Journal of Low Carbon Economy 低碳经济, 2026, 15(4), 297-309
Published Online November 2026 in Hans. https://www.hanspub.org/journal/jlce
https://doi.org/10.12677/jlce.2026.154031

Hans汉斯

全国碳市场收益率多期分位数预测与区间校准

张子旭，古勇毅，徐榛杉*

广东财经大学统计与数据科学学院，广东 广州

收稿日期：2026年8月18日；录用日期：2026年9月10日；发布日期：2026年11月5日

摘 要

本文以中国全国碳市场碳排放配额(CEA)收益率为研究对象，在严格样本外框架下对未来5个和20个交易日的累计对数收益率开展条件分位数预测，并利用0.05和0.95分位数构造名义90%预测区间。针对不同预测期限可能具有不同信息需求的特点，本文构建分期限差异化分位数预测方法：5日预测采用S、SL和SLM三个信息集的同分位数预测等权组合，20日预测采用基于vol_{20}阈值的波动状态门控，以此作为提高多期限分位数预测精度的核心预测结构。在此基础上，进一步采用序贯共形化分位数回归(Sequential Conformalized Quantile Regression, Sequential CQR)作为区间校准工具，根据历史已实现预测误差对原始区间进行动态修正。相较于采用相同Sequential CQR校准规则的基准方法(S_{CQR})，本文方法在全滚动样本中的Mean Pinball Loss分别下降约2.13%和7.00%，且相对于基准方法(S_{CQR})的平均损失差在多种稳健检验下均达到统计显著水平。未经校准的5日和20日预测区间Coverage分别为80.79%和65.79%，经Sequential CQR校准后提高至89.27%和88.55%，接近90%的名义覆盖水平。进一步研究发现，本文方法的预测优势在2025~2026后段样本中明显减弱，且不再具有统计显著性。研究结果表明，分期限差异化预测结构是提升多期分位数预测精度的主要来源，而Sequential CQR主要用于改善原始预测区间的覆盖可靠性。

关键词

全国碳市场，碳收益率，分位数预测，Sequential CQR，样本外预测

Multi-Horizon Quantile Forecasting and Interval Calibration of Returns in China's National Carbon Market

Zixu Zhang, Yongyi Gu, Yuanshan Xu*

School of Statistics and Data Science, Guangdong University of Finance and Economics, Guangzhou Guangdong

Received: August 18, 2026; accepted: September 10, 2026; published: November 5, 2026

*通讯作者。

文章引用: 张子旭, 古勇毅, 徐榛杉. 全国碳市场收益率多期分位数预测与区间校准[J]. 低碳经济, 2026, 15(4): 297-309. DOI: 10.12677/jlce.2026.154031

Abstract

This study examines the cumulative log returns of China Emission Allowances (CEA) in China's national carbon market and conducts conditional quantile forecasting for 5- and 20-trading-day horizons under a strict out-of-sample framework. Nominal 90% prediction intervals are constructed using the 0.05 and 0.95 conditional quantiles. To accommodate horizon-specific information needs, this study first constructs a horizon-specific quantile forecasting structure as the core forecasting layer: same-quantile forecasts from S, SL and SLM are equally weighted for the 5-day horizon, whereas a threshold-based volatility-state gating rule is used for the 20-day horizon. On this basis, Sequential CQR is employed as a post-calibration tool to adjust the raw prediction intervals using previously realized forecast errors. Relative to the benchmark method (S_{CQR}) under the same Sequential CQR calibration rule, the proposed horizon-specific forecasting structure reduces Mean Pinball Loss by approximately 2.13% and 7.00% for the 5-day and 20-day horizons, respectively, and the corresponding average loss differentials are statistically significant under multiple robustness checks. Before calibration, the interval coverage rates are 80.79% and 65.79%; after Sequential CQR calibration, they increase to 89.27% and 88.55%, respectively, approaching the nominal 90% coverage level. Further analysis shows that the method's forecasting advantage weakens in the later evaluation subsample covering 2025~2026 and is no longer statistically significant. These findings indicate that the horizon-specific differentiated forecasting structure is the primary source of improved multi-horizon quantile forecasting accuracy, whereas Sequential CQR is mainly used to improve the coverage reliability of the raw prediction intervals.

Keywords

China's National Carbon Market, CEA Returns, Quantile Forecasting, Sequential CQR, Out-of-Sample Forecasting

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

近年来，在应对气候变化和加快能源低碳转型的大背景下，碳排放权交易市场持续发展和完善。中国碳市场伴随着钢铁、水泥、铝冶炼等高排放行业的正式纳入[1]，其资源配置、减排激励作用将进一步突显。碳价作为一个关键市场变量，受政策、供需、宏观经济等不确定因素所影响，往往难以预测。因此，准确刻画碳价动态特征不仅关系到市场主体的履约决策和碳资产管理，也有助于认识全国碳市场的价格形成与风险特征。

既有碳价预测研究主要聚焦于预测精度的提升。Zhu 和 Wei 将 ARIMA 与最小二乘支持向量机相结合，以同时刻画碳价中的线性与非线性特征[2]；随着后续研究对碳价非线性和复杂性的关注不断增加，以及机器学习的快速发展，神经网络、支持向量机等机器学习方法被逐渐应用于碳价预测[3]。针对碳价的非线性和多尺度特征，Qi 等进一步将历史价格信息与相关的影响因素结合，构建多尺度集成预测模型[4]。总体而言，这类研究主要关注价格点预测及其预测精度，对预测结果不确定性及收益率条件分布的刻画仍相对有限。

分位数预测能够从条件分布的不同位置描述未来收益率，为分析下行风险、中心位置和上行变化提

供统一框架[5]。已有研究发现，CEA 收益率的预测因子及其预测作用可能随分位位置而变化[6]，能源价格对碳价的影响也具有明显的非对称性和分位数异质性[7]。据此，本文选取 0.05、0.50 和 0.95 三个分位点，分别刻画全国碳市场收益率的下尾、中位和上尾特征。

仅取得条件分位数预测还不足以保证区间预测可靠。由 0.05 和 0.95 分位数构成的名义 90%预测区间，其实际覆盖率可能出现不足或过覆盖。因此，仅评价单个分位数的预测精度仍不足以全面衡量预测区间的可靠程度，还需关注区间的校准性，即实际覆盖率与名义覆盖率的一致程度。依此，引入 Sequential CQR 对原始分位数区间进行动态校准，以改善样本外区间的覆盖表现。

本文以全国碳市场 CEA 收益率作为研究对象，分别开展 5 日和 20 日条件分位数预测，并利用上下分位数构造 90%预测区间。引入 Sequential CQR 进行区间校准，并从预测精度、覆盖率、区间锐度和统计显著性等方面评价样本外表现。

基于上述考虑，本文将预测问题划分为两个层次。第一层为分期限差异化分位数预测，其核心目标是针对 5 日和 20 日预测期限的不同信息需求提高条件分位数预测精度；第二层为区间后校准，在维持预测层结构不变的基础上，利用 Sequential CQR 根据历史已实现预测误差修正原始区间边界，以改善预测区间的覆盖可靠性。由此，本文将“预测精度改善”与“区间可靠性校准”在方法作用上予以区分。

相较于现有研究，本文贡献主要体现在以下三个方面。

第一，针对不同预测期限可能存在差异化信息需求的问题，构建分期限差异化分位数预测结构。5 日预测通过 S、SL 和 SLM 三个嵌套信息集的同分位数等权组合整合不同层级信息，20 日预测则依据样本外开始前冻结的波动率阈值，在 S 和 SL 信息集之间实施确定性状态门控，从而避免对不同预测期限机械采用统一的信息结构。

第二，在上述分期限预测结构基础上引入 Sequential CQR 作为区间后校准工具，仅利用预测时点已经实现的历史预测误差动态调整区间边界，以改善原始分位数区间的欠覆盖问题，并在方法作用上区分预测精度提升与区间覆盖校准。

第三，从 Mean Pinball Loss、Coverage、Interval Width 和 Interval Score 等维度分别评价预测精度、覆盖可靠性与区间锐度，并结合 Newey-West、区块 Bootstrap 及 Holm 校正考察预测优势的统计显著性和时间稳定性。

2. 数据与研究方法

2.1. 数据来源与变量说明

研究对象为 CEA。数据起始于 2021 年 7 月 16 日，截止至 2026 年 5 月 6 日，主要包括 CEA 价格、成交量及其金额，以及挂牌协议交易等市场信息。鉴于收益率、滚动波动率等变量构造以及不同市场数据间的交易日匹配，完成数据整理后的共同建模面板覆盖 2021 年 10 月 25 日至 2026 年 4 月 29 日。本文样本外预测自 2023 年 1 月 3 日开始，并分别设置 5 个与 20 个交易日两个预测期限。具体样本划分与实验设计见表 1。

Table 1. Sample and experimental design
表 1. 样本与实验设计

预测期限	样本外开始日期	样本外总样本数	CQR 可用样本数	后段评价样本数
5	2023-01-03	799	755	315
20	2023-01-03	784	725	300

注：预测分位点 0.05/0.50/0.95；每 20 个预测交易日重拟合；Sequential CQR 最低历史样本数 40。

为直观展示样本期内全国 CEA 价格变化及不同评价阶段的位置，图 1 给出 CEA 收盘价时间序列。

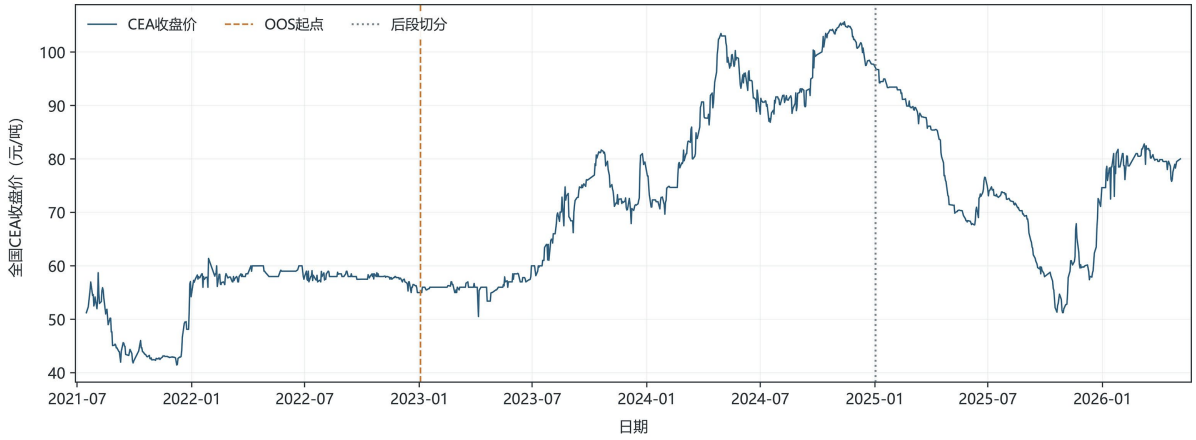


Figure 1. Time series of national CEA closing prices
图 1. 全国 CEA 收盘价时间序列

按照预测信息范围，本文构建 S、SL 和 SLM 三个严格嵌套的信息集。S 包含 13 个 CEA 市场短期内部变量，主要刻画近期价格、波动和交易活跃度；SL 在 S 基础上新增 10 个 20 日和 60 日尺度的中长期市场变量，共计 23 个；SLM 再加入金融市场、能源市场、欧盟碳排放配额(European Union Allowance, EUA)以及电力基本面等 10 个外部变量，共计 33 个，满足 S⊂SL⊂SLM。各信息集及对应模型的定义和正则化参数见表 2。

Table 2. Definitions of models, information sets, and regularization parameters
表 2. 模型与信息集定义及正则化参数

模型/信息集	定义	变量数	$h=5\alpha$ (q05/q50/q95)	$h=20\alpha$ (q05/q50/q95)	CQR
S	13 个短期全国 CEA 市场内部变量上的线性分位数回归	13	0.01/0.0001/0.001	0.1/0.0001/0	否
SL	S 增加 10 个 20 日和 60 日尺度的中长期市场变量	23	0.01/0.1/0.01	0.1/0.1/0.01	否
SLM	SL 增加 10 个金融、能源、EUA 和月度电力外部变量	33	0.1/0.1/0.1	0.1/0.1/0.01	否
S_{CQR}	对 S 的 q05 和 q95 采用 Sequential CQR 校准，q50 保持不变	13	0.01/0.0001/0.001	0.1/0.0001/0	是
等权组合–RAW $(h=5)$	S、SL、SLM 三个 RAW 预测在相同分位点上进行 1/3 等权平均	组合	沿用各组成模型	——	否
状态门控–RAW $(h=20)$	$\mathrm{vol}_{20,t} \geq c$ 时选择 SL，否则选择 S	状态切换	——	沿用所选择的 S 或 SL	否
分期限方法–CQR	对于对应 RAW 预测采用 Sequential CQR 校准，q50 保持不变	同对应 RAW 预测方案	沿用对应 RAW 预测所用模型	沿用对应 RAW 预测所用模型	是

注：S_{CQR} 与分期限方法_CQR 采用相同的 Sequential CQR 校准规则，但分别根据各自原始预测区间的历史预测误差计算校准量。$h=20$ 的状态门控阈值在样本外评价开始前根据历史 vol_{20} 中位数一次性确定并冻结，评价期内不重新估计。

为避免前视偏误，所有变量均按照预测时点的实际可得性进行时间对齐。CEA 内部变量仅使用预测时点及以前的信息；沪深 300 指数、电力行业指数和美元兑人民币汇率按照 CEA 交易日进行 as-of 匹配；动力煤、Brent 原油和 EUA 期货使用严格早于预测日的最近可用观测；对于全国用电量和火电发电量等指标，则依据实际信息发布日期确定可见时点，且仅在信息公开后进入预测模型。

2.2. 预测目标与分位数预测模型

将预测时点定义为CEA 交易日当日收盘后，以全国碳市场 CEA 未来多期对数收益率作为预测目标。设 P_t 为预测时点 t 的 CEA 价格，则未来 h 个交易日的对数收益率定义为：

$$y_{t,h} = \ln P_{t+h} - \ln P_t = \ln \frac{P_{t+h}}{P_t}, \tag{1}$$

其中，$h \in \{5,20\}$，分别对应未来 5 个和 20 个交易日的预测期限，$t+h$ 表示预测时点之后第 h 个有效 CEA 交易日，而非自然日。

为描述 CEA 收益率条件分布的不同位置，分别对 0.05、0.50 和 0.95 三个条件分位数进行预测[8]。其中 0.05、0.50 和 0.95 分位数分别刻画收益率条件分布的下尾、中位和上尾。对于给定分位点 τ，设 X_t 为预测时点 t 可获得的预测变量，则未来 h 期收益率的条件分位数表示为：

$$Q_\tau\left(y_{t,h} \mid X_t\right) = \beta_{0,\tau,h} + X_t^{\mathrm{T}} \beta_{\tau,h}. \tag{2}$$

采用线性 QuantileRegressor 进行估计，其优化目标为：

$$\min_{\beta_0,\beta} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau\left(y_i - \beta_0 - x_i^{\mathrm{T}} \beta\right) + \alpha \|\beta\|_1, \tag{3}$$

其中，α 为 L1 正则化系数，$\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$ 表示回归系数向量的 L1 范数，截距项 β_0 不参与 L1 惩罚。

Pinball Loss 定义为：

$$\rho_\tau(u) = \max\{\tau u, (\tau - 1)u\}, \tag{4}$$

其中，$u = y - \hat{Q}\tau$。

模型估计前，首先将正负无穷值处理为缺失值，并利用当次训练样本的列中位数进行填补；随后仅基于当次训练样本拟合 StandardScaler 标准化参数，并采用 QuantileRegressor (solver = 'highs')分别估计 0.05、0.50 和 0.95 三个条件分位数。L1 正则化参数 α 按照预测期限、信息集和分位点分别选择，集合为 {0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}}，并以验证样本平均 Pinball Loss 最小为选择标准。在参数选择中，训练样本满足预测日早于 2022 年 7 月 1 日且目标日不晚于 2022 年 7 月 1 日；验证样本满足预测日在 2022 年 7 月 1 日至 2023 年 1 月 3 日之前且目标日不晚于 2023 年 1 月 3 日。由此保证参数选择仅使用样本外评价开始前可获得的信息，确定后的 α 在后续样本外评价中沿用。

为避免分位数交叉，在获得 0.05、0.50 和 0.95 三个分位数预测结果后，按分位点顺序对预测值进行排序，以保证 q05 ≤ q50 ≤ q95。考虑到不同预测期限可能对应不同的信息需求，便采用分期限差异化分位数预测构造方式。

对于 5 个交易日预测期限，分别基于 S、SL 和 SLM 三个信息集得到相同分位点下的预测结果，并采用等权平均构造 5 日预测期限下的组合分位数预测，公式如下：

$$\hat{Q}_{\tau,t}^{C} = \frac{1}{3}\left(\hat{Q}_{\tau,t}^{S} + \hat{Q}_{\tau,t}^{SL} + \hat{Q}_{\tau,t}^{SLM}\right), \tag{5}$$

其中，$\hat{Q}^{S}_{\tau,t}$、$\hat{Q}^{SL}_{\tau,t}$ 以及 $\hat{Q}^{SLM}_{\tau,t}$ 分别表示基于三个信息集得到的分位数预测值。

该组合权重固定为 1/3，评价期内不重新选择；完成等权组合后，再次按照分位点顺序进行排序，以保证组合后的分位数预测满足 q05 ≤ q50 ≤ q95。

对于 20 个交易日预测期限，本文采用基于市场波动状态的确定性门控规则。具体而言，在样本外评价开始之前，根据历史样本中 20 日滚动波动率 vol_{20} 的中位数确定阈值约为 c = 0.008956451613091 (约 0.008956)。当预测时点的 vol_{20} 不低于该阈值时，采用 SL 信息集对应的分位数预测；否则采用 S 信息集预测。即：

$$\hat{Q}^{C}_{\tau,t}=\begin{cases}\hat{Q}^{SL}_{\tau,t}, & \mathrm{vol}_{20,t}\geq c,\\ \hat{Q}^{S}_{\tau,t}, & \mathrm{vol}_{20,t}<c.\end{cases} \tag{6}$$

其中，c 为样本外评价开始前预先确定并冻结的波动率阈值。

需要指出的是，该规则属于事先设定的确定性状态门控，而非在评价期内根据预测结果重新识别或优化市场状态。样本外评价自 2023 年 1 月 3 日开始，训练样本采用扩展方式更新，每 20 个预测交易日重新拟合模型。每次重新拟合仅使用在该 20 日预测块起点已经实现目标值的历史样本，同一预测块内不再追加随后新实现的目标值。本文将上述针对 5 日与 20 日预测期限的差异化分位数预测构造统称为“本文方法”。其中，5 日未经过 Sequential CQR 校准的分位数预测及其区间记为等权组合_RAW，20 日未经校准的预测结果记为状态门控_RAW；经 Sequential CQR 校准后的结果统称为分期限方法_CQR。上述分期限差异化预测结构构成本文的核心预测层，目的是针对不同预测期限的信息需求差异提高条件分位数预测精度。

2.3. 预测区间构建与序贯共形校准

在保持上述分期限预测层结构及其分位数预测结果不变的基础上，进一步设置区间校准层，对由 0.05 和 0.95 条件分位数形成的原始预测区间实施 Sequential CQR 校准。该步骤的目标是提高区间覆盖可靠性，而非重新选择预测模型或重新优化分位数预测结构。

利用 0.05 和 0.95 条件分位数构造未经校准的名义 90%预测区间。对于预测时点 t 的原始预测区间的上界和下界如下：

上界：

$$U_{t,\mathrm{raw}}=\hat{Q}_{0.95,t}, \tag{7}$$

下界：

$$L_{t,\mathrm{raw}}=\hat{Q}_{0.05,t}. \tag{8}$$

因此，原始 90%区间可表示为：

$$C_{t,\mathrm{raw}}=\left[L_{t,\mathrm{raw}},U_{t,\mathrm{raw}}\right]. \tag{9}$$

由 0.05 和 0.95 分位数构成名义覆盖率为 90%的原始预测区间，但其实际样本外覆盖率可能偏离名义水平，因此进一步采用 Sequential CQR 进行校准[9]。

具体而言，对于已经实现的历史预测样本，定义非一致性得分为：

$$s_i=\max\left\{L_{i,\mathrm{raw}}-y_i,y_i-U_{i,\mathrm{raw}}\right\}, \tag{10}$$

其中，y_i 表示第 i 个最终实现的实际收益率，$L_{i,\mathrm{raw}}$ 和 $U_{i,\mathrm{raw}}$ 分别表示对应时点的原始预测区间上下界。当实际收益率落在原始预测区间之外时，非一致性得分为正；当实际收益率落在原始预测区间之内时，该

得分可以为负。

在预测时点 t，仅使用目标值在当前预测时点已经实现的历史样本计算非一致性得分，即校准样本满足 $\text{target}_{\text{date}_i} \le \text{predictor}_{\text{date}_t}$。当可用历史非一致性得分数量 $n_t \ge 40$ 时，实施 Sequential CQR 校准。

对于当前可获得的 n_t 个历史非一致性得分，首先按照从小到大的顺序进行排列，并确定 90%名义覆盖水平下的有限样本校准位置：

$$k_t = \lceil (n_t + 1) \times 0.90 \rceil, \tag{11}$$

其中，$\lceil \cdot \rceil$表示向上取整。

设排序后的第 k_t 个非一致性得分为 $s_{(k_t)}$，则当前预测时点的区间校准量定义为：

$$\hat{q}_t = \max\left\{0, s_{(k_t)}\right\}. \tag{12}$$

通过上述处理，校准量被限制为非负值。CQR 只允许保持或扩张原始区间，不主动收窄区间。

经过校准后，预测区间上下界分别调整为：

上界：

$$U_{t,\text{CQR}} = U_{t,\text{raw}} + \hat{q}_t, \tag{13}$$

下界：

$$L_{t,\text{CQR}} = L_{t,\text{raw}} - \hat{q}_t. \tag{14}$$

因此，校准后的预测区间为：

$$C_{t,\text{CQR}} = \left[L_{t,\text{CQR}}, U_{t,\text{CQR}}\right]. \tag{15}$$

需指出，本文仅对 0.05 和 0.95 分位数对应的预测区间边界进行校准，而不调整 0.50 分位数预测结果。这一处理既保留了 0.50 分位数预测对条件分布中心位置的描述，又能够利用历史预测误差对尾部预测区间进行动态修正。

需要说明的是，本文将 Sequential CQR 作为时间序列样本外环境下的序贯经验校准方法，主要考察其实际覆盖表现，不据此主张在存在时间依赖条件下仍具有经典交换性假设下的有限样本覆盖保证。

2.4. 预测评价指标与统计检验

为全面评价不同预测方法的样本外表现，本文将从分位数预测精度、预测区间校准性、区间锐度和统计显著性四个方面进行比较。采用 Pinball Loss 和 Mean Pinball Loss 评价条件分位数预测精度，同时采用覆盖率(Coverage)评价预测区间的校准程度，采用平均区间宽度(Interval Width)评价区间锐度，并利用 Interval Score 综合衡量预测区间的宽度与漏覆损失。同时采用 Newey-West 稳健检验[10]、Circular Block Bootstrap [11]、Stationary Bootstrap [12]以及 Holm 多重检验校正[13]考察预测表现差异的统计显著性。

首先，对于给定分位点 τ，采用式(4)所定义的 Pinball Loss 衡量条件分位数预测误差，损失越小表示分位数预测误差越低。

鉴于同时预测 0.05、0.50 和 0.95 三个分位点，进一步采用 Mean Pinball Loss (MPL)评价三个条件分位数的整体预测表现。对于预测时点 t，Mean Pinball Loss 由三个分位点的 Pinball Loss 进行等权平均，其总体样本外指标表示为：

$$\text{MPL} = \frac{1}{3} \sum_{\tau \in \{0.05, 0.50, 0.95\}} \left[\frac{1}{n} \sum_{t=1}^{n} \rho_\tau \left(y_t - \hat{Q}_{\tau,t} \right) \right], \tag{16}$$

其中，τ 取 0.05、0.50 和 0.95，n 表示有效样本外预测数量。MPL 越小，表示模型对收益率条件分布整体的预测精度越高。

其次，采用覆盖率 Coverage 衡量预测区间的校准程度。设 L_t 和 U_t 分别表示预测时点 t 对应的预测区间下界和上界，y_t 为实际收益率，则 Coverage 定义为：

$$\text{Coverage} = \frac{1}{n}\sum_{t=1}^{n} I\left(L_t \le y_t \le U_t\right), \tag{17}$$

其中，$I(\cdot)$ 为示性函数。Coverage 越接近 0.90，区间校准程度越高。

在评价 Coverage 的同时，本文采用平均区间宽度(Interval Width)衡量预测区间的锐度，其定义为：

$$\text{Width} = \frac{1}{n}\sum_{t=1}^{n}\left(U_t - L_t\right). \tag{18}$$

在 Coverage 相近时，Width 越小表示区间锐度越高。

为综合考虑预测区间宽度与漏覆情况，本文进一步计算 90%预测区间对应的 Interval Score 如下：

$$\text{IS}_{0.10} = \left(U - L\right) + 20\left(L - y\right)I\left(y < L\right) + 20\left(y - U\right)I\left(y > U\right). \tag{19}$$

同时惩罚区间宽度和漏覆，数值越小越优。

在本文方法与基准方法的比较中，采用经过相同 Sequential CQR 校准程序的 S 信息集模型 S_{CQR} 作为可比基准，并将分期限方法_CQR 与 S_{CQR} 在相同样本外评价日期、相同三个分位点、相同 CQR 校准规则以及相同最低校准样本要求下进行比较。分期限方法_CQR 与 S_{CQR} 分别基于自身未经校准区间的历史预测误差计算校准量 $\hat{q}_t$，因此二者采用相同校准规则，但具体校准量不要求相同。

对于任一预测损失指标，本文方法相对于基准方法的损失改善率定义为：

$$\text{Improvement} = \frac{\text{Loss}_{\text{baseline}} - \text{Loss}_{\text{candidate}}}{\text{Loss}_{\text{baseline}}} \times 100\%, \tag{20}$$

其中，$\text{Loss}_{\text{baseline}}$ 表示 S_{CQR} 对应的预测损失，$\text{Loss}_{\text{candidate}}$ 表示分期限方法_CQR 对应的预测损失。当 $\text{Improvement} > 0$ 时，表示本文方法的预测损失低于基准方法。

为进一步评价本文方法与基准模型之间的预测损失差异，定义预测时点 t 的损失差为：

$$d_t = \text{Loss}_{\text{candidate},t} - \text{Loss}_{\text{baseline},t}. \tag{21}$$

因此，当 $d(t) < 0$ 时，表示本文方法在对应时点的预测损失低于基准方法。本文针对平均损失差进行单侧检验，检验假设设定为：

$$H_0 : E\left(d_t\right) \ge 0, \tag{22}$$

$$H_1 : E\left(d_t\right) < 0. \tag{23}$$

首先，采用 Newey-West 异方差与自相关稳健方法估计损失差序列均值的标准误，并使用 Bartlett 权重，最大滞后阶数设为 20。相应统计量为：

$$T_{\text{NW}} = \frac{\bar{d}}{\sqrt{\widehat{\text{LRV}}_{\text{NW}}/n}}, \tag{24}$$

其中，$\bar{d}$ 表示样本平均损失差，$\widehat{\text{LRV}}_{\text{NW}}$ 表示 Newey-West 方法估计的长期方差。按照零均值边界计算左尾单侧 p 值，以检验本文方法是否具有显著更低的平均预测损失。

除 Newey-West 稳健推断外，进一步采用 Circular Block Bootstrap 和 Stationary Bootstrap 对主要预测

损失比较进行补充检验。主比较中的 Bootstrap 检验用于全滚动样本和 2025~2026 后段评价样本的 Mean Pinball Loss，参数选择阶段不实施 Bootstrap。其中，Circular Block Bootstrap 固定块长为 20，Stationary Bootstrap 平均块长为 20、重启概率为 1/20，两种方法均重复 5000 次。

考虑到本文同时考察 5 日和 20 日两个预测期限，在同一样本期、同一评价指标和同一检验方法下，仅对 5 日和 20 日两个预测期限的 p 值按 Holm 方法校正($m=2$)；Coverage 与 0.90 的差异采用精确二项检验。

3. 实证结果与分析

3.1. 分位数预测结果

表 3 比较的是在相同 Sequential CQR 校准规则下，分期限预测方法与 S_{CQR} 基准的 Mean Pinball Loss，因此本节主要检验分期限预测结构本身的样本外预测价值，而非 Sequential CQR 的校准贡献。见表 3。

Table 3. Quantile forecasting results for the full rolling sample
表 3. 全滚动样本分位数预测结果

预测期限	有效样本数	分期限方法_CQR MPL	S_{CQR} MPL	MPL 改善率
$h=5$	755	0.007560	0.007724	2.1292%
$h=20$	725	0.018782	0.020196	7.0009%

注：改善率 $=(S_{CQR}-$ 分期限方法_CQR$)/S_{CQR}\times 100\%$，正值表示本文方法的 Mean Pinball Loss 低于基准。

从总体预测损失上看，两个预测期限下本文方法的 Mean Pinball Loss 均低于 S_{CQR}。对于 5 日预测，分期限方法_CQR 的 MPL 为 0.007560，低于 S_{CQR} 的 0.007724，损失下降约 2.13%；对于 20 日预测，MPL 由 0.020196 下降至 0.018782，降幅约为 7.00%。从相对改善率看，20 日预测的损失改善幅度高于 5 日预测。

上述 2.13%和 7.00%的改善均来自分期限方法_CQR 与 S_{CQR} 之间在统一校准条件下比较，因而应当主要归因于 5 日等权组合与 20 日状态门控所构成的差异化预测结构。Sequential CQR 的作用将在下一节通过 RAW 与 CQR 区间的 Coverage、Width 和 Interval Score 变化单独考察。

3.2. 预测区间与 Sequential CQR 校准结果

与上节分期限方法_CQR 和 S_{CQR} 之间的比较不同，本节在同一预测结构下来比较校准前的 RAW 区间与校准后的 CQR 区间，从而考察序贯共形校准前后预测区间的变化。表 4 报告了两个预测期限下的 Coverage、平均区间宽度和 Interval Score。具体结果见表 4。

Table 4. Prediction interval performance before and after sequential CQR calibration
表 4. Sequential CQR 校准前后的预测区间表现

预测期限	RAW Coverage	CQR Coverage	RAW Width	CQR Width	RAW Interval Score	CQR Interval Score
$h=5$	80.7947%	89.2715%	0.098225	0.134160	0.194259	0.181142
$h=20$	65.7931%	88.5517%	0.191903	0.334761	0.574219	0.447245

注：RAW 表示未经 Sequential CQR 校准的预测区间，CQR 表示采用 Sequential CQR 校准后的预测区间；名义覆盖率为 90%。

从覆盖率来看，未经校准的 RAW 预测区间均存在不同程度的欠覆盖。5 日预测的 RAW Coverage 为 80.7947%，低于 90%的名义覆盖水平；经过 Sequential CQR 校准后，Coverage 提高至 89.2715%，增加了

约 8.48 个百分点。20 日预测的原始欠覆盖则更加严重，RAW Coverage 仅为 65.7931%；校准后提高至 88.5517%，增加了约 22.76 个百分点。精确二项检验表明，两个预测期限下 RAW 区间的 Coverage 均显著低于 90%的名义水平，而 Sequential CQR 校准后的 Coverage 与 90%之间均不存在显著差异，其中 h = 5 和 h = 20 的双侧 p 值分别为 0.5044 和 0.1935。综上，Sequential CQR 明显改善了原始预测区间的欠覆盖问题，使实际覆盖率接近预设的 90%水平。

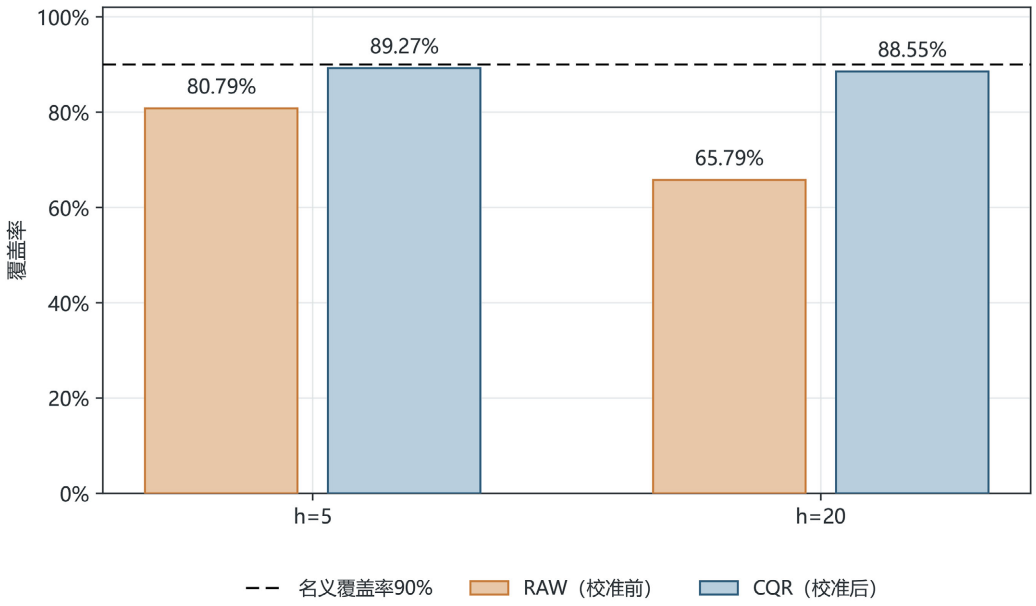


Figure 2. Prediction interval coverage before and after sequential CQR calibration
图 2. Sequential CQR 校准前后的预测区间覆盖率比较

从图 2 可知，Coverage 的改善伴随着预测区间宽度的增加。5 日预测的平均区间宽度由 0.098225 扩大至 0.134160；20 日预测则由 0.191903 扩大至 0.334761。由于校准量被限制为非负，Sequential CQR 仅保持或扩张原始预测区间，因此区间宽度增加与其校准机制一致。特别对于 RAW 欠覆盖程度较高的 20 日预测，校准后的区间扩张幅度也更大。

从 Interval Score 考察 Coverage 与区间宽度之间的综合权衡。5 日预测的 Interval Score 由 0.194259 下降至 0.181142，20 日预测则由 0.574219 下降至 0.447245。尽管 Sequential CQR 扩大了预测区间，但 Interval Score 在两个预测期限下均呈数值下降，这说明增加的区间宽度在一定程度上由漏覆损失的减少所补偿。因此，CQR 的作用并非简单扩大预测区间，而是在牺牲一定区间锐度的情况下改善原始区间的覆盖校准表现。

进一步从覆盖率–锐度权衡来看，Sequential CQR 在两个预测期限下均通过扩大区间换取更高的覆盖可靠性，但这种代价在不同期限上存在明显差异。5 日预测的 Coverage 由 80.7947%提高至 89.2715%，增加约 8.48 个百分点，同时平均区间宽度由 0.098225 增至 0.134160，相对增加约 36.58%；20 日预测的 Coverage 由 65.7931%提高至 88.5517%，增加约 22.76 个百分点，而平均区间宽度由 0.191903 增至 0.334761，相对增加约 74.44%。

尽管区间锐度有所下降，Interval Score 在 5 日和 20 日预测下分别由 0.194259 降至 0.181142、由 0.574219 降至 0.447245，相对下降约 6.75%和 22.11%，说明漏覆损失的减少在数值上能够补偿部分区间扩张成本。进一步的显著性检验显示，两个期限下 Interval Score 改善经 Holm 校正后的 Newey–West 的 p

值均为 0.1069，尚未达到 5%的显著性水平。

3.3. 稳定性与显著性分析

为进一步考察这种平均预测优势是否具有时间稳定性，并避免仅依据样本平均损失判断模型优劣，下面将进一步结合 2025~2026 后段评价样本以及多种时间序列稳健检验进行分析。

首先考察 2025~2026 后段评价样本。该样本属于全滚动评价期的后段子样本，并非独立于模型开发过程的确认集，因此本节主要用于研究预测优势在样本后段是否仍能维持。具体结果见表 5。

Table 5. Stability results for the 2025~2026 late evaluation sample
表 5. 2025~2026 后段稳定性结果

预测期限	有效样本数	分期限方法_CQR MPL	S_{CQR} MPL	MPL 改善率	Holm-NW p 值	分期限方法_CQR Coverage
$h = 5$	315	0.007972	0.007870	−1.2886%	0.7901	89.2063%
$h = 20$	300	0.019964	0.020267	1.4939%	0.5711	90.6667%

注：改善率为正表示分期限方法_CQR 的 Mean Pinball Loss 低于 S_{CQR}；2025~2026 为后段稳定性样本，不属于独立确认集。

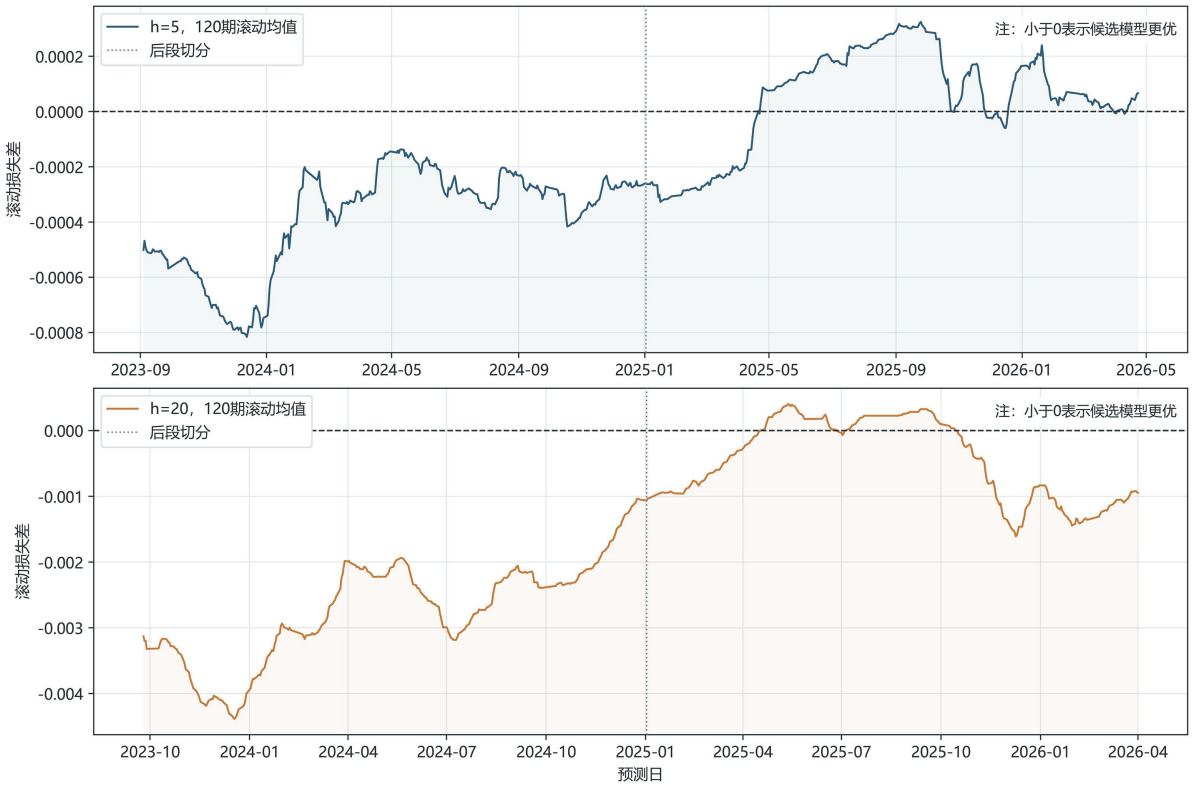


Figure 3. Rolling mean loss differential between the proposed method and the benchmark
图 3. 本文方法与基准模型的滚动平均损失差

从后段样本表现看，本文分期限预测结构在两个预测期限下均未继续表现出显著的平均预测优势。$h = 5$ 的 Mean Pinball Loss 为 0.007972，高于 S_{CQR} 的 0.007870，对应改善率为−1.2886%，即本文方法在该阶段出现约 1.29%的数值恶化，Holm 校正后的 Newey-West p 值为 0.7901。$h = 20$ 的 Mean Pinball Loss 由基准的 0.020267 下降至 0.019964，改善约 1.49%，但 Holm-NW p 值为 0.5711，同样未达到 5%的显著

性水平。综上，与全滚动样本相比，本文方法的精度优势在后段样本中明显减弱，说明分期限预测结构的平均优势具有一定阶段依赖性。

为进一步观察本文方法相对预测表现的时间变化，图 3 给出 120 期滚动平均损失差。损失差定义同式(21)。其中，负值表示本文方法的平均预测损失较低。该图仅用于描述时间变化，不作为模型重新选择的依据。

对于 Mean Pinball Loss，$h=5$ 的 Newey-West、Circular Block Bootstrap 和 Stationary Bootstrap 经 Holm 校正后的 p 值分别为 0.0377、0.0380 和 0.0402，均低于 0.05；$h=20$ 对应的 Holm 校正 p 值分别为 0.000856、0.0012 和 0.0064，同样均低于 0.05。三种检验方法得到一致结论，即在 2023~2026 全滚动评价样本中，本文方法相对于 S_{CQR} 具有显著更低的平均分位数预测损失。具体结果见表 6。

Table 6. Robust significance tests for Mean Pinball Loss
表 6. Mean Pinball Loss 的稳健显著性检验结果

样本期	预测期限	NW Holm p 值	Circular Holm p 值	Stationary Holm p 值	5%水平判断
全滚动 2023~2026	$h=5$	0.037681	0.0380	0.0402	显著
全滚动 2023~2026	$h=20$	0.000856	0.0012	0.0064	显著
2025~2026 后段样本	$h=5$	0.790147	0.7870	0.8174	不显著
2025~2026 后段样本	$h=20$	0.571143	0.5468	0.5120	不显著

注：损失差定义为 $d_t=\mathrm{Loss}_{\mathrm{candidate},t}-\mathrm{Loss}_{\mathrm{baseline},t}$，采用左尾单侧检验；Circular Block Bootstrap 与 Stationary Bootstrap 均重复 5000 次。表中 p 值均为在同一样本期、同一评价指标和同一检验方法下，对 $h=5$ 和 $h=20$ 两个预测期限进行 Holm 校正后的结果。

与 Mean Pinball Loss 相比，预测区间综合表现的统计证据相对较弱。在全滚动样本中，本文方法相对于 S_{CQR} 的 90% Interval Score 在 $h=5$ 和 $h=20$ 下均表现为数值改善，但 Newey-West 检验经 Holm 校正后的 p 值均为 0.1069，未能达到 5%的显著性水平。因此，Interval Score 结果应解释为数值上的改善，而不能表述为统计显著改善。

综上，本文预测结构在整体评价期具有平均预测优势，但这种优势并非在所有阶段均保持稳定。因而，样本外模型比较既需要考察全期平均表现，也应同时报告不同阶段的稳定性，避免将整体平均结果外推为持续性的预测优势。

4. 结论

本文以中国全国碳市场 CEA 收益率为研究对象，在严格样本外框架下开展 5 日和 20 日条件分位数预测，并利用 0.05 和 0.95 分位数构造名义 90%预测区间，进一步采用 Sequential CQR 对原始区间进行动态校准。

本文的核心发现首先体现在分期限差异化预测结构的样本外精度改善。在采用相同 Sequential CQR 校准规则的条件下，分期限预测方法在全滚动样本中的 Mean Pinball Loss 在 5 日和 20 日预测下分别较 S_CQR 基准下降约 2.13%和 7.00%，且多种稳健检验均表明这种平均损失优势具有统计显著性。这表明，不同预测期限采用差异化信息结构具有一定的预测价值，其中 20 日预测对信息结构调整表现出更明显的敏感性。

Sequential CQR 则主要发挥区间可靠性校准作用。未经校准的 5 日和 20 日预测区间存在明显欠覆盖，校准后 Coverage 分别提高至 89.27%和 88.55%，接近 90%的名义覆盖水平；与此同时，平均区间宽度相应增加，表明覆盖可靠性的改善以部分锐度损失为代价。虽然两个期限下 Interval Score 均表现为数值改

善，但相关显著性检验尚不足以认定其综合区间表现存在统计显著改善。因此，Sequential CQR 应理解为提升覆盖可靠性的后校准工具，而非本文分位数预测精度改善的主要来源。

后段样本结果表明，本文方法在 2025~2026 年的预测优势明显减弱，且未继续表现出显著预测优势。这意味着全滚动样本中的平均优势不能简单外推至所有市场阶段。未来研究可在保持严格样本外设计的前提下，进一步比较不同时间序列区间校准机制在覆盖率与锐度之间的差异。

参考文献

[1] 关于印发《全国碳排放权交易市场覆盖钢铁、水泥、铝冶炼行业工作方案》的通知[EB/OL]. https://www.mee.gov.cn/xxgk2018/xxgk/xxgk03/202503/t20250326_1104736.html, 2025-03-21.

[2] Zhu, B.Z. and Wei, Y.M. (2013) Carbon Price Forecasting with a Novel Hybrid ARIMA and Least Squares Support Vector Machines Methodology. *Omega*, **41**, 517-524. https://doi.org/10.1016/j.omega.2012.06.005

[3] Lu, H.F., Ma, X., Huang, K. and Azimi, M. (2020) Carbon Trading Volume and Price Forecasting in China Using Multiple Machine Learning Models. *Journal of Cleaner Production*, **249**, Article ID: 119386. https://doi.org/10.1016/j.jclepro.2019.119386

[4] Qi, S.Z., Cheng, S.H., Tan, X.J., Feng, S.H. and Zhou, Q. (2022) Predicting China's Carbon Price Based on a Multi-Scale Integrated Model. *Applied Energy*, **324**, Article ID: 119784. https://doi.org/10.1016/j.apenergy.2022.119784

[5] Baruník, J. and Hanus, L. (2025) Taming Data-Driven Probability Distributions. *Journal of Forecasting*, **44**, 676-691. https://doi.org/10.1002/for.3208

[6] Ren, X.H., Duan, K., Tao, L.Z., Shi, Y.K. and Yan, C. (2022) Carbon Prices Forecasting in Quantiles. *Energy Economics*, **108**, Article ID: 105862. https://doi.org/10.1016/j.eneco.2022.105862

[7] Duan, K., Ren, X.H., Shi, Y.K., Mishra, T. and Yan, C. (2021) The Marginal Impacts of Energy Prices on Carbon Price Variations: Evidence from a Quantile-on-Quantile Approach. *Energy Economics*, **95**, Article ID: 105131. https://doi.org/10.1016/j.eneco.2021.105131

[8] Koenker, R. and Bassett, G. (1978) Regression Quantiles. *Econometrica*, **46**, 33-50. https://doi.org/10.2307/1913643

[9] Romano, Y., Patterson, E. and Candès, E.J. (2019) Conformalized Quantile Regression. *Advances in Neural Information Processing Systems*, **32**, 3538-3548.

[10] Newey, W.K. and West, K.D. (1987) A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**, 703-708. https://doi.org/10.2307/1913610

[11] Politis, D.N. and Romano, J.P. (1992) A Circular Block-Resampling Procedure for Stationary Data. In: LePage, R. and Billard, L., Eds., *Exploring the Limits of Bootstrap*, John Wiley & Sons, 263-270.

[12] Politis, D.N. and Romano, J.P. (1994) The Stationary Bootstrap. *Journal of the American Statistical Association*, **89**, 1303-1313. https://doi.org/10.1080/01621459.1994.10476870

[13] Holm, S. (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, **6**, 65-70.