left(y_i^{(k)} - \boldsymbol{x}_i^{(k)T}\boldsymbol{\beta}\right) + \lambda_0 \|\boldsymbol{\beta}\|_1.$$

Step 2: For each $k \in \mathcal{A} \cup \{0\}$, the kernel density estimator of the error density at zero point is

$$\hat{f}^{(k)}(0) = \frac{1}{n_k h^{(k)}} \sum_{i=1}^{n_k} K_h\left(\frac{y_i^{(k)} - \boldsymbol{x}_i^{(k)T}\hat{\boldsymbol{\beta}}^{(k)}}{h^{(k)}}\right),$$

where K_h is the kernel function and $h^{(k)}$ is the bandwidth.

The pseudo response variables are constructed as $\tilde{\boldsymbol{y}}^{(k)} = \left(\tilde{y}_1^{(k)}, \ldots, \tilde{y}_{n_k}^{(k)}\right)^T$,where

$$\tilde{y}_i^{(k)} = \boldsymbol{x}_i^{(k)T}\hat{\boldsymbol{\beta}}^{(k)} - \frac{1}{S}\{\hat{f}^{(k)}(0)\}^{-1} \sum_{s=1}^{S}\left[I\left(y_i^{(k)} - \boldsymbol{x}_i^{(k)T}\hat{\boldsymbol{\beta}}^{(k)} \le 0\right) - \tau_s\right], i = 1, \ldots, n_k.$$

Step 3: Compute $\hat{\boldsymbol{\beta}}^{\mathcal{A}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \frac{1}{2(n_A+n_0)} \sum_{k\in\mathcal{A}\cup\{0\}} \left\|\tilde{\boldsymbol{y}}^{(k)} - \boldsymbol{X}^{(k)}\boldsymbol{\beta}\right\|^2 + \lambda_1\|\boldsymbol{\beta}\|_1.$

Step 4: Let $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{\mathcal{A}} + \hat{\boldsymbol{\delta}}^{\mathcal{A}}$, where

$$\hat{\boldsymbol{\delta}}^{\mathcal{A}} = \arg\min_{\boldsymbol{\delta}\in\mathbb{R}^p} \frac{1}{2n_0}\|\tilde{\boldsymbol{y}}^{(0)} - \boldsymbol{X}^{(0)}(\hat{\boldsymbol{\beta}}^{\mathcal{A}} + \boldsymbol{\delta})\|^2 + \lambda_2\|\boldsymbol{\delta}\|_1.$$

compared with the traditional method, the estimation method proposed in Algorithm 1 not only enhances estimation accuracy but also achieves significant gains in computational efficiency. In the second step, the composite quantile regression problem is transformed into a more tractable Lasso problem by using the pseudo response. The third step is known as fusion learning, where the target and source data are combined to obtain a rough estimation. Finally, a debiasing step is conducted. The target dataset is subjected to l_1 penalty composite quantile regression again, and the bias $\hat{\boldsymbol{\delta}}^{\mathcal{A}}$ of the coefficient vector estimation is learned by offsetting $\hat{\boldsymbol{\beta}}^{\mathcal{A}}$.

3.3 Informative set detection

In practical applications, we do not know whether the source can be transferred, which makes the problem more difficult. This section introduces an informative set detection algorithm based on the sample splitting. To reliably identify transferable sources, we propose an informative set detection algorithm based on cross-validated composite quantile loss. The core idea is that whether a source is 'transferable' depends on the similarity of its conditional distribution to that of the target. By comparing the relative magnitude of the cross-validated loss of each source against a baseline (pure target model) loss, we can make a statistical judgement on its transferability. The advantage of this approach is that it does not rely on specific distributional conditions, is robust to outliers and is therefore well-suited for aviation data scenarios characterised by heterogeneity and heavy-tailed errors. The proposed algorithm is inspired by Tian and Feng [8].

Following this idea, we constructed pseudo response via Steps 1–2 of Algorithm 1. Similar to the idea of cross-validation, first, the target data is divided into q subsets. One subset is selected as the test set each time, and the remaining subsets are used as the training set. Secondly, under this subset, the training set data is used to fit the composite quantile regression with l_1 penalty, and the loss on the test set is calculated. Each source dataset and training set are combined to estimate, and then the test set and coefficient estimation are added to the loss $\hat{G}_0^{(0)}$. Similarly, the average cross-validation loss $\hat{G}_0^{(k)}$ of each source is calculated. This process is repeated over all q folds, and the average cross-validation losses $\hat{G}_0^{(0)}$ and $\hat{G}_0^{(k)}$ are obtained. Finally, the informative set is obtained by comparing $\hat{G}_0^{(0)}$ and $\hat{G}_0^{(k)}$ of each k.

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

Algorithm 2. Informative set detection

Input: $\{\boldsymbol{X}^{(0)}, \boldsymbol{y}^{(0)}\}$, $\{\boldsymbol{X}^{(k)}, \boldsymbol{y}^{(k)}\}_{=1}^{K}$, threshold value ϵ_0, q,
penalty parameter $\{\{\lambda^{(k)[a]}\}_{k=0}^{K}\}_{a=1}^{q}$;
Output: informative set $\widehat{\mathcal{A}}$;
Step 1: Construct pseudo response $\tilde{\boldsymbol{y}}^{(k)}$, $i = 1, \ldots, n_k, k = 1, \ldots, K$.
Step 2: $\{\boldsymbol{X}^{(0)}, \boldsymbol{Y}^{(0)}\}$is randomly divided into q equal sets $\{\boldsymbol{X}^{(0)[i]}, Y^{(0)[i]}\}_{i=1}^{q}$.
Step 3: For each $a = 1, \ldots, q$,
$\hat{\boldsymbol{\beta}}^{(0)[a]} \leftarrow$ fit the penalised composite quantile regression with penalty parameter $\lambda^{(0)[a]}$ on $\{\boldsymbol{X}^{(0)[i]}, \boldsymbol{y}^{(0)[i]}\}_{i=1}^{q} \backslash (\boldsymbol{X}^{(0)[a]}, \boldsymbol{y}^{(0)[a]})$,
$\hat{\boldsymbol{\beta}}^{(k)[a]} \leftarrow$ run the transferring step in Algorithm 1 with $\lambda^{(0)[a]}$ on $\{\boldsymbol{X}^{(0)[i]}, \boldsymbol{y}^{(0)[i]}\}_{i=1}^{q} \backslash (\boldsymbol{X}^{(0)[a]}, \boldsymbol{y}^{(0)[a]}) \cup (\boldsymbol{X}^{(k)}, \boldsymbol{y}^{(k)})$ and penalty parameter $\lambda^{(k)[a]}$ for $k = 1, \ldots, K$.
Calculate the loss function $\hat{G}_0^{[a]}(\hat{\boldsymbol{\beta}}^{(k)[a]})$ on $(\boldsymbol{X}^{(0)[a]}, \boldsymbol{y}^{(0)[a]})$.
Step 4: $\hat{G}_0^{(k)} \leftarrow \sum_{a=1}^{q} \hat{G}_0^{[a]}(\hat{\boldsymbol{\beta}}^{(k)[a]})/q$, $\hat{G}_0^{(0)} \leftarrow \sum_{a=1}^{q} \hat{G}_0^{[a]}(\hat{\boldsymbol{\beta}}^{(0)[a]})/q$.
$\widehat{\mathcal{A}} \leftarrow \left\{k \neq 0 : \hat{G}_0^{(k)} \leq (1+\epsilon_0)\hat{G}_0^{(0)}\right\}$.

Compared with the model aggregation method of Li et al. [6], which requires repeatedly executing an Oracle transfer learning algorithm over a set of candidate estimates, our Algorithm 2 offers a simpler and more efficient alternative. Notably, our Algorithm 2 does not require input bandwidth, which significantly reduces the computational burden. This efficiency gain is particularly beneficial when the target sample size n_0 is small. For instance, the bootstrap method [17] can be used to enrich sample, which can solve the statistical problem of small sample size. Assuming that we have an original sample dataset X (sample size is n), we can put back random sampling N data, so that we can obtain a new dataset with a sample size of N. When the dataset class imbalance can also use the ROSE oversampling method for enrichment.

4.0 Theoretical results

In this section we provide the theoretical results for our transfer learning method. We assume the following regularity conditions.

Assumption 1. *In the k-th source, $\boldsymbol{\beta}_i^{(k)}$ is assumed to be sparse with s_k nonzero components, and $s_k = O(n_k^b)$ for some $0 < b < 1/3$. Besides, $\|\boldsymbol{\beta}^{(k)}\|$ is bounded.*

Assumption 2. *For the k-th source, we assume the rows of $\boldsymbol{X}^{(k)}$ are sub-Gaussian and its covariance matrix $\boldsymbol{\Sigma}^{(k)}$ is bounded away from 0 and ∞.*

Given the potential for distributional shifts in features, where the $\boldsymbol{\Sigma}^{(k)}$ matrices may differ, the following quantity is defined to measure the maximum discrepancy between $\boldsymbol{\Sigma}^{(0)}$ and $\boldsymbol{\Sigma}^{(k)}$ for $k \in \mathcal{A}$:

$$C_{\boldsymbol{\Sigma}} = 1 + \max_{1 \leq j \leq p} \max_{k \in \mathcal{A}} \left\| \boldsymbol{e}_j^{\top} (\boldsymbol{\Sigma}^{(k)} - \boldsymbol{\Sigma}^{(0)}) \left(\sum_{k \in \mathcal{A} \cup \{0\}} \alpha_k \boldsymbol{\Sigma}^{(k)} \right)^{-1} \right\|_1,$$

where $\boldsymbol{e}_j$ denotes the unit vector with its jth component equal to 1, and the weight $\alpha_k = n_k/(n_{\mathcal{A}} + n_0)$. Under Assumption 2, $C_{\boldsymbol{\Sigma}}$ is bounded by a constant that does not depend on p or on any of the n_k.

Assumption 3. *Assume*

$$\max \left\{ s_0 \log p / \left(\min_{k \in \mathcal{A} \cup \{0\}} n_k + n_0 \right), C_{\boldsymbol{\Sigma}} d (\log p / n_0)^{1/2} \right\} = o(1).$$

Besides, we assume $K = O(n_0)$ and $|\mathcal{A}|$ is bounded.

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

Assumption 4. *Let* $S_k = \{0 \le i \le p : \boldsymbol{\beta}_i^{(k)} \neq 0\}$. *Suppose that* $\boldsymbol{\Sigma}^{(k)} = \mathbb{E}\left(\boldsymbol{X}^{(k)}\boldsymbol{X}^{(k)\mathrm{T}}\right)$ *satisfies*

$$\left\| \boldsymbol{\Sigma}^{(k)}_{S_k^c \times S_k} \left(\boldsymbol{\Sigma}^{(k)}_{S_k \times S_k} \boldsymbol{g}\right)^{-1} \right\|_1 \le 1 - r,$$

for some $0 < r < 1$. *Also assume that* $c_0^{-1} \le \lambda_{\min}(\boldsymbol{\Sigma}^{(k)}) \le \lambda_{\max}(\boldsymbol{\Sigma}^{(k)}) \le c_0$ *for some constant* $c_0 > 0$.

Assumption 5. *The error density function* $f(\cdot)$ *is bounded and Lipschitz continuous. Moreover, we assume* $f(0) > c > 0$ *for some constant c.*

Assumption 6. *Assume the kernel function* $K_h(\cdot)$ *is integrable, and* $\int_{-\infty}^{+\infty} K_h(x)dx = 1$. *At the same time,* $K_h(x) = 0$ *if* $|x| \ge 1$, *and* $K_h(\cdot)$ *is differentiable and bounded.*

Assumption 7. *For each source* $k = 0, 1, \ldots, \mathrm{K}$, *let* $\boldsymbol{\beta}_\tau^{(k)}$ *denote the true coefficient at quantile level* τ. *There exists a constant* $L > 0$ *such that for any* $\tau, \tau' \in \{\tau_1, \ldots, \tau_S\}$,

$$\left\| \boldsymbol{\beta}_\tau^{(k)} - \boldsymbol{\beta}_{\tau'}^{(k)} \right\|_2 \le L|\tau - \tau'|.$$

Moreover, the average parameter $\bar{\boldsymbol{\beta}}^{(k)} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\beta}_{\tau_s}^{(k)}$ *satisfies* $\|\bar{\boldsymbol{\beta}}^{(k)}\|_2 = O(1)$.

Assumption 8. *For each* $k = 0, \ldots, K$ *and each quantile level* τ_s $(s = 1, \ldots, S)$, *the error density* $f^{(k)}(\cdot)$ *satisfies* $f^{(k)}(F^{-1}(\tau_s)) > c > 0$ *and is Lipschitz continuous in a neighbourhood of* $F^{-1}(\tau_s)$.

Assumption 9. *The initial estimator* $\hat{\boldsymbol{\beta}}_0$ *satisfies* $\|\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}^*\|_2 = O_{\mathcal{P}}\left(\sqrt{\frac{s_0 \log \max(n,p)}{n_0}}\right)$.

Assumption 1 specifies the sparsity level s_k. The requirement $s_k = O(n_k^b)$ ensures that the proposed algorithm attains a near-oracle convergence rate within a finite number of iterations. Assumption 2 imposes regularity assumptions on the distribution of $\boldsymbol{X}$ and the covariance matrix $\boldsymbol{\Sigma}$. Assumption 3 pertains to the convergence rate and support recovery properties required of the initial estimator. Assumption 4 corresponds to the classical irrepresentable condition, frequently employed in high-dimensional statistics to guarantee support recovery [4]. Assumption 5 ensures that the error density function is sufficiently smooth near zero, a standard requirement for establishing the consistency and asymptotic normality of quantile regression estimators. Assumption 6 outlines standard assumptions on the kernel function $K_h(\cdot)$, which approximates the Dirac delta (the derivative of the indicator function). The specific form of the kernel is given by

$$K_h(x) = \begin{cases} 0 & \text{if } |x| \ge 1; \\ -\frac{315}{64}x^6 + \frac{735}{64}x^4 - \frac{525}{64}x^2 + \frac{105}{64} & \text{if } |x| < 1. \end{cases} \tag{11}$$

The composite quantile regression model can be expressed as a finite linear combination of quantile regression models, and its theoretical properties (such as consistency and asymptotic normality) remain valid under appropriate regularity conditions. Condition 7 and Condition 8 collectively ensure the regularity of parameters as they vary across quantile levels in composite quantile regression, as well as the well-behaved nature of error densities at the target quantile points, providing the necessary mathematical foundation for bias control and the asymptotic behaviour of estimators in theoretical derivations. Condition 9 specifies requirements on the convergence rate and support recovery capability of the initial estimator. We remark that in the pseudo response construction, the initial estimator $\widehat{\boldsymbol{\beta}}_0$ is constructed by solving a high-dimensional composite quantile regression problem using only the target data.

Subsequently, this paper refers to Chen's method [2], which establishes a connection between quantile regression and ordinary linear regression. Under these assumptions, Theorem 1 gives the error rates for the estimation and prediction with the oracle Trans-Lasso CQR estimator in Section 3.2.

Theorem 1 *Suppose Conditions 1–8 hold. We take* $\lambda_1 \asymp \sqrt{\log\left(p \vee \max_{k \in \mathcal{A}\cup\{0\}} n_k\right) / (n_{\mathcal{A}} + n_0)}$ *and* $\lambda_0 \asymp \lambda_2 \asymp \sqrt{\log(p \vee n_0)/n_0}$, *then with*

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

$$\xi_1\Big(n_0, n_{\mathcal{A}}, \max_{k\in\mathcal{A}\cup\{0\}} n_k, s_0, p\Big) = \frac{s_0 \log\Big(p \vee \max\limits_{k\in\mathcal{A}\cup\{0\}} n_k\Big)}{n_{\mathcal{A}} + n_0} + \frac{s_0 \log(p \vee n_0)}{n_0} \wedge C_{\boldsymbol{\Sigma}} d \sqrt{\frac{\log(p \vee n_0)}{n_0}} \wedge (C_{\boldsymbol{\Sigma}} d)^2,$$

we have

$$\mathbb{P}\left(\frac{1}{n_0}\left\|\boldsymbol{X}^{(0)}\Big(\widehat{\boldsymbol{\beta}}^{\mathrm{CQR}} - \bar{\boldsymbol{\beta}}^{(0)}\Big)\right\|^2 \vee \left\|\widehat{\boldsymbol{\beta}}^{\mathrm{CQR}} - \bar{\boldsymbol{\beta}}^{(0)}\right\|^2 \lesssim \xi_1\Big(n_0, n_{\mathcal{A}}, \max_{k\in\mathcal{A}\cup\{0\}} n_k, s_0, p\Big)\right)$$
$$\geq 1 - C_1 \exp\Big(-C_2 \min_{k\in\mathcal{A}\cup\{0\}} n_k\Big) - C_3\Big(\min_{k\in\mathcal{A}\cup\{0\}} n_k\Big)^{-C_4 \underline{s}_{\mathcal{A}}} - \exp(-C_5 \log p),$$

for some positive constants $C_1, \ldots, C_5$, *where* $\underline{s}_{\mathcal{A}} = \min_{k\in\mathcal{A}\cup\{0\}} s_k$, *mean parameter* $\bar{\boldsymbol{\beta}}^{(0)} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\beta}_{\tau_s}^{(0)}$.

Relative to the target-only estimator whose squared ℓ_2 error rate for $\boldsymbol{\beta}$ is $s_0 \log(p \vee n_0)/n_0$ [11], the oracle transfer learning estimator attains a faster convergence rate when $C_{\boldsymbol{\Sigma}} d \leq \sqrt{s_0 \log(p \vee n_0)/n_0}$ and $n_{\mathcal{A}} \gg n_0$. The first inequality essentially requires a small heterogeneity d when $C_{\boldsymbol{\Sigma}}$ remains bounded, which is naturally satisfied under favourable covariance structures of informative source domains.

In the least favourable scenario where no informative source is available, the oracle estimator reduces to the target-only estimator and achieves the same convergence rate $s_0 \log(p \vee n_0)/n_0$. Theorem 1 further indicates that smaller heterogeneity between informative sources and the target, as quantified by $C_{\boldsymbol{\Sigma}}$, leads to improved estimation accuracy.

The informative set detection method proposed in Algorithm 2 is primarily based on a cross-validation strategy, the theoretical reliability of which has been substantiated by authoritative existing research. Specifically, Tian and Feng [8] systematically demonstrate that, under certain assumptions, the set of sources selected through cross-validation loss is theoretically consistent. They further prove the convergence rate advantages of transfer learning based on this selected set. Therefore, the theoretical justification for informative set detection in this work mainly relies on the theoretical framework established in that literature, which serves as the validity foundation for the source detection step in our methodology. For detailed proofs, refer to Ref. [8].

5.0 Simulation studies

To validate the performance of the proposed algorithm in terms of transfer learning and heavy-tailed error handling, we conducted a numerical simulation study. The l_1 penalty composite quantile regression (L CQR) is fitted on the target data to confirm the potential loss of unused source data. Compared to the Naive Trans-Lasso composite quantile regression (Naive TL CQR) using all source data without discrimination, the algorithm proposed in this paper is Oracle TL CQR and TL CQR using only transferable source data. In addition, the difference between the composite quantile regression estimation and the quantile regression estimation (TL QR) is also compared.

When the number of informative set is known to be m, the informative set estimation in the Pseudo TL CQR is

$$\widehat{\mathcal{A}} = \big\{k \neq 0 : \hat{G}_0^{(k)} \leq \hat{G}_0^{(k)}(m)\big\}, \tag{12}$$

where $\hat{G}_0^{(k)}(m)$ represents the mth order statistic among all the empirical loss $\hat{G}_0^{(k)}$ calculated by Step 3 in the informative set detection.

There exists a small value $\epsilon_0 \in (0, c_\epsilon]$ given in Algorithm 2, where c_ϵ is a constant. It is used to control the strictness of the selection. We can estimate c_ϵ by

$$\widehat{c}_\varepsilon = \inf_{k\in\{1,\ldots,K\}} \frac{\big(\hat{\boldsymbol{\beta}}^{(k)} - \hat{\boldsymbol{\beta}}^{(0)}\big)^\top \hat{\boldsymbol{\Sigma}}^{(0)} \big(\hat{\boldsymbol{\beta}}^{(k)} - \hat{\boldsymbol{\beta}}^{(0)}\big)}{\hat{G}_0^{(0)}}, \tag{13}$$

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

Table 1. *Summary of hyperparameter settings*

Hyperparameter	Notation	Values/range
Similarity	d	$\{2, 10, 20\}$
Informative set size	$\lvert\mathcal{A}\rvert$	$\{0, 4, 8, 12, 16, 20\}$
Composite quantile level	τ	$c = (0.7, 0.8, 0.9)$
Bandwidth	h	chosen via 10-fold cross-validation
Fusion penalty parameter	λ_1	chosen via 10-fold cross-validation
Debiasing penalty parameter	λ_2	$\lambda_1\sqrt{\frac{n_{\mathcal{A}}+n_0}{n_0}}$ Li et al. [6]
Kernel function	$K_h(\cdot)$	Equation (11)

where $\hat{\boldsymbol{\beta}}^{(k)}$, $k = 0, \ldots, K$ are the initial single-source estimators from Step 1 in Algorithm 1, and $\hat{\boldsymbol{\Sigma}}^{(0)}$ is an empirical estimator of $\boldsymbol{\Sigma}^{(0)}$ with the target data, e.g. the sample covariance matrix of $\boldsymbol{X}^{(0)}$. According to the simulation results, it can be found that when $\varepsilon_0 = 0.01$, the efficiency of the algorithm is optimal. In order to simplify the calculation, the numerical simulation part ϵ_0 is set as 0.01. Detailed results can be found in the Appendix B.

5.1 Homogeneous design

Assuming that the informative set is known, we select the homogeneous and light-tailed data to simulate the process of data generation. Referring to the design of Ref. [9], for the target and all sources, the covariates are generated by the Gaussian distribution $N_p(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = [0.5^{|i-j|}]_{p\times p}$. Consider the following composite quantile regression model, from which the target and source datasets are generated independently by

$$\boldsymbol{y}^{(k)} = \boldsymbol{X}^{(k)T}\boldsymbol{\beta}^{(k)} + \varepsilon^{(k)}, k = 0, 1, \ldots, K. \tag{14}$$

We take $p = 150, n_0 = 150$ and $n_1 = n_2 = \cdots = n_K = 150$ with $K = 20$. In the composite quantile regression model, the sparsity is set to $s_0 = 20$. The true parameter vector $\boldsymbol{\beta}^{(0)}$ can be expressed as $\boldsymbol{\beta}^{(0)} = (I_{s_0}, \mathbf{0}_{p-s_0})^T$, where I_{s_0} is the s_0-dimensional unit vector and $\mathbf{0}_{p-s_0}$ is the $(p - s_0)$-dimensional zero vector. For the error, $\varepsilon_i^{(k)} \sim N(\Phi^{-1}(0.2), \boldsymbol{\beta}^{(k)\top}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(k)}/20)$, where Φ^{-1} is the quantile function of normal distribution $N(0, \boldsymbol{\beta}^{(k)\top}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(k)}/20)$.

For the kth source dataset, $\boldsymbol{\beta}_j^{(k)}$ is the jth element of $\boldsymbol{\beta}^{(k)}$. Let $\boldsymbol{\beta}_j^{(k)} = \boldsymbol{\beta}_j^{(0)} + r_j I(j \in H_k)$, where H_k is a set of $p/2$ randomly selected from $\{1, \ldots, p\}$. For the random variable, we set

$$r_j \overset{i.i.d.}{\sim} \begin{cases} Laplace(0, 2d/p) & \text{if k } \in \mathcal{A}; \\ Laplace(0, 140/p) & \text{if k } \notin \mathcal{A}. \end{cases} \tag{15}$$

Consider the similarity $d \in \{2, 10, 20\}$ and the informative set $|\mathcal{A}| \in \{0, 4, 8, 12, 16, 20\}$. When the similarity d increases, the variance of r_j also increases, which means that the degree of random fluctuation increases. The quantile level is set to $\tau = c(0.7, 0.8, 0.9)$. In the process of minimising the loss function, appropriate penalty parameters and smoothing kernels must be selected. The kernel function is set as (11). To choose the bandwidth h, we employ 10-fold cross-validation (CV) by minimising the composite quantile loss based on pseudo-responses. Similarly, in the fusion step, λ_1 is tuned via 10-fold CV. The use of CV is motivated by its ability to provide a data-driven, out-of-sample assessment of model performance, thereby enhancing generalisation and avoiding overfitting – a crucial advantage in high-dimensional settings with complex data structures such as aviation QAR records. For the parameter λ_2, we adopt the theoretically derived scaling $\lambda_2 = \lambda_1\sqrt{\frac{n_{\mathcal{A}}+n_0}{n_0}}$ from Li et al. [6]. This relation ensures that the regularisation strength adapts to the sample-size ratio between sources and target, preserving the theoretical convergence guarantees while avoiding an additional computational burden. A summary table of hyperparameters is presented in Table 1.

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

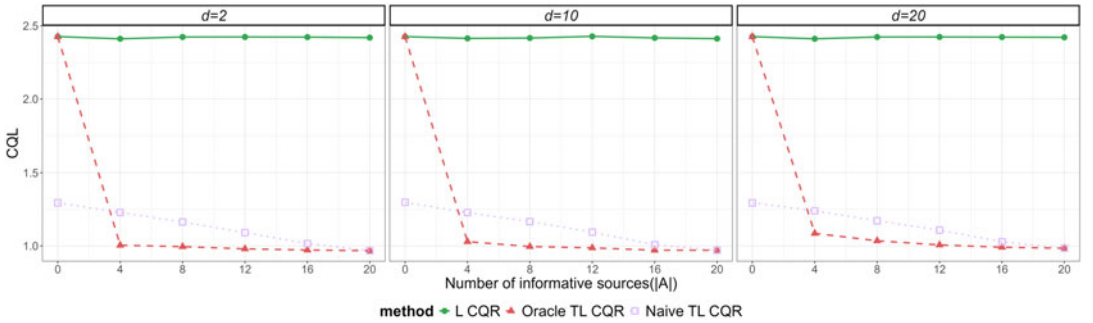


Figure 2. *Under the condition of different similarity d with a homogeneous design for normal error, average composite quantile loss between the number of transferable sources.*

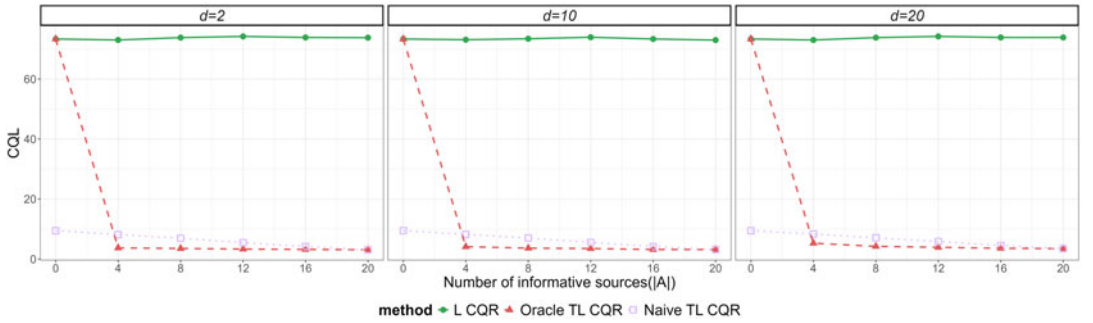


Figure 3. *Under the condition of different similarity d with a homogeneous design for normal error, average MSE between the number of transferable sources.*

We choose two measurement criteria to measure the statistical performance of each estimate, the mean square error (MSE) of the target coefficient vector $\boldsymbol{\beta}^{(0)}$, i.e. $\|\hat{\boldsymbol{\beta}}^{(0)} - \boldsymbol{\beta}^{(0)}\|^2$, and the out-of-sample composite quantile loss (CQL), i.e. $\frac{1}{n_t S}\sum_{s=1}^{S}\sum_{i=1}^{n_t}\rho_\tau(y_i^{(t)} - \boldsymbol{x}_i^{(t)\top}\hat{\boldsymbol{\beta}}^{(0)})$, where $(y_i^{(t)}, \boldsymbol{x}_i^{(t)})_{i=1}^{n_t}$ is an independently generated testing data of size $n_t = 150$ from the target model. A total of 100 independent repeated simulation experiments were carried out, and the mean and standard deviation of the composite quantile loss were trimmed by 10% to reduce the influence of extreme values in the heavy-tailed error setting. The MSE values were not processed. Finally, the mean values of all experimental results were taken. In the case of $p = 150$ and $n = 20$, Figs. 2–3 show the CQL and MSE values, respectively.

The simulation result in Fig. 2 demonstrates that the transfer learning method shows better estimation performance than L CQR, indicating that transfer learning method plays an active role in improving model performance. When $|\mathcal{A}|$ is the same, d becomes the standard to measure the similarity between the source datasets and the target dataset. The greater the value of d, the lower the similarity between the two datasets, and the worse the effect of transfer learning. Compared with Naive TL CQR using all sources, Oracle TL CQR estimation using only informative sets can reduce the negative transfer effects caused by other sources, and the results in Fig. 3 also support this conclusion.

5.2 Heterogeneous design

This section mainly evaluates the parameter estimation performance of each method under heterogeneous data. For the target, covariates are generated by multivariate normal distribution $N_p(0, \boldsymbol{I})$. For the sources, the kth covariate of the source comes from $\boldsymbol{x}_i^{(k)} \overset{i.i.d.}{\sim} N_p(0, \boldsymbol{\Sigma}_k)$, where the covariance matrix has

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

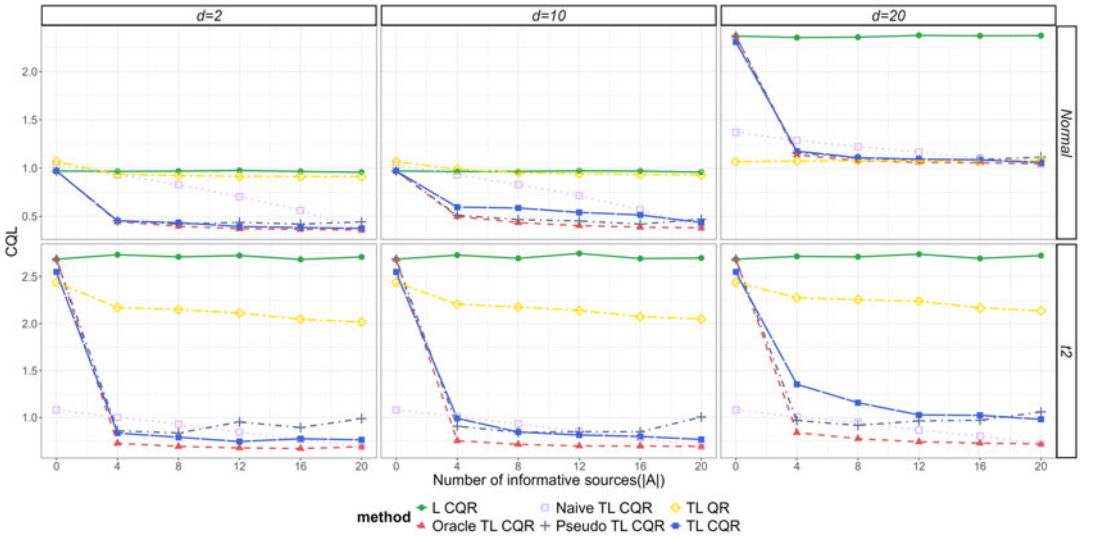


Figure 4. *Under the condition of different similarity d with heterogeneous design, the average composite quantile loss between the number of transferable sources.*

a Toeplitz structure, that is,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \cdots & \sigma_{p-2} & \sigma_{p-1} \\ \sigma_1 & \sigma_0 & \sigma_1 & & & \sigma_{p-2} \\ \sigma_2 & \sigma_1 & \sigma_0 & & & \\ \vdots & & & \ddots & & \vdots \\ \sigma_{p-2} & & & & \sigma_0 & \sigma_1 \\ \sigma_{p-1} & \sigma_{p-2} & & \cdots & \sigma_1 & \sigma_0 \end{pmatrix}.$$

with its first row given as

$$(\sigma_0, \ldots, \sigma_{p-1}) = (1, \underbrace{1/(k+1), \ldots, 1/(k+1)}_{2k-1}, \mathbf{0}_{p-2k}).$$

For the error distribution, we consider the light-tailed and heavy-tailed cases to study whether the transfer learning under composite quantile regression can solve the heavy-tailed problem. Other settings are the same as in Section 5.1.

- Case 1: $\epsilon_i^{(k)} \sim N(\Phi^{-1}(0.2), \boldsymbol{\beta}^{(k)\top}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(k)}/20)$, where Φ^{-1} is the quantile function of normal distribution $N(0, \boldsymbol{\beta}^{(k)\top}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(k)}/20)$.
- Case 2: $\epsilon_i^{(k)} \sim t(2)$.

Figure 4 shows the trend of the loss of different estimation methods with the informative set $|\mathcal{A}|$ when d is fixed. Under the normal setting, the loss value of each estimation method is less different, and the loss value under the complex setting of heavy-tailed error is larger. The difference of each estimation method is more obvious, and the performance of the proposed algorithm is the best. The loss of composite quantile regression transfer learning is lower than that of quantile regression transfer learning, which proves that considering multiple quantile points can improve the performance of the model.

In the case of $d = 2$ and $d = 10$, as the informative set increases, the effective data increases, and the estimation performance gradually increases. When $|\mathcal{A}|$ is small, the loss of Naive TL CQR estimation is worse than that of L CQR estimation. At this time, the amount of information is less and there is

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

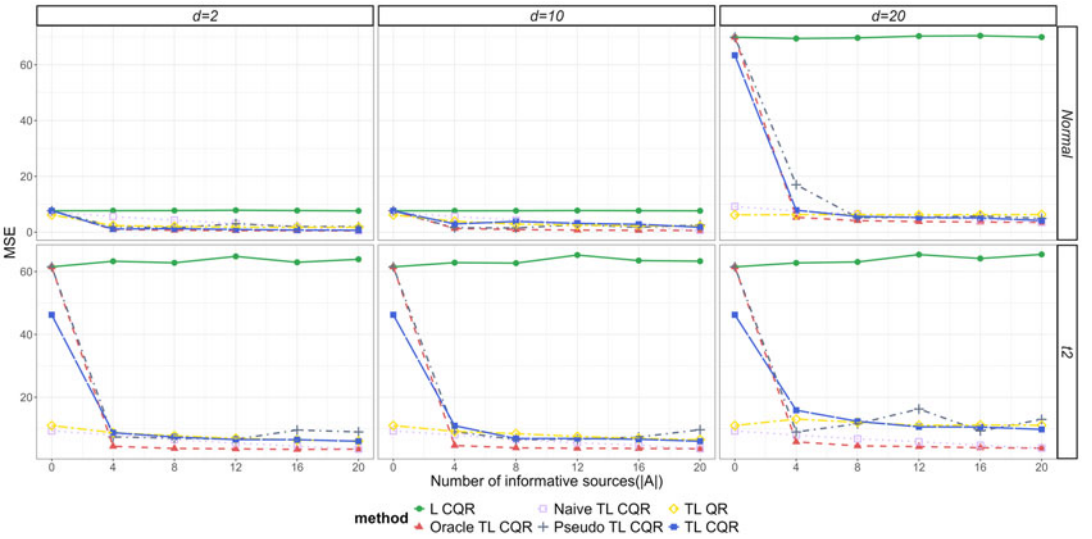


Figure 5. *Under the condition of different similarity d with heterogeneous design, the average MSE between the number of transferable sources.*

a negative transfer. With the increase of $|\mathcal{A}|$, the effective information of the sources increases, and the loss of the Naive TL CQR method is greater than that of the L CQR method. The performance of the Pseudo TL CQR is better than that of the Naive TL CQR estimation due to a priori knowledge of the size of $|\mathcal{A}|$. The estimation accuracy of TL CQR method is close to that of Oracle method, which proves the effectiveness of informative set detection algorithm. In the case of $d = 20$, the similarity between the source datasets and the target dataset is increased, and the informative set detection is more difficult. The TL CQR estimation shows strong robustness, and its loss value is lower than that of Naive TL CQR, indicating that it can effectively use the informative set for positive transfer and avoid being misled by false information.

Figure 5 illustrates the trend of MSE of different estimation methods with the informative set $|\mathcal{A}|$ when d is fixed. Compared with L CQR and other methods, it can be found that transfer learning can significantly reduce the error. Under the setting of heavy-tailed error and large similarity, our proposed algorithm still maintains robustness. In general, the performance of all methods under the normal error condition is better than that under the t_2 error condition, and the effect of each method is more significant under the smaller d value. As long as the d value is not too large, different transfer learning methods usually perform better in the case of a large number of sources. Compared with the loss value, the MSE value is less different, and the results also support the conclusion of Fig. 4.

6.0 Real data application

According to the accident statistics report published by Airbus, although the landing phase accounts for only 1% of the flight duration, the proportion of aircraft damage accidents at this stage exceeds 50% [18]. Based on the real-time data obtained by the QAR, the machine learning method can realise the quantitative risk assessment and accident prevention during the landing phase. For example, Wu et al. [19] studied the relationship between landing distance and aircraft performance parameters based on the linear regression model, and Atasoy and Cetek [20] proposed a high-accuracy flight fuel model based on QAR data.

We use the QAR dataset derived from the supplementary materials of Jin et al. [9] to evaluate the performance of the proposed algorithm. The dataset contains three types of aircraft, namely Airbus A320 (A320), Airbus A380 (A380) and Boeing 737 (B737). There are a total of 102 variable values,

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

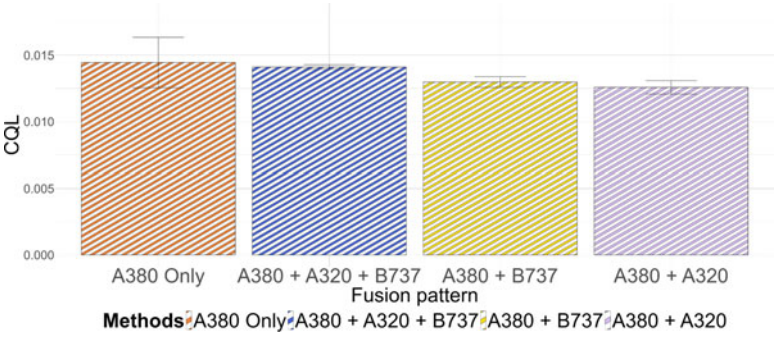


Figure 6. *Composite quantile loss on A380 as the target with all different fusion combinations.*

which record the values of each aircraft performance parameter and their descriptive statistics at different stages. The number of flight records for A380 is 1,218, which is much smaller than that of B737 and A320, which exceed 680,000. This imbalance in sample sizes presents a natural scenario for applying transfer learning to augment the limited A380 data with information from the larger sources (B737 and A320). The maximum vertical acceleration (MVA) increases rapidly when the aircraft undergoes a severe dive or a large deceleration, and the aircraft bears a large vertical acceleration load. Based on this, this paper selects MVA as the response variable [21, 22] quantitatively analysed the relationship between the various performance parameters in the QAR data and the MVA to understand the aircraft load status.

Safety risks often reside in the tail of the distribution, hence the τ values are set relatively high. In the case of $\tau = c(0.85, 0.9, 0.95)$, A380 is used as the target, and A320 and B737 are used as the sources. The target data is randomly divided into 80% training set and 20% test set, and then transfer learning with informative set detection is performed. Transfer learning methods under different informative sets were compared, and the mean and standard deviation of out-of-sample composite quantile loss were calculated respectively.

Figure 6 shows the results of using different sources with A380 as the target. L CQR only uses A380 data without considering other source information. Naive TL CQR uses all sources without informative set detection. The TL CQR algorithm proposed in this paper only retains the A320 source as the informative source after informative set detection.

The results show that using only the A320 data more similar to the target data can make the transfer learning model more accurate and avoid the noise introduced by the B737 data. The data distribution of B737 may be quite different from that of A380. The mixing of B737 data into the training process leads to poor matching between the source and the target of the transfer learning algorithm, which reduces the performance of the model. These findings validate the effectiveness of the informative set detection algorithm in filtering out non-transferable sources. The underlying engineering context further explains these results. Both the A320 and A380 are Airbus's large civilian aircraft, using the same aircraft control system, similar architecture and parts, so the A320's flight data is more likely to help analyse the A380 data. In contrast, the B737 is Boeing's medium-sized airliner, which is different from the A380 in terms of shape, control system, fuel setting and so on. Comparing Naive TL CQR and TL CQR, it can be found that even if B737 has a large amount of data, the modeling of A380 by the source still shows a negative transfer.

Building on the optimal TL CQR model with A320 as the informative source, we proceed to analyse the factors influencing A380 hard landing risk. The features are divided into several classifications, namely speed, fuel, distance, time, acceleration (non-vertical), weight, attitude, wind speed and temperature. The detailed type of each variable is shown in the 'type' column of the Tables A3 and A4.

In order to quantify the impact of different types on the model results, this paper first sums the absolute values of the estimated coefficients of the variables contained in various types of features, and then

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

Table 2. *Features contribution from A380 TL CQR with* $\tau = \mathrm{c}(0.85, 0.9, 0.95)$

Type	Coefficient	Proportion
Speed	0.090	0.372
Acceleration	0.069	0.284
Attitude	0.038	0.155
Time	0.018	0.074
Distance	0.009	0.036
Fuel	0.008	0.033
Wind	0.007	0.027
Temperature	0.003	0.012
Weight	0.002	0.008

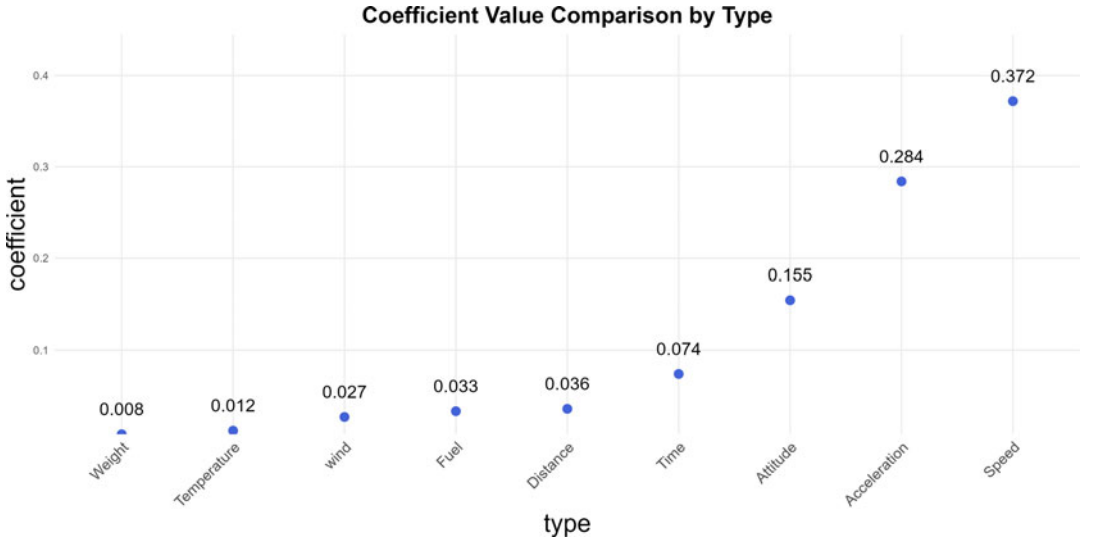


Figure 7. *Plot for features contribution from A380 TL CQR with* $\tau = c(0.85, 0.9, 0.95)$.

converts them into relative data with a sum of 1 by normalisation. Finally, the relative contribution value of the type is visualised in Fig. 7 and Table 2. The results show that the controllable factors such as speed, acceleration and attitude have a greater impact on the hard landing risk, while the external conditions such as time, distance, fuel and wind speed have less impact. In summary, the focus of reducing the risk of hard landing is on the optimisation and management of 'controllable factors'. Reflected in real life, it is necessary to further improve the refined management mechanism of the aircraft control system, improve aircraft performance at the technical level, and strengthen the training of pilots. Reduce the adverse effects of controllable factors from two aspects of human operation and equipment control, thus effectively reducing the risk of the landing phase.

7.0 Summary

In the field of flight safety monitoring, traditional machine learning methods rely on the condition of error distribution, which has limitations when applied to QAR data. Aiming at this application difficulty, we develop a transfer learning framework based on composite quantile regression. By constructing

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

pseudo response variables, the algorithm transforms the non-smooth composite quantile loss into a linear loss form. A large number of numerical simulations show that the algorithm significantly improves the robustness and applicability of the model. In addition, for the case where the informative set in the actual data is unknown, an informative set detection algorithm based on cross-validation is proposed, which can effectively detect the source that can be used for migration. In the QAR data, we establish a composite quantile regression model for the landing risk of the A380 aircraft, and other types of aircraft data can provide additional useful information.

Follow-up research can be extended to other aspects. For example, Li et al. [23] considered the problem of data privacy, and an extension direction is to study the privacy protection problem in combination with distributed computing, so as to apply to the multi-source situation. The research focus of the example part is QAR data. In the future, it can also be extended to medical, financial and other fields with heterogeneity and heavy-tailed problems to verify the domain applicability of the algorithm.

Acknowledgements. The authors sincerely thank the editor and the anonymous reviewers for their insightful comments and constructive suggestions, which greatly improved the quality of this manuscript. This research is supported by National Natural Science Foundation of China (11771215), Young Talent Fund of Association for Science and Technology in Shaanxi, China (20260518).

References

[1] Wang, L., Wu, C. and Sun, R. An analysis of flight quick access recorder (QAR) data and its applications in preventing landing incidents, *Reliab. Eng. Syst. Saf.*, 2014, **127**, pp 86–96.
[2] Chen, X., Liu, W., Mao, X., et al. Distributed high dimensional regression under a quantile loss function, *J. Mach. Learn. Res.*, 2020, **21**, pp 1–43.
[3] Zhao J., Rong C., Dang X., et al. Qar data imputation using generative adversarial network with self-attention mechanism, *Big Data Mining and Analytics*, 2023, **7**, (1), pp 12–28.
[4] Tibshirani, R. Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B (Methodol.)*, 1996, **58**, (1), pp 267–288.
[5] Fan, J. and Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Am. Stat. Assoc.*, 2001, **96**, (456), pp 1348–1360.
[6] Li, S., et al. Estimation and inference for high-dimensional generalized linear models with knowledge transfer, *J. Am. Stat. Assoc.*, 2024, **119**, (546), pp 1274–1285.
[7] Li, S., Cai, T.T. and Li, H. Transfer learning for high-dimensional linear regression: Prediction, estimation, and minimax optimality, *J. R. Stat. Soc. Ser. B*, 2022, **84**, pp 149–173.
[8] Tian, Y. and Feng, Y. Transfer learning under high-dimensional generalized linear models, *J. Am. Stat. Assoc.*, 2023, **118**, pp 2684–2697.
[9] Jin, J., Yan, J., Aseltine, R.H., et al. Transfer learning with large-scale quantile regression, *Technometrics*, 2024, **66**, (3), pp 381–393.
[10] Fan, J., Li, Q. and Wang, Y. Estimation of high dimensional mean regression in the absence of symmetry and light tail conditions, *J. R. Stat. Soc. Ser. B*, 2017, **79**, (1), pp 247–265.
[11] Sun, Q., Zhou, W.X. and Fan, J. Adaptive Huber regression, *J. Am. Stat. Assoc.*, 2020, **115**, (529), pp 254–265.
[12] Koenker, R. and Hallock, K.F. Quantile regression, *J. Econ. Perspect.*, 2001, **15**, (4), pp 143–156.
[13] Koenker, R. and Bassett, G. Regression quantiles, *Econometrica*, 1978, **46**, (1), pp 33–50.
[14] Zou, H. and Yuan, M. Composite quantile regression and the oracle model selection theory, *Ann. Stat.*, 2008, **36**, (3), pp 1108–1126.
[15] Qiao, S., He, Y., Zhou, W., et al. Transfer learning for high-dimensional quantile regression with statistical guarantee, *Transact. Mach. Learn. Res*, 2024. https://openreview.net/forum?id=d3xwrfAG4V
[16] Li, J. and Song, Y. Transfer learning with high dimensional composite quantile regression, *J. Stat. Comput. Simul.*, 2024, **94**, (10), pp 2273–2290.
[17] Efron, B. Bootstrap methods: another look at the jackknife, *Ann. Stat.*, 1979, **7**, (1), pp 1–26.
[18] Airbus. A Statistical Analysis of Commercial Aviation Accidents 1958-2022, X00D17008863 Issue 7[R]. France: AIRBUS, 27, 2023.
[19] Wu, Z., Song, W., Qi, Y., et al. A mission fuel performance model based on hybrid flight physics and QAR data, *Aerosp. Syst.*, 2024, **2025**, (8), pp 83–103.
[20] Atasoy, V.E. and Cetek, C. Enhanced cruise range prediction for narrow-body turbofan commercial aircraft based on QAR data, *Aeronaut. J.*, 2020, **125**, (1286), pp 672–701.
[21] Huang, J., Wang, M. and Wu, Y. Estimation and Inference for Transfer Learning with High-Dimensional Quantile Regression. *arXiv preprint*, 2022. arXiv:2211.14578.
[22] Gu, T., Han, Y., Duan, R., et al. Robust angle-based transfer learning in high dimensions, *J. R. Stat. Soc. Ser. B*, 2024, **2024**, (87), pp 723–745.

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

[23] Li, B.S., Cai, T. and Duan, R. Targeting underrepresented populations in precision medicine: a federated transfer learning approach, *Ann. Appl. Stat.*, 2023, **17**, (4), pp 2970–2992.
[24] Raskutti, G., Wainwright, M.J. and Yu, B. Restricted eigenvalue properties for correlated gaussian designs, *J. Mach. Learn. Res.*, 2010, **11**, pp 2241–2259.
[25] Boucheron, S., Lugosi, G. and Bousquet, O. Concentration inequalities, in Bousquet, O., von Luxburg, U. and Rätsch, G. (Eds), *Summer School on Machine Learning*, Springer, 2003, Berlin, pp. 208–240.
[26] Zhou, S. Restricted eigenvalue conditions on subgaussian random matrices, 2009. *arXiv preprint* arXiv:0912.4045.

Appendix A. Proofs of theorem

The theoretical proof of the convergence rate for CQR is a natural extension of that for the QR case. Consequently, the proof of Theorem 1 follows a structure analogous to that of Theorem 1 in Jin et al. [9]. The principal technical difficulty lies in dealing with the quantile loss, which we address by introducing a pseudo-response variable following the approach of Chen et al. [2]. To maintain a self-contained presentation, we include here the necessary background results from the literature, adapted to our framework. For notational simplicity, we keep the same symbol for constants that appear in convergence rates or probabilistic bounds; for instance, c_1 consistently denotes a generic positive constant in expressions such as $1-\exp\{-c_1 n_0\}$. To facilitate the proof, we first establish the convergence rate of the quantile regression estimator under the setting of a single quantile level, and then extend the result to the composite quantile regression estimator with multiple quantile points.

Lemma 1 (Theorem 1 in Jin et al. [9]). *Suppose Conditions 1–6 hold true. We take* $\lambda_1 \asymp \sqrt{\log(p \vee \max_{k\in\mathcal{A}\cup\{0\}} n_k)/(n_{\mathcal{A}}+n_0)}$ *and* $\lambda_0 \asymp \lambda_2 \asymp \sqrt{\log(p\vee n_0)/n_0}$, *then with*

$$\xi_1\left(n_0, n_{\mathcal{A}}, \max_{k\in\mathcal{A}\cup\{0\}} n_k, s_0, p\right) = \frac{s_0 \log\left(p \vee \max_{k\in\mathcal{A}\cup\{0\}} n_k\right)}{n_{\mathcal{A}}+n_0} + \frac{s_0 \log(p\vee n_0)}{n_0} \wedge C_{\boldsymbol{\Sigma}} d\sqrt{\frac{\log(p\vee n_0)}{n_0}} \wedge (C_{\boldsymbol{\Sigma}} d)^2,$$

we have

$$\mathbb{P}\left(\frac{1}{n_0}\left\|\boldsymbol{X}^{(0)}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(0)})\right\|^2 \vee \left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(0)}\right\|^2 \lesssim \xi_1\left(n_0, n_{\mathcal{A}}, \max_{k\in\mathcal{A}\cup\{0\}} n_k, s_0, p\right)\right)$$
$$\geq 1 - C_1\exp\left(-C_2 \min_{k\in\mathcal{A}\cup\{0\}} n_k\right) - C_3\left(\min_{k\in\mathcal{A}\cup\{0\}} n_k\right)^{-C_4 \underline{s}_{\mathcal{A}}} - \exp\left(-C_5 \log p\right),$$

for some positive constants $C_1,\ldots, C_5$, *where* $\underline{s}_{\mathcal{A}} = \min_{k\in\mathcal{A}\cup\{0\}} s_k$.

Lemma 2 (Proposition 11 in Chen et al. [2]). *Denote* $\widehat{\boldsymbol{\Sigma}}^{(k)} = \sum_{i=1}^{n_k} \boldsymbol{X}^{(k)\top}\boldsymbol{X}^{(k)}/n_k$ *for* $k=0,\ldots,K$. *Recall* $\tilde{\boldsymbol{y}}^{(k)}$ *is the surrogate response. Under Conditions 1, 2, 4, 5 and 6, for any* $k=0,\ldots,K$, *we have*

$$\left\|\frac{1}{n_k}\boldsymbol{X}^{(k)\top}\tilde{\boldsymbol{y}}^{(k)} - \widehat{\boldsymbol{\Sigma}}^{(k)}\boldsymbol{\beta}^{(k)}\right\|_\infty \lesssim \sqrt{\frac{\log(p\vee n_k)}{n_k}} + \frac{s_k \log(p\vee n_k)}{n_k}.$$

with probability at least $1 - c_1 n_k^{-c_2 s_k} - c_3\exp(-c_4 n_k)$ *for positive constants* c_1 *to* c_4.

Remark 1. The necessity of Condition 1 can be seen from the following argument. Applying the proof technique of Proposition 11 in Chen et al. [2] gives

$$\left\|\frac{1}{n_k}\boldsymbol{X}^{(k)\top}\tilde{\boldsymbol{y}}^{(k)} - \widehat{\boldsymbol{\Sigma}}^{(k)}\boldsymbol{\beta}^{(k)}\right\|_\infty \lesssim \sqrt{\frac{\log p}{n_k}} + \sqrt{\frac{s_k z_n \log n_k}{n_k}} + z_n^2,$$

where $z_n = \sqrt{s_k \log n_k/n_k}$ is the convergence rate of the initial estimator of $\boldsymbol{\beta}_k$. Consequently, to attain the rate required in Lemma 1 we must have $s_k z_n = O(1)$, i.e. $\sqrt{s_k^3 \log n_k/n_k} = O(1)$. This forces $s_k = O\left(n_k^b\right)$

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

with $b \in (0, 1/3)$, which is precisely what Condition 1 states.

$$\widehat{f}^{(k)}(0) = \frac{1}{n_k h^{(k)}} \sum_{i=1}^{n_k} K_h \left(\frac{y_i^{(k)} - \boldsymbol{x}_i^{(k)\top} \widehat{\boldsymbol{\beta}}^{(k)}}{h^{(k)}} \right), \tag{16}$$

Namely, with $\widehat{f}^{(k)}(0)$ defined in 16, Condition 4 contributes to obtain

$$\left| \widehat{f}^{(k)}(0) - f^{(k)}(0) \right| = O_P \left(\sqrt{\frac{s_k \log n_k}{n_k h^{(k)}}} + \sqrt{\frac{s_k \log n_k}{n_k}} \right).$$

Specifically, denote

$$\mathcal{B}n, h^{(k)(k)}(\boldsymbol{\beta}) = \frac{1}{nh^{(k)}} \sum_{i=1}^{n_k} K_h \left(\frac{y_i^{(k)} - \boldsymbol{x}_{i,S_k}^{(k)\top} \boldsymbol{\beta}_{S_k}}{h^{(k)}} \right),$$

in which $\boldsymbol{x}_i^{(k)\top} \widehat{\boldsymbol{\beta}}^{(k)}$ is replaced by $\boldsymbol{x}_{i,S_k}^{(k)\top} \boldsymbol{\beta}_{S_k}$. If $\hat{S}_k \subseteq S_k$ holds true with high probability (for which we would need the irrepresentable condition), then we have $\widehat{f}^{(k)}(0) = \mathcal{B}n, h^{(k)(k)}(\hat{\boldsymbol{\beta}}^{(k)})$ with high probability and can bound $|\widehat{f}^{(k)}(0) - f^{(k)}(0)|$ by

$$\left| \widehat{f}^{(k)}(0) - f^{(k)}(0) \right| \leq \sup_{\left\| \boldsymbol{\beta}_{S_k} - \boldsymbol{\beta}_{S_k}^{(k)} \right\| \leq z_n} \left| \mathcal{B}n, h^{(k)(k)}(\boldsymbol{\beta}) - f^{(k)}(0) \right|,$$

where $z_n = \sqrt{s_k \log n_k / n_k}$ is the convergence rate of the initial estimator of $\boldsymbol{\beta}_k$ provided by Sun et al. [11].

As demonstrated in Chen et al. [2], for every coordinate i belonging to the support set S_k, we partition the interval $[\boldsymbol{\beta}_i^{(k)} - z_n, \boldsymbol{\beta}_{\mathrm{i}}^{(\mathrm{k})} + \mathrm{z_n}]$ into n^M small sub-intervals of length $2z_n/n^M$. In this way, we construct a collection of points $\{\boldsymbol{\beta}_j, 1 \leq j \leq q_n\}$ in $\mathbb{R}^p$ whose cardinality satisfies $q_n \leq n^{Ms_k}$, and such that for any $\boldsymbol{\beta}$ inside the ball $\|\boldsymbol{\beta}_{S_k} - \boldsymbol{\beta}_{S_k}^{(k)}\| \leq z_n$ there exists an index j $(1 \leq j \leq q_n)$ with $\|\boldsymbol{\beta}_{S_k} - \boldsymbol{\beta}_{j,S_k}\| \leq 2\sqrt{s_k} \frac{\mathrm{z_n}}{\mathrm{n^M}}$, $\|\boldsymbol{\beta}_{j,S_k} - \boldsymbol{\beta}_{S_k}^{(k)}\| \leq z_n$. Consequently, we can relate $\sup_{\|\boldsymbol{\beta}_{S_k} - \boldsymbol{\beta}_{S_k}^{(k)}\| \leq z_n} |\mathcal{B}_{n,h^{(k)}}^{(k)}(\boldsymbol{\beta}) - f^{(k)}(0)|$ to $\sup_{1 \leq j \leq n^{Ms_k}} \left| \mathcal{B}_{n,h^{(k)}}^{(k)}(\boldsymbol{\beta}_j) - f^{(k)}(0) \right|$, and the latter can be bounded using exponential inequalities. This finally yields the required bound on the error of estimating the density functions.

Lemma 3 (Theorem 1 in Raskutti et al. [24]). *Denote* $\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}} = \sum_{k \in \mathcal{A} \cup \{0\}} \alpha_k \widehat{\boldsymbol{\Sigma}}^{(k)}$, $\boldsymbol{\Sigma}^{\mathcal{A}} = \sum_{k \in \mathcal{A} \cup \{0\}} \alpha_k \boldsymbol{\Sigma}^{(k)}$ *and* $\rho^2(\boldsymbol{\Sigma}) = \max_{j=1,\ldots,p} \Sigma_{jj}$, *where* $\alpha_k = \frac{n_k}{n_{\mathcal{A}} + n_0}$. *Under Condition 2, we have*

$$u^\top \widehat{\boldsymbol{\Sigma}}^{\mathcal{A}} u \geq \frac{1}{32} u^\top \boldsymbol{\Sigma}^{\mathcal{A}} u - 81 \rho^2 \left(\boldsymbol{\Sigma}^{\mathcal{A}} \right) \frac{\log p}{n_{\mathcal{A}} + n_0} \|u\|_1^2$$

for any $u \in \mathbb{R}^p$ *with probability at least* $1 - \exp(-c_1(n_{\mathcal{A}} + n_0))$ *and*

$$u^\top \widehat{\boldsymbol{\Sigma}}^{(0)} u \geq \frac{1}{32} u^\top \boldsymbol{\Sigma}^{(0)} u - 81 \rho^2 \left(\boldsymbol{\Sigma}^{(0)} \right) \frac{\log p}{n_0} \|u\|_1^2$$

for any $u \in \mathbb{R}^p$ *with probability at least* $1 - \exp(-c_1 n_0)$ *for positive constant* c_1.

Lemma 4 (Bernstein's inequality (Theorem 1.10 of Boucheron et al. [25])). *A random variable* X *is said to follow a sub-exponential distribution with parameter* λ *(written* $X \sim \mathrm{subE}(\lambda)$*) if* $\mathbb{E}[X] = 0$, *and its moment generating function satisfies*

$$\mathbb{E}\left[e^{sX}\right] \leq \exp\left(\frac{1}{2} s^2 \lambda^2\right), \qquad \forall |s| \leq \frac{1}{\lambda}.$$

If $X_1, \ldots, X_n$ *are independent random variables with* $\mathbb{E}X_i = \mu_i$ *and* $X_i - \mu_i \sim subE(\lambda)$, *then for any* $t > 0$ *we have*

$$\mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^{n} (X_i - \mu_i) \right| \geq t \right) \leq \exp\left(-\frac{n}{2} \min\left\{ \frac{t^2}{\lambda^2}, \frac{t}{\lambda} \right\} \right).$$

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

We now outline the proof of Lemma 1, with the detailed derivations to follow. Let $\boldsymbol{\beta}^{\mathcal{A}}$ denote the probabilistic limit of $\widehat{\boldsymbol{\beta}}^{\mathcal{A}}$ in Step 3 of the oracle QR algorithm [9]. It solves the moment equation

$$\mathbb{E}\left[\sum_{k\in\mathcal{A}\cup\{0\}}\left(\boldsymbol{X}^{(k)}\right)^{\top}\left(\tilde{y}^{(k)}-\boldsymbol{X}^{(k)}\boldsymbol{\beta}^{\mathcal{A}}\right)\right]=0, \tag{17}$$

where $\tilde{y}^{(k)}=\boldsymbol{x}^{(k)\top}\boldsymbol{\beta}^{(k)}-\left\{f^{(k)}\left(0\right)\right\}^{-1}\left\{I(y^{(k)}-\boldsymbol{x}^{(k)\top}\boldsymbol{\beta}^{(k)}\le 0)-\tau\right\}$. Thus, $\boldsymbol{\beta}^{\mathcal{A}}$ has the explicit form as $\boldsymbol{\beta}^{\mathcal{A}}=\boldsymbol{\beta}^{(0)}+\boldsymbol{\delta}^{\mathcal{A}}$, where $\boldsymbol{\delta}^{\mathcal{A}}=(\boldsymbol{\Sigma}^{\mathcal{A}})^{-1}\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\boldsymbol{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}$. It implies $\boldsymbol{\beta}^{\mathcal{A}}$ is a pooled version of $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\beta}^{(k)}$ for all $k\in\mathcal{A}$ such that $\boldsymbol{\beta}^{\mathcal{A}}=(\boldsymbol{\Sigma}^{\mathcal{A}})^{-1}\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\boldsymbol{\Sigma}^{(k)}\boldsymbol{\beta}^{(k)}$.

Define

$$B_1=\left\{\frac{1}{n_{\mathcal{A}}+n_0}\left\|\sum_{k\in\mathcal{A}\cup\{0\}}\boldsymbol{X}^{(k)\top}(\tilde{\boldsymbol{y}}^{(k)}-\boldsymbol{X}^{(k)}\boldsymbol{\beta}^{\mathcal{A}})\right\|_{\infty}\le\frac{\lambda_1}{2}\right\},$$
$$B_2=\left\{\frac{1}{n_0}\left\|\boldsymbol{X}^{(0)\top}\left(\tilde{y}^{(0)}-\boldsymbol{X}^{(0)}\boldsymbol{\beta}^{(0)}\right)\right\|_{\infty}\le\frac{\lambda_2}{2}\right\}.$$

We will show the following statements:

- (1) Denote $\widehat{\boldsymbol{u}}^{\mathcal{A}}=\widehat{\boldsymbol{\beta}}^{\mathcal{A}}-\boldsymbol{\beta}^{\mathcal{A}}$. On the event of B_1, we have
$$\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}}\widehat{\boldsymbol{u}}^{\mathcal{A}}\vee\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2\lesssim s_0\lambda_1^2+\lambda_1C_{\boldsymbol{\Sigma}}d,$$
$$\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|_1\lesssim s_0\lambda_1+C_{\boldsymbol{\Sigma}}d$$
with probability at least $1-\exp(-c_1(n_{\mathcal{A}}+n_0))$.
- (2) Denote $\widehat{\boldsymbol{v}}^{\mathcal{A}}=\widehat{\boldsymbol{\delta}}^{\mathcal{A}}-\boldsymbol{\delta}^{\mathcal{A}}$. On the event of B_2, we have
$$\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}}\vee\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|^2\lesssim s_0\lambda_1^2+\lambda_1C_{\boldsymbol{\Sigma}}d,$$
$$\widehat{\boldsymbol{v}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{v}}^{\mathcal{A}}\lesssim\lambda_2C_{\boldsymbol{\Sigma}}d,$$
$$\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|^2\lesssim\lambda_2C_{\boldsymbol{\Sigma}}d\wedge(C_{\boldsymbol{\Sigma}}d)^2$$
with probability at least $1-3\exp(-c_1n_0)$.

We then show that $\mathbb{P}(B_1\cap B_2)\to 1$ as $p\to\infty$ and $\min_{k\in\mathcal{A}\cup\{0\}}n_k\to\infty$, which is the minimum sample size across the informative sources and the target. Finally, the proof of Lemma 1 is proved by appealing to Lemma 3 together with the observation that

$$\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{(0)}\right\|=\left\|\left(\widehat{\boldsymbol{\beta}}^{\mathcal{A}}-\boldsymbol{\beta}^{\mathcal{A}}\right)-\left(\widehat{\boldsymbol{\delta}}^{\mathcal{A}}-\boldsymbol{\delta}^{\mathcal{A}}\right)\right\|\le\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\|+\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\|.$$

We now proceed to prove statement (1). The following oracle inequality is identical to (A.1) in the supplementary material of Li et al. [7]. Expanding the loss function in Step 3 of the oracle QR algorithm in Jin et al. [9] via Taylor expansion gives

$$\begin{aligned}
&\frac{1}{2}\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}}\widehat{\boldsymbol{u}}^{\mathcal{A}}\\
&\le\lambda_1\left\|\boldsymbol{\beta}^{\mathcal{A}}\right\|_1-\lambda_1\left\|\widehat{\boldsymbol{\beta}}^{\mathcal{A}}\right\|_1+\left|\frac{\widehat{\boldsymbol{u}}^{\mathcal{A}\top}}{(n_{\mathcal{A}}+n_0)}\sum_{k\in\mathcal{A}\cup\{0\}}\boldsymbol{X}^{(k)\top}\left(\tilde{\boldsymbol{y}}^{(k)}-\boldsymbol{X}^{(k)}\boldsymbol{\beta}^{\mathcal{A}}\right)\right|\\
&\le\lambda_1\left\|\boldsymbol{\beta}^{\mathcal{A}}\right\|_1-\lambda_1\left\|\widehat{\boldsymbol{\beta}}^{\mathcal{A}}\right\|_1+\frac{\lambda_1}{2}\left\|\hat{\boldsymbol{u}}^{\mathcal{A}}\right\|_1\\
&\le\frac{3}{2}\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0}\right\|_1+\lambda_1\left\|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\right\|_1-\lambda_1\left\|\widehat{\boldsymbol{\beta}}^{\mathcal{A}}_{S_0^c}\right\|_1+\frac{1}{2}\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0^c}\right\|_1\\
&\le\frac{3}{2}\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0}\right\|_1+\frac{3}{2}\lambda_1\left\|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\right\|_1-\frac{1}{2}\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0^c}\right\|_1.
\end{aligned}$$

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

(1) If $\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0}\|_1 \geq \|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\|_1$, with a positive constant c_4, we can further have

$$\begin{aligned}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2 - \frac{\log p}{n_{\mathcal{A}}+n_0}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|_1^2 &\lesssim \frac{1}{2}\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}}\widehat{\boldsymbol{u}}^{\mathcal{A}} \\ &\leq 3\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0}\right\|_1 - \frac{1}{2}\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0^c}\right\|_1 \\ &\leq 3\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0}\right\|_1 \\ &\leq 3\lambda_1\sqrt{s_0}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0}\right\| \\ &\leq 3\lambda_1\sqrt{s_0}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|,\end{aligned}$$

where the first line holds true with probability at least $1-\exp(-c_1(n_{\mathcal{A}}+n_0))$ due to Lemma 3 and Condition 4. It further leads to

$$\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2 \lesssim s_0\lambda_1^2, \tag{18}$$

$$\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|_1 \lesssim s_0\lambda_1 \tag{19}$$

by utilising $s_0 \log p/(n_{\mathcal{A}}+n_0) = o(1)$ derived from Condition 3.

(2) If $\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0}\|_1 \leq \|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\|_1$, we have

$$0 \leq \frac{1}{2}\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}}\widehat{\boldsymbol{u}}^{\mathcal{A}} \leq 3\lambda_1\left\|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\right\|_1 - \frac{1}{2}\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0^c}\right\|_1.$$

It follows that

$$\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0^c}\right\|_1 \leq 6\left\|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\right\|_1,$$

and

$$\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|_1 = \left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0}\right\|_1 + \left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0^c}\right\|_1 \leq 7\left\|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\right\|_1 \leq 7\left\|\boldsymbol{\delta}^{\mathcal{A}}_{S_0^c}\right\|_1 \leq 7C_{\boldsymbol{\Sigma}}d. \tag{20}$$

This implies $\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\| \leq 7C_{\boldsymbol{\Sigma}}d$ by the ℓ_1-ℓ_2 inequality. By Lemma 3, we have

$$\begin{aligned}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2 - \frac{\log p}{n_{\mathcal{A}}+n_0}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|_1^2 &\lesssim \frac{1}{2}\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}}\widehat{\boldsymbol{u}}^{\mathcal{A}} \\ &\leq 3\lambda_1\left\|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\right\|_1 - \frac{1}{2}\lambda_1\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}_{S_0^c}\right\|_1 \\ &\leq 3\lambda_1\left\|\boldsymbol{\beta}^{\mathcal{A}}_{S_0^c}\right\|_1 \\ &\leq 3\lambda_1 C_{\boldsymbol{\Sigma}}d\end{aligned}$$

with probability at least $1-\exp(-c_1(n_{\mathcal{A}}+n_0))$. Noticing that $C_{\boldsymbol{\Sigma}}d\sqrt{\log p/(n_A+n_0)} = o(1)$ as derived from Condition 3, we get

$$\begin{aligned}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2 &\leq \frac{\log p}{n_{\mathcal{A}}+n_0}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|_1^2 + 3\lambda_1 C_{\boldsymbol{\Sigma}}d \\ &\leq \frac{\sqrt{s_0}\log p}{n_{\mathcal{A}}+n_0}\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2 + 3\lambda_1 C_{\boldsymbol{\Sigma}}d,\end{aligned}$$

and then we have

$$\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2 \lesssim \lambda_1 C_{\boldsymbol{\Sigma}}d. \tag{21}$$

Statement (1) then follows by combining the results from Equations (18), (19), (20) and (21) together with Lemma 3.

We now turn to statement (2). Conditional on the event B_2, the following oracle inequality holds, which is identical to the one appearing in the proofs of Theorems 1 and 4 in Li et al. [7]. Expanding the

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

loss of $\widehat{\boldsymbol{\delta}}^{\mathcal{A}}$ via Taylor expansion gives

$$\begin{aligned}
&\frac{1}{2}\widehat{\boldsymbol{v}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{v}}^{\mathcal{A}} \\
&\le \lambda_2\left\|\boldsymbol{\delta}^{\mathcal{A}}\right\|_1 - \lambda_2\left\|\widehat{\boldsymbol{\delta}}^{\mathcal{A}}\right\|_1 + \left|\frac{\widehat{\boldsymbol{v}}^{\mathcal{A}\top}}{n_0}\boldsymbol{X}^{(0)\top}\left\{\tilde{\boldsymbol{y}}^{(0)} - \boldsymbol{X}^{(0)}\left(\widehat{\boldsymbol{\beta}}^{\mathcal{A}} - \boldsymbol{\delta}^{\mathcal{A}}\right)\right\}\right| \\
&= \lambda_2\left\|\boldsymbol{\delta}^{\mathcal{A}}\right\|_1 - \lambda_2\left\|\widehat{\boldsymbol{\delta}}^{\mathcal{A}}\right\|_1 + \left|\frac{\widehat{\boldsymbol{v}}^{\mathcal{A}\top}}{n_0}\boldsymbol{X}^{(0)\top}\left\{\tilde{\boldsymbol{y}}^{(0)} - \boldsymbol{X}^{(0)}\boldsymbol{\beta}^{(0)} - \boldsymbol{X}^{(0)}\left(\widehat{\boldsymbol{\beta}}^{\mathcal{A}} - \boldsymbol{\delta}^{\mathcal{A}} - \boldsymbol{\beta}^{(0)}\right)\right\}\right| \\
&= \lambda_2\left\|\boldsymbol{\delta}^{\mathcal{A}}\right\|_1 - \lambda_2\left\|\widehat{\boldsymbol{\delta}}^{\mathcal{A}}\right\|_1 + \left|\frac{\boldsymbol{v}^{\mathcal{A}\top}}{n_0}\boldsymbol{X}^{(0)\top}\left\{\left(\tilde{\boldsymbol{y}}^{(0)} - \boldsymbol{X}^{(0)}\boldsymbol{\beta}^{(0)}\right) - \boldsymbol{X}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\}\right| \\
&\le \lambda_2\left\|\boldsymbol{\delta}^{\mathcal{A}}\right\|_1 - \lambda_2\left\|\widehat{\boldsymbol{\delta}}^{\mathcal{A}}\right\|_1 + \frac{\lambda_2}{2}\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|_1 + \widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}} + \frac{1}{4}\widehat{\boldsymbol{v}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{v}}^{\mathcal{A}}.
\end{aligned}$$

Hence, by noticing $\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\|_1 = \|\widehat{\boldsymbol{\delta}}^{\mathcal{A}} - \boldsymbol{\delta}^{\mathcal{A}}\|_1 \ge \|\widehat{\boldsymbol{\delta}}^{\mathcal{A}}\|_1 - \|\boldsymbol{\delta}^{\mathcal{A}}\|_1$, we have

$$\frac{1}{4}\widehat{\boldsymbol{v}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{v}}^{\mathcal{A}} \le 2\lambda_2\left\|\boldsymbol{\delta}^{\mathcal{A}}\right\|_1 - \frac{\lambda_2}{2}\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|_1 + \widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}}.$$

(1) If $\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}} \le \lambda_2\|\boldsymbol{\delta}^{\mathcal{A}}\|_1$, then

$$\begin{aligned}
0 \le \frac{1}{4}\widehat{\boldsymbol{v}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{v}}^{\mathcal{A}} &\le 3\lambda_2\left\|\boldsymbol{\delta}^{\mathcal{A}}\right\|_1 - \frac{\lambda_2}{2}\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|_1 \\
&\le 3\lambda_2\left\|\boldsymbol{\delta}^{\mathcal{A}}\right\|_1 \\
&\le 3\lambda_2 C_{\boldsymbol{\Sigma}} d,
\end{aligned} \tag{22}$$

which further leads to $\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\|_1 \le 6\|\boldsymbol{\delta}^{\mathcal{A}}\|_1 \le 6C_{\boldsymbol{\Sigma}}d$ and

$$\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\| \le \left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|_1 \le 6C_{\boldsymbol{\Sigma}}d. \tag{23}$$

Meanwhile, from Lemma 3, we have

$$\begin{aligned}
\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|^2 - \frac{\log p}{n_0}\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|_1^2 \lesssim \frac{1}{4}\widehat{\boldsymbol{v}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{v}}^{\mathcal{A}} &\le 3\lambda_2\left\|\boldsymbol{\delta}^{\mathcal{A}}\right\|_1 \\
&\le 3\lambda_2 C_{\boldsymbol{\Sigma}} d,
\end{aligned}$$

with probability at least $1 - \exp(-c_1 n_0)$, and then

$$\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|^2 \lesssim \lambda_2 C_{\boldsymbol{\Sigma}} d \tag{24}$$

by noticing $C_{\boldsymbol{\Sigma}} d(\log(p \vee n_0)/n_0)^{1/2} = o(1)$ as derived from Condition 3.
(2) If $\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}} \ge \lambda_2\|\boldsymbol{\delta}^{\mathcal{A}}\|_1$, we have

$$0 \le \frac{1}{4}\widehat{\boldsymbol{v}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{v}}^{\mathcal{A}} \le 3\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}} - \frac{\lambda_2}{2}\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|_1.$$

It follows that

$$\begin{aligned}
\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|_1 &\le \frac{6}{\delta_2}\widehat{\boldsymbol{u}}^{\mathcal{A}\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}} \\
&\lesssim \frac{\left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2}{\lambda_2} \\
&\lesssim \frac{s_0\lambda_1^2 + \lambda_1 C_{\boldsymbol{\Sigma}} d}{\lambda_2}
\end{aligned} \tag{25}$$

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

with probability at least $1-2\exp(-c_1 n_0)$, where the second line holds true with probability at least $1-\exp(-c_1 n_0)$ due to Theorem 1.6 of Zhou [26], and the third line holds true with probability at least $1-\exp(-c_1(n_{\mathcal{A}}+n_0))$ due to (18) and (21). By Lemma 3 and Theorem 1.6 of Zhou [26], once again, we have

$$\begin{aligned}\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|^2-\frac{\log p}{n_0}\left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|_1^2 &\lesssim \frac{1}{4}\widehat{\boldsymbol{v}}^{\mathcal{A}^\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{v}}^{\mathcal{A}}\\ &\leq 3\widehat{\boldsymbol{u}}^{\mathcal{A}^\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}}\\ &\lesssim \left\|\widehat{\boldsymbol{u}}^{\mathcal{A}}\right\|^2\\ &\lesssim s_0\lambda_1^2+\lambda_1 C_{\boldsymbol{\Sigma}} d\end{aligned}$$

with probability at least $1-3\exp(-c_1 n_0)$. Thus, from (25) that shows $\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\|_1^2\log p/n_0 \leq (\lambda_2\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\|_1)^2 = o(1)$, we have

$$\widehat{\boldsymbol{u}}^{\mathcal{A}^\top}\widehat{\boldsymbol{\Sigma}}^{(0)}\widehat{\boldsymbol{u}}^{\mathcal{A}} \vee \left\|\widehat{\boldsymbol{v}}^{\mathcal{A}}\right\|^2 \lesssim s_0\lambda_1^2+\lambda_1 C_{\boldsymbol{\Sigma}} d. \tag{26}$$

Statement (2) is established by combining results from (22), (23), (24), (25) and (26).

We now prove that $\mathbb{P}(B_1\cap B_2)\geq 1-c_1(\min_{k\in\mathcal{A}\cup\{0\}} n_k)^{-c_2 s_{\mathcal{A}}}-c_3\exp(-c_4\min_{k\in\mathcal{A}\cup\{0\}} n_k)-\exp(-c_5\log p)$ for some positive constants c_1 to c_5. We utilise the inequality $\mathbb{P}(B_1\cap B_2)\geq \mathbb{P}(B_1)+\mathbb{P}(B_2)-1$ and then bound $\mathbb{P}(B_1)$ and $\mathbb{P}(B_2)$ separately. For $\mathbb{P}(B_2)$, we have $\mathbb{P}(B_2)\geq 1-c_1 n_0^{-c_2 s_0}-c_3\exp(-c_4 n_0)$ by Lemma 2. For $\mathbb{P}(B_1)$, we bound it by the following decomposition:

$$\begin{aligned}&\frac{1}{n_A+n_0}\left\|\sum_{k\in\mathcal{A}\cup\{0\}} X^{(k)^\top}(\tilde{y}^{(k)}-X^{(k)}\boldsymbol{\beta}^{\mathcal{A}})\right\|_\infty\\ &=\left\|\frac{1}{n_{\mathcal{A}}+n_0}\sum_{k\in\mathcal{A}\cup\{0\}}\boldsymbol{X}^{(k)^\top}\tilde{\boldsymbol{y}}^{(k)}-\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\widehat{\boldsymbol{\Sigma}}^{(k)}\boldsymbol{\beta}^{(k)}+\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\widehat{\boldsymbol{\Sigma}}^{(k)}\boldsymbol{\beta}^{(k)}-\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}}\boldsymbol{\beta}^{\mathcal{A}}\right\|_\infty\\ &=\left\|\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\left(\frac{1}{n_k}\boldsymbol{X}^{(k)^\top}\tilde{\boldsymbol{y}}^{(k)}-\widehat{\boldsymbol{\Sigma}}^{(k)}\boldsymbol{\beta}^{(k)}\right)+\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\left(\widehat{\boldsymbol{\Sigma}}^{(k)}-\boldsymbol{\Sigma}^{(k)}\right)\delta^{(k)}-\left(\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}}-\boldsymbol{\Sigma}^{\mathcal{A}}\right)\delta^{\mathcal{A}}\right\|_\infty\\ &\leq\left\|\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\left(\frac{1}{n_k}\boldsymbol{X}^{(k)^\top}\tilde{\boldsymbol{y}}^{(k)}-\widehat{\boldsymbol{\Sigma}}^{(k)}\boldsymbol{\beta}^{(k)}\right)\right\|_\infty+\left\|\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\left(\widehat{\boldsymbol{\Sigma}}^{(k)}-\boldsymbol{\Sigma}^{(k)}\right)\delta^{(k)}\right\|_\infty\\ &\quad+\left\|\left(\widehat{\boldsymbol{\Sigma}}^{\mathcal{A}}-\boldsymbol{\Sigma}^{\mathcal{A}}\right)\delta^{\mathcal{A}}\right\|_\infty\\ &\stackrel{\text{def}}{=} D_1+D_2+D_3\,.\end{aligned}$$

For D_1, based on Lemma 2, we have

$$\begin{aligned}\alpha_k\left\|\frac{1}{n_k}\boldsymbol{X}^{(k)^\top}\tilde{\boldsymbol{y}}^{(k)}-\widehat{\boldsymbol{\Sigma}}^{(k)}\boldsymbol{\beta}^{(k)}\right\|_\infty &\lesssim \frac{n_k}{n_{\mathcal{A}}+n_0}\left(\sqrt{\frac{\log(p\vee n_k)}{n_k}}+\frac{s_k\log(p\vee n_k)}{n_k}\right)\\ &\lesssim \frac{\sqrt{n_k\log(p\vee n_k)}}{n_{\mathcal{A}}+n_0}+\frac{s_k\log(p\vee n_k)}{n_{\mathcal{A}}+n_0}\\ &\lesssim \frac{\sqrt{n_k\log(p\vee n_k)}}{n_{\mathcal{A}}+n_0}\\ &\lesssim \sqrt{\frac{\log(p\vee n_k)}{n_{\mathcal{A}}+n_0}}\end{aligned}$$

for any $k\in\mathcal{A}\cup\{0\}$ with probability at least $1-c_1 n_k^{-c_2 s_k}-c_3\exp(-c_4 n_k)$, where the third line holds true due to $s_k=O(n_k^b)$ with $0<b<1/3$ from Condition 1. Hence, considering that $|\mathcal{A}|$ is bounded,

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

we have

$$D_1 \lesssim \max_{k\in\mathcal{A}\cup\{0\}} \sqrt{\frac{\log(p\vee n_k)}{n_{\mathcal{A}}+n_0}} \lesssim \sqrt{\frac{\log(p\vee n_{\mathcal{A}^*})}{n_{\mathcal{A}}+n_0}} \tag{27}$$

with probability at least $1-c_1\min_{k\in\mathcal{A}\cup\{0\}} n_k^{-c_2 s_{\mathcal{A}}} - c_3\exp(-c_4\min_{k\in\mathcal{A}\cup\{0\}} n_k)$. For D_2 and D_3, similar with proofs of Theorems 1 and 4 of Li et al. [7], with a positive constant c_5, we have

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{k\in\mathcal{A}\cup\{0\}}\alpha_k\left(\boldsymbol{\Sigma}^{(k)}-\boldsymbol{\Sigma}^{(k)}\right)\boldsymbol{\delta}^{(k)}\right\|_\infty \geq t\right)\\
&\leq \mathbb{P}\left(\frac{1}{n_{\mathcal{A}}+n_0}\left\|\sum_{k\in\mathcal{A}\cup\{0\}}\left(\boldsymbol{X}^{(k)^\top}\boldsymbol{X}^{(k)}-E\left[\boldsymbol{X}^{(k)^\top}\boldsymbol{X}^{(k)}\right]\right)\boldsymbol{\delta}^{(k)}\right\|_\infty \geq t\right)\\
&\leq p\cdot\mathbb{P}\left(\max_{j\leq p}\frac{1}{n_{\mathcal{A}}+n_0}\left|\left\{\sum_{k\in\mathcal{A}\cup\{0\}}\left(\boldsymbol{X}^{(k)^\top}\boldsymbol{X}^{(k)}-E\left[\boldsymbol{X}^{(k)^\top}\boldsymbol{X}^{(k)}\right]\right)\boldsymbol{\delta}^{(k)}\right\}_j\right| \geq t\right)\\
&\leq p\max_{j\leq p}\exp\left(-c_4\min\left\{\frac{(n_{\mathcal{A}}+n_0)\,t^2}{2\max\limits_{k\in\mathcal{A}\cup\{0\}}\boldsymbol{\delta}^{(k)^\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}}, \frac{(n_{\mathcal{A}}+n_0)\,t}{\max\limits_{k\in\mathcal{A}\cup\{0\}}\sqrt{\boldsymbol{\delta}^{(k)^\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}}}\right\}\right)\\
&= p\max_{j\leq p,k\in\mathcal{A}\cup\{0\}}\exp\left(-c_4\min\left\{\frac{(n_{\mathcal{A}}+n_0)\,t^2}{2\boldsymbol{\delta}^{(k)^\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}}, \frac{(n_{\mathcal{A}}+n_0)\,t}{\sqrt{\boldsymbol{\delta}^{(k)^\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}}}\right\}\right),
\end{aligned}$$

where the fourth line holds true by combining Lemma 4 and the fact that $x_{i,j}^{(k)}\boldsymbol{x}_i^{(k)^\top}\boldsymbol{\delta}^{(k)}$ is sub-exponential. Hence, applying similar procedure on D_3 with D_2, by taking $t \geq \max_{k\in\mathcal{A}\cup\{0\}}\sqrt{\boldsymbol{\delta}^{(k)^\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}\log p/(n_{\mathcal{A}}+n_0)}$ and noticing that $\log p = o(n_{\mathcal{A}}+n_0)$, we have

$$\mathbb{P}\left(\max(D_2,D_3) \gtrsim \max_{k\in\mathcal{A}\cup\{0\}}\sqrt{\boldsymbol{\delta}^{(k)^\top}\boldsymbol{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}\log p/(n_{\mathcal{A}}+n_0)}\right) \leq \exp(-c_4\log p). \tag{28}$$

Combining (27), (28), we have

$$\mathbb{P}(B_1) \geq 1-c_1\min_{k\in\mathcal{A}\cup\{0\}} n_k^{-c_2 s_{\mathcal{A}}} - c_3\exp\left(-c_4\min_{k\in\mathcal{A}\cup\{0\}} n_k\right) - \exp(-c_5\log p),$$

which further leads to the desired conclusion that $\mathbb{P}(B_1\cap B_2)\to 1$ as $p\to\infty$ and $\min_{k\in\mathcal{A}\cup\{0\}} n_k\to\infty$. The proof of Lemma 1 is completed.

Consequently, we extend the convergence results established for quantile regression to the composite quantile regression setting. The key insight is that CQR can be viewed as a weighted average of multiple QR estimators, allowing us to leverage the existing QR theory.

Define the CQR estimator as $\widehat{\boldsymbol{\beta}}^{\mathrm{CQR}} = \frac{1}{S}\sum_{s=1}^S \widehat{\boldsymbol{\beta}}_{\tau_s}$, where $\widehat{\boldsymbol{\beta}}_{\tau_s}$ is the QR estimator at quantile level τ_s. The estimation error decomposes as:

$$\|\widehat{\boldsymbol{\beta}}^{\mathrm{CQR}} - \bar{\boldsymbol{\beta}}^{(0)}\| \leq \underbrace{\frac{1}{S}\sum_{s=1}^S\|\widehat{\boldsymbol{\beta}}_{\tau_s}-\boldsymbol{\beta}_{\tau_s}^{(0)}\|}_{F1} + \underbrace{\frac{1}{S}\sum_{s=1}^S\|\boldsymbol{\beta}_{\tau_s}^{(0)}-\bar{\boldsymbol{\beta}}^{(0)}\|}_{F2}. \tag{29}$$

(1) Bounding Term F1 (estimation error).

For each τ_s, under Conditions 1–6 and 8, Lemma 1 implies that with probability at least $1-\delta_s$,

$$\|\widehat{\boldsymbol{\beta}}_{\tau_s}-\boldsymbol{\beta}_{\tau_s}^{(0)}\| \lesssim \xi_1\left(n_0, n_{\mathcal{A}}, \max_{k\in\mathcal{A}\cup\{0\}} n_k, s_0, p\right), \tag{30}$$

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

where $\delta_s = C_1\exp(-C_2 \min_{k\in\mathcal{A}\cup\{0\}} n_k) + C_3(\min_{k\in\mathcal{A}\cup\{0\}} n_k)^{-C_4 s_{\mathcal{A}}} + \exp(-C_5 \log p)$. By the union bound and the fact that S is finite (typically small in practice), we obtain that with probability at least $1 - S\delta_s$,

$$F1 \lesssim \xi_1\left(n_0, n_{\mathcal{A}}, \max_{k\in\mathcal{A}\cup\{0\}} n_k, s_0, p\right). \tag{31}$$

(2) Bounding Term F2 (approximation bias).

Under Condition 7, we have

$$F2 \leq \frac{1}{S}\sum_{s=1}^{S} L|\tau_s - \bar{\tau}| \leq L \cdot \max_{1\leq s\leq S} |\tau_s - \bar{\tau}|, \tag{32}$$

where $\bar{\tau} = \frac{1}{S}\sum_{s=1}^{S}\tau_s$. For equally spaced quantiles, this term is $O(1/S)$. Since S is fixed and $\xi_1(\cdot) \gg 1/S$ under the sample size conditions in Condition 3, Term F2 is dominated by the estimation error and can be absorbed into the $\lesssim$ notation.

Substituting (31) and (32) into (29), we obtain

$$\left\|\widehat{\boldsymbol{\beta}}^{\mathrm{CQR}} - \bar{\boldsymbol{\beta}}^{(0)}\right\| \lesssim \xi_1\left(n_0, n_{\mathcal{A}}, \max_{k\in\mathcal{A}\cup\{0\}} n_k, s_0, p\right).$$

The prediction error bound follows similarly by noting the linearity of the prediction operator. The final probability bound follows from the union bound over S quantile levels. Since S is finite and typically small, the factor S in front of δ_s can be absorbed into the constants $C_1, \ldots, C_5$.

Remark 2. The proof strategy employed here – decomposing the CQR error into estimation error and approximation bias – is general and can be adapted to other loss functions that are linear combinations of convex losses. While the convergence rate remains the same as in the QR case, CQR often exhibits better finite-sample performance due to its robustness to heavy-tailed errors and its ability to capture broader features of the conditional distribution.

Appendix B. Sensitivity analysis for ϵ_0

During the informative set detection process in transfer learning, the parameter ϵ_0 controls the estimation accuracy of the informative set. In this section, we investigate its influence on model performance by selecting different values of ϵ_0. If ϵ_0 is set too large, it may introduce noise; if it is too small, the information may be insufficient. Therefore, we design several ϵ_0 values in the simulation and compare the model performance under different settings, such as changes in loss values and running time. By plotting the relationship curves between ϵ_0 and various metrics, we aim to identify the optimal trade-off point. Here, we assume the relationship between the estimated composite quantile loss under a homogeneous design and ϵ_0. We consider the following cases and plot the curves of ϵ_0 versus model performance (e.g. prediction error) to observe inflection points or stable intervals:

1. **Relatively small ϵ_0 values, e.g. 0.001, 0.005.** These make the detection process stricter. Only sources very close to the target model are allowed into the informative set, which may result in a smaller but more accurate informative set and reduce the risk of negative transfer.
2. **The commonly used baseline value of 0.01 in the literature.**
3. **Relatively large ϵ_0 values, e.g. 0.05, 0.1.** These make the criterion more lenient. More source models may be included, increasing the information, but irrelevant models may also be introduced, potentially degrading performance.

In practical applications, the choice of ϵ_0 can be tailored to the scenario. For instance, in high-reliability settings (e.g. medical applications), smaller ϵ_0 values are preferred, whereas in scenarios

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

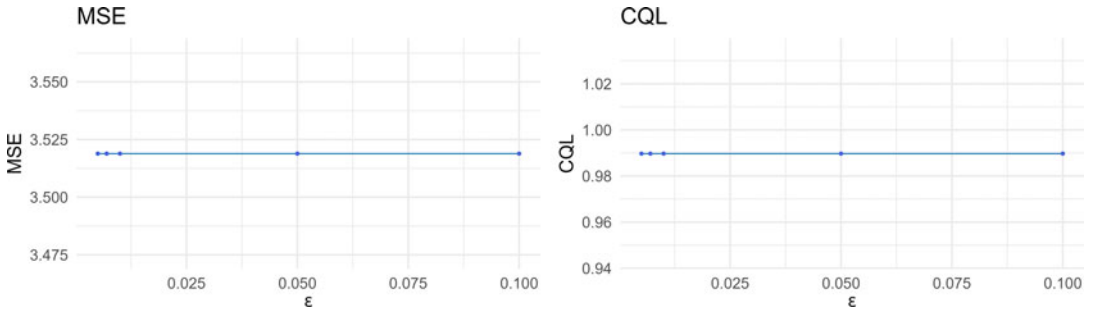


Figure A8. *Average MSE and composite quantile loss versus the threshold ϵ_0 (homogeneous design, normal errors).*

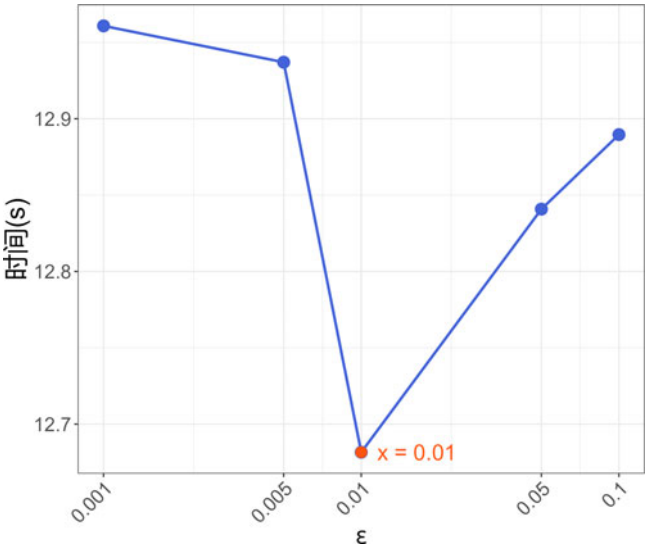


Figure A9. *Average running time versus the threshold ϵ_0 (homogeneous design, normal errors).*

requiring high generalisation (e.g. recommendation systems), larger ϵ_0 values can be tested. The comparison of model evaluation metrics for different ϵ_0 values is shown in Figures A8–A9.

Figures A8 and A9 display the MSE, composite quantile loss and running time under different values of ϵ_0. Due to the large variation in the horizontal coordinate, a logarithmic transformation is applied to the running time axis to better visualise the results across different ϵ_0 values. From the figures, it can be observed that the choice of ϵ_0 has little impact on model performance, with the main difference being in running time. The plots show that the running time is minimised at $\epsilon_0 = 0.01$. Therefore, in TL-CQR, to simplify the computation, we set the parameter in Step 5 of Algorithm 2 to $\epsilon_0 = 0.01$.

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

Appendix C. Features dictionary of flight QAR analysis

Table A3. *Features dictionary of flight QAR analysis*

Code	Define	Type
3000	Max airspeed during flaps extended	Speed
3001	Max airspeed during landing gear extended	Speed
3002	Max airspeed during speed brakes extended	Speed
3005	Average airspeed during climb	Speed
3007	Average airspeed during cruise	Speed
3009	Average airspeed during descent	Speed
3013	Average airspeed during engine stop	Speed
3023	Average airspeed during preflight	Speed
3027	Average airspeed during taxi in	Speed
3029	Average airspeed during taxi out	Speed
3035	Max groundspeed during turn	Speed
3036	Max groundspeed during pre lift off	Speed
3037	Max groundspeed during straight line	Speed
3057	Time during lift off to landing gear selected up	Time
3058	Time during landing gear selected down to touchdown	Time
3070	Time during climb	Time
3071	Time during cruise	Time
3072	Time during descent	Time
3089	Time during touchdown to start of all engine reverser deployed	Time
3090	Time during final 200 kn CAS to touchdown	Time
3093	Max vertical acceleration during flight	Acceleration
3094	Max vertical acceleration during landing	Response
3096	Max vertical acceleration during take off	Acceleration
3104	Max vertical acceleration during pre take off	Acceleration
3105	Max vertical acceleration during post landing	Acceleration
3112	Average wind direction during first 20 ft AFE to first 100 ft AFE	Wind
3113	Average angle-of-attack during cruise	Attitude
3119	Difference in fuel consumed engine 1 during lift off to first cruise	Fuel
3123	Difference in fuel consumed engine 1 during climb	Fuel
3124	Difference in fuel consumed engine 1 during cruise	Fuel
3125	Difference in fuel consumed engine 1 during descent	Fuel
3192	Average fuel flow engine 1 during climb	Fuel
3193	Average fuel flow engine 1 during cruise	Fuel
3194	Average fuel flow engine 1 during descent	Fuel
3195	Average fuel flow engine 1 during engine start	Fuel
3201	Average fuel flow engine 1 during preflight	Fuel
3203	Average fuel flow engine 1 during taxi in	Fuel
3204	Average fuel flow engine 1 during taxi out	Fuel
3253	Total ground distance during climb	Distance
3254	Total ground distance during cruise	Distance
3255	Total ground distance during descent	Distance
3256	Total ground distance during engine start	Distance

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

Table A3. *Continued.*

Code	Define	Type
3265	Total ground distance during taxi out	Distance
3268	Max absolute longitudinal acceleration during landing	Acceleration
3271	Average Mach during climb	Speed
3273	Average Mach during Cruise	Speed
3275	Average Mach during descent	Speed
3279	Average Mach during engine stop	Speed
3293	Average Mach during taxi in	Speed
3295	Average Mach during taxi out	Speed
3344	Max absolute bank angle during final 50 ft AFE to touchdown	Attitude
3361	Max airspeed – take off safety speed during slats only extended	Speed
3362	Max airspeed – take off safety speed during flap setting 1	Speed
3363	Max airspeed – take off safety speed during flap setting 2	Speed
3364	Max airspeed – take off safety speed during flap setting 3	Speed
3365	Max airspeed – take off safety speed during flap setting 4	Speed
3366	Max airspeed – take off safety speed during flap setting 5	Speed
3367	Max airspeed – take off safety speed during flap setting 6	Speed
3368	Max airspeed – take off safety speed during flap setting 7	Speed
3384	Max wind speed during first 20 ft AFE to first 100ft AFE	Wind
3385	Max absolute lateral acceleration during landing	Acceleration
3387	Min maximum operating speed – airspeed during flight	Speed
3418	Max airspeed – landing reference speed during slats only extended	Speed
3419	Max airspeed – landing reference speed during flap setting 1	Speed
3420	Max airspeed – landing reference speed during flap setting 2	Speed
3421	Max airspeed – landing reference speed during flap setting 3	Speed
3422	Max airspeed – landing reference speed during flap setting 4	Speed
3423	Max airspeed – landing reference speed during flap setting 5	Speed
3424	Max airspeed – landing reference speed during flap setting 6	Speed
3425	Max airspeed – landing reference speed during flap setting 7	Speed
3426	Max airspeed – landing reference speed during flap setting 8	Speed

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press

Table A4. *Feature dictionary for flight QAR data analysis of scalar measurements*

Feature dictionary		
Code	Define	Type
1000	Airspeed at final wing change configuration before touchdown	Speed
1001	Airspeed at first wing change configuration after lift off	Speed
1002	Airspeed at first wing clean configuration after lift off	Speed
1005	Airspeed at landing gear selected down	Speed
1006	Airspeed at landing gear selected up	Speed
1018	Altitude at landing gear selected down	Attitude
1019	Altitude at landing gear selected up	Attitude
1026	Altitude at final 200 kn cas	Attitude
1034	Gross weight at start of flight	Weight
1050	Gross weight at start of taxi out	Weight
1056	Groundspeed at start of off runway	Speed
1057	Landing reference speed at threshold	Speed
1058	Landing reference speed at touchdown	Speed
1062	Landing reference speed at final 50 ft AFE	Attitude
1065	Pitch at lift off	Attitude
1070	Pitch rate at lift off	Attitude
1075	Rotation speed at rotation	Speed
1076	Take off decision speed, V1 at lift off	Speed
1077	Take off safety speed, V2 at lift off	Speed
1080	Temperature at lift off	Temperature
1081	Temperature at touchdown	Temperature
1087	Time at start of climb	Time
1088	Time at start of cruise	Time
1096	Time at start of preflight	Time
1103	Vertical speed at touchdown	Speed
1109	Wind direction at lift off	Wind
1110	Wind direction at touchdown	Wind
1113	Bank angle at touchdown	Attitude
1114	Stabiliser angle at rotation	Attitude
1115	Wind speed at touchdown	Wind

Cite this article: Huang L., Wang L. and Lai P. (2026). Transfer learning in hard landing risk prediction. *The Aeronautical Journal*, **130**, 2080–2107. https://doi.org/10.1017/aer.2026.10166

https://doi.org/10.1017/aer.2026.10166 Published online by Cambridge University Press