

基于 Copula 函数特征筛选的电力物资供应商 投标价格预测

刘 达,刘雨萌,许晓敏

(华北电力大学 经济与管理学院,北京 102206)

摘 要:分析电力物资供应商投标价格特征是招标企业估计投标价格并制定合理采购策略的基础。传统的 Spearman 秩相关系数特征筛选方法只单纯考虑了特征的数值排列顺序,不能挖掘特征变量与目标的内在关联关系,会导致预测效果差。本文在电力企业招标价格预测建模中引入 Copula 函数,通过特征变量与目标变量的联合概率分布来分析变量间的相依关系。首先确定特征变量与目标变量的边缘分布,进行 Copula 参数估计,选取合适的 Copula 函数并计算相关系数来筛选供应商投标价格的特征,并以多种预测方法进行预测来验证引入 Copula 函数后预测精度是否提升。结果表明,引入 Copula 函数进行特征筛选后,预测精度更高、效果更好。

关键词:Copula 函数;投标价格;特征筛选;预测

中图分类号:F252;F272 **文献标志码:**A **文章编号:**1002—980X(2021)10—0001—09

一、引言

采购招标是国际上广泛采用的较为完善的采购方式之一(徐永锋,2018)。现有研究表明供应商的动机、偏好和投标经验等行为模式都会影响到采购方的采购策略(Sepehr et al, 2020)。有学者尝试根据供应商的投标特点分析其行为模式(徐娟等,2014)。在采购招标过程中,投标价格是企业招标中考虑的核心因素。投标价格直接影响了供应商是否会中标,也会影响采购方的最终采购价格(Shafiei Kisomi et al, 2016)。标价格的预测是招标企业分析供应商的投标行为并制定合适的采购策略的基础(Akamp 和 Müller, 2013)。

目前投标价格预测建模大多采用时间序列模型(李永泉等,2017)。这种方法突出了时间因素在预测中的作用,但缺少对外界具体因素的考虑,不能动态反映投标价格外界影响因素的波动。若要保证预测的准确性,就必须全面考虑外界影响因素。在预测投标价格的时候,首先需要对投标价格影响因素进行特征筛选(刘玉敏等,2021)。虽然特征越多越可以全面覆盖影响因素,但过多的特征也会带来大量的冗余信息,对后续计算带来负担甚至影响预测精度。从众多特征中优选出数量少且敏感度高的代表性特征,对提高预测精度有着十分重要的意义(何胜美和谢家泉,2020)。

目前有许多特征提取方法,比如利用 Pearson 相关系数、Spearman 相关系数(张国祯等,2021)和 Kendall 秩相关系数公式来选取强相关因素。但是,计算 Pearson 相关系数的前提条件是变量的总体要符合正态分布,或为接近正态分布的单峰分布,这与本文的实际情况不符。Spearman 秩相关系数是一种无参数的检验方法,用于筛选特征时对变量分布要求不严格,但它只简单考虑了每个数值在变量内的排列顺序,只能简单度量变量之间联系的强弱。Kendall 秩相关系数方法通常用于度量分类变量的相关性,是一种非参数方法,适用于两个变量均为有序分类的情况,对于连续变量,可能会有信息损失的情况发生。

Copula 函数是一种利用联合分布来分析变量间关系的方法,既包含了边缘分布的信息也包含了两变量之间的结构信息,可以刻画变量之间的非线性相依关系(柴尚蕾和周鹏,2019),而且 Copula 函数对于变量的分布没有严格要求(Virbickaitė et al, 2020)。在引入 Copula 函数后可以更准确地分析变量之间的相关性

收稿日期:2021—06—25

基金项目:国家重点研发计划“制造业多价值链协同数据空间设计理论与方法”(2020YFB1707802);国家自然科学基金“促进大规模新能源消纳的电力市场管理研究”(71804045)

作者简介:刘达,博士,华北电力大学经济与管理学院教授,博士研究生导师,研究方向:供应链大数据分析;刘雨萌,华北电力大学经济与管理学院硕士研究生,研究方向:供应链管理;许晓敏,博士,华北电力大学经济与管理学院副教授,硕士生导师,研究方向:电力市场管理与分析、技术经济及管理。

(Chang et al, 2021),可更好地筛选特征(刘剑锋和何文祥, 2020)。Copula 函数较多应用于金融领域(郭利宁和黄运成, 2021),在金融产品定价(Garcia-Jorcano 和 Benito, 2020)和风险评估等方面应用广泛。近几年, Copula 函数的应用领域有了扩展,在气象领域(Sahana et al, 2020)、物流和供应链领域(周赛玉等, 2020)等领域也被广泛使用。

基于以上分析,本文以电力物资供应商的投标价格为研究对象,引入 Copula 函数,深入分析价格驱动因素,对投标价格进行特征筛选,并基于筛选结果预测投标价格。研究成果可以为电力物资招标策略的优化提供依据,提升招标企业采购的合理化和智能化水平。

本文第二部分介绍主要研究方法及特征筛选与价格预测模型的构建过程;第三部分进行供应商投标价格特征筛选与投标价格的预测;第四部分进行总结,结合研究结论为企业提出建议,并对后续研究进行展望。

二、研究方法及模型构建

(一)研究方法

1. Copula 函数

Copula 函数是联合分布函数的另外一种转换形式(周利国等, 2019)。假设常见的双变量联合分布函数写作 $F(X, Y)$, 当把 X 和 Y 根据其分布函数概率积分转换成 u 和 v 之后,原来的函数 F 就可以写作 u 和 v 的函数 C 。即

$$F(X, Y) = C[F_1(X), F_2(Y)] \quad (1)$$

$$C(u, v) = F[F_1^{-1}(u), F_2^{-1}(v)] \quad (2)$$

这个 C 就是 Copula 函数。常见的 Copula 函数有许多,阿基米德 Copula 函数由于构造方便,计算简单而被广泛研究和讨论,接下来对阿基米德 Copula 函数进行介绍。

假设函数 φ 为在 $[0, \infty)$ 上连续且单调递减的凸函数,满足 $\varphi(1) = 0, \varphi^{[-1]}: [0, \infty) \rightarrow [0, 1]$ 是函数 φ 的伪逆函数,按照伪逆函数的定义,

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0) \\ 0, & \varphi(0) \leq t < \infty \end{cases} \quad (3)$$

则具有 $C(u, v) = \varphi^{[-1]}[\varphi(u) + \varphi(v)]$ 形式的 C 为阿基米德 Copula 函数,其中函数 φ 称为生成元。如果 $\varphi(0) = \infty$, 此时, $\varphi^{[-1]} = \varphi^{-1}$, $C(u, v) = \varphi^{-1}[\varphi(u) + \varphi(v)]$ 称为严格的阿基米德 Copula 函数。阿基米德 Copula 函数模型有许多种,常用的有三种,见表 1。对阿基米德 Copula 函数来说,相关性的度量可以使用尾部系数。

表 1 常用阿基米德 Copula 函数

Copula	$C(u, v)$	$\varphi(t)$	尾部系数
Frank	$\frac{1}{\alpha} \ln \left(1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right)$	$-\ln \left(\frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1} \right)$	无
Clayton	$\max \left[(u^{-\beta} + v^{-\beta} - 1)^{-1/\beta}, 0 \right]$	$\frac{1}{\beta} (t^{-\beta} - 1)$	$2^{-1/\beta}$
Gumbel	$\exp \left\{ - \left[(-\ln u)^{1/\gamma} + (-\ln v)^{1/\gamma} \right]^\gamma \right\}$	$(-\ln t)^\gamma$	$2 - 2^{1/\gamma}$

对于不同的 Copula 函数来说,其参数的含义和取值范围不同,不能直接进行比较,可以使用 Spearman 秩相关系数来进行相关性的度量。

Spearman 秩相关系数反映了变量之间联系的强弱,计算公式如式(4)所示。在引入 Copula 函数后,可假设随机变量 X, Y 的边缘分布分别为 $F(X)$ 和 $G(Y)$, 设 $u = F(X)$ 和 $v = G(Y)$, 对应的 Copula 函数为 $C(u, v)$, 其中 $u, v \in [0, 1]$, 则 Spearman 秩相关系数(ρ)计算公式如式(5)所示(高雅, 2017)。

$$\rho = \frac{\sum_{i=1}^n (R_i - \bar{R})(Q_i - \bar{Q})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (Q_i - \bar{Q})^2}} \quad (4)$$

$$\rho = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3 \quad (5)$$

根据相关性系数 ρ 可以确定两者间的相关程度,相关性系数 ρ 越大,变量间相关性越强。 ρ 值用来评估相关程度计算结果的显著程度,如果计算出的 ρ 值比较小,则可说明两变量间存在显著的相关关联。一般情况,选择 0.05 作为阈值, $\rho < 0.05$ 则说明存在显著的相关性。

为了评价 Copula 模型的优劣,可以将原始数据观察得到的经验 Copula 函数和计算得到的 Copula 函数进行比较,计算两函数的欧氏距离。欧式距离越小表明,构建的模型拟合原始数据的效果越好,模型越佳。假设经验 Copula 函数为 $\hat{C}(u, v)$, 计算得到的各个因素的 Copula 函数为 $\hat{C}_n(u, v)$, 那么欧氏距离 d^2 的计算公式如式(6)所示。

$$d^2 = \sum_{i=1}^n \left| \hat{C}_n(u_i, v_i) - \hat{C}(u_i, v_i) \right|^2 \quad (6)$$

2. 预测模型

(1)多元线性回归预测。多元线性回归预测是指用两个或两个以上的影响因素作为自变量来解释因变量的变化。当多个自变量与因变量之间是线性关系时,所进行的预测分析就是多元线性预测(Mumtaz et al, 2020)。

设 y 为因变量, $x_1, x_2, \dots, x_k$ 为自变量,则多元线性预测模型如式(7)所示。

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + e \quad (7)$$

其中: b_0 为常数项; $b_1, b_2, \dots, b_k$ 为系数; e 为随机误差项。

(2)随机森林预测。随机森林是基于分类树和袋装法的组合方法。袋装法是一个简单而强大的集成、聚合算法,袋装法的基本原理是将多个独立的分类树的预测结果进行平均聚合。随机森林预测是一种利用集成学习思想建立多个互不影响的决策树集合来解决预测问题的机器学习方法,该方法可以减少预测结果的方差(方若男和骆品亮, 2020)。随机森林预测算法的基础是单一决策树,但同时它也是对单一决策树算法的延伸和改进。随机森林算法训练速度快,能够处理高维度的数据且对数据的适应能力强,且不容易出现过拟合的现象(Weiran et al, 2018)。各个子节点不纯度的加权和 $G(x_i, v_{ij})$ 的计算公式如式(8)所示。

$$G(x_i, v_{ij}) = \frac{n_{\text{left}}}{N_s} H(X_{\text{left}}) + \frac{n_{\text{right}}}{N_s} H(X_{\text{right}}) \quad (8)$$

其中: $H(X)$ 为衡量节点不纯度的函数; x_i 为某一个切分变量; v_{ij} 为切分变量的一个切分值; n_{left} 和 n_{right} 分别为切分后左子节点的训练样本个数和右子节点的训练样本个数; N_s 为当前节点所有训练样本个数; X_{left} 和 X_{right} 分别为左右子节点的训练样本集合; $G(x_i, v_{ij})$ 为衡量节点不纯度的函数,预测时一般采用不同的不纯度函数为均方误差(MSE),即针对某一切分点,计算公式如式(9)所示

$$G(x, v) = \frac{1}{N_s} \left[\sum_{y_i \in X_{\text{left}}} (y_i - \bar{y}_{\text{left}})^2 + \sum_{y_j \in X_{\text{right}}} (y_j - \bar{y}_{\text{right}})^2 \right] \quad (9)$$

(3)支持向量机预测

支持向量机预测是人工智能领域中一种基于统计学习的机器学习方法(王玲等, 2020),核心的思想是找到一个回归平面,让一个集合的所有数据到该平面的距离最近(Durgam et al, 2020)。

给定训练数据 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, 希望得到一个回归模型 $f(x) = w^T x + b$ 使得 $f(x)$ 与 y 尽可能接近, w 和 b 是模型参数。支持向量机回归(Support Vector Machine-Regression, SVR)问题可以写成

$$\min_{w, b} \frac{1}{2} \|w\|^2 + D \sum_{i=1}^m l_{\epsilon} [f(x_i) - y_i] \quad (10)$$

其中: D 为正则化常数; l_{ϵ} 为损失函数(ϵ 为误差),

$$l_{\epsilon}(z) = \begin{cases} 0, & |z| < \epsilon \\ |z| - \epsilon, & \text{其他} \end{cases} \quad (11)$$

引入松弛变量、拉格朗日乘子($\hat{\alpha}_i$ 和 α_i)后求偏导,得到 SVR 模型 $f(x)$ 如式(12)所示。

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) x_i^T x + b \quad (12)$$

3. 均方误差

均方误差(MSE)是常用的误差衡量指标,计算方法如式(13)所示, y_i 为实际值, $\hat{y}_i$ 为预测值。 MSE 值越小,说明预测模型描述实验数据具有更好的精确度。

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (13)$$

(二)特征筛选与预测模型

根据上述理论构建基于Copula函数的投标价格特征筛选模型。首先确定潜在变量的边缘分布;然后进行Copula参数估计,并结合经验Copula函数计算出欧氏距离,选取欧氏距离最小的Copula函数进行相关系数的计算;选取潜在变量中的强相关变量作为筛选出的特征变量;最后使用了多种预测方法进行预测并对比了预测精度。图1为技术路线图。

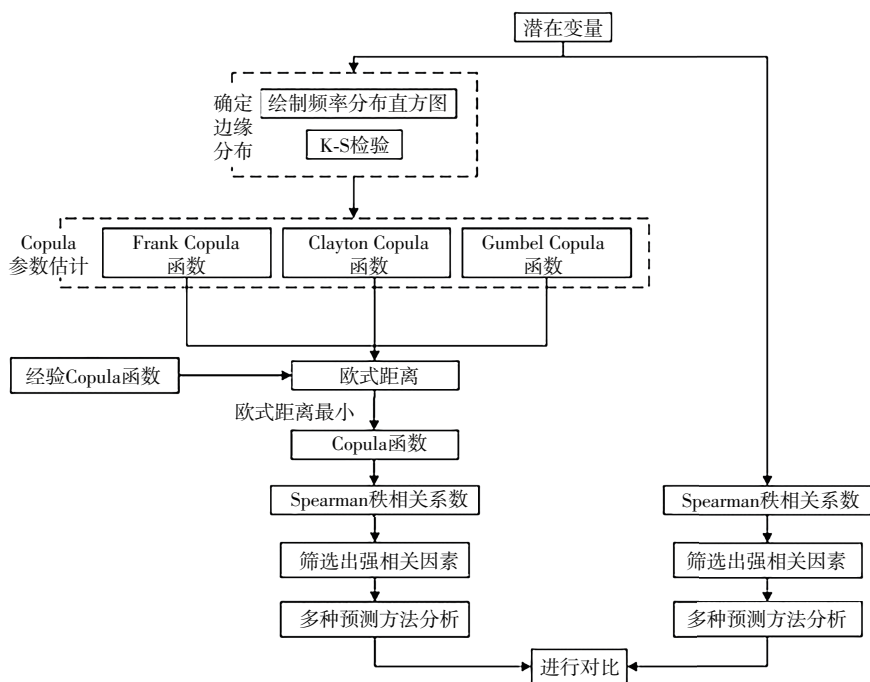


图1 技术路线图

三、投标价格特征筛选及预测

本节将影响投标价格的外部因素作为特征筛选模型的潜在变量,利用特征筛选模型对潜在变量进行特征筛选,筛选出强相关变量作为筛选出的特征变量,并用这些特征变量对投标价格进行多种方法的预测。

(一)投标价格基本情况和数据预处理

1. 投标价格基本情况

在计算过程中涉及的投标价格数据来自某电力企业2015—2020年上半年共37次招标采购活动的记录。整理后绘制了如图2所示的各年份投标价格及其变化情况。总体来说,投标价格均值处于上升趋势,但不平稳,每年的不同批次的投标价格均值也有波动。投标价格均值不稳定,招标企业在进行招标采购活动时难以预测价格变化,所以进行投标价格特征筛选及合理预测投标价格对企业来说是至关重要的。

2. 数据预处理

由于收集到的原始数据的量纲不同,数值数量级也有较大的差别,直接利用原始数据,无法进行比较,需要对数据进

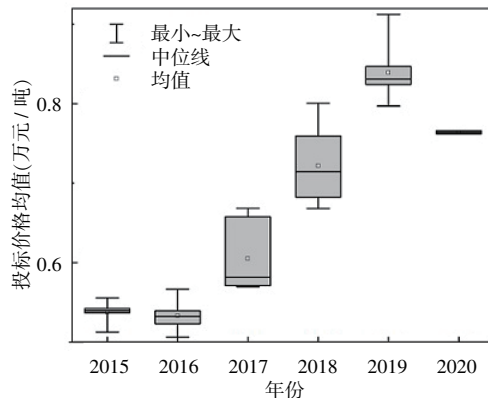


图2 投标价格变化图

行预处理。本文对所有的数据的预处理采用极差归一化变换,假设数据记为 h ,归一化公式如式(14)所示。

$$h' = \frac{t - \min(h)}{\max(h) - \min(h)} \quad (14)$$

经过极差归一化后,所有数值都在0~1之间。

(二)潜在变量分析

供应商在报价时主要考虑的是生产成本,所以在选择投标价格的潜在变量时,要聚焦于钢材和锌锭价格等原材料产业链相关因素。与原材料相关的因素选取了5个。

此外,供应商的报价行为也与全国的经济形势有关,所以考虑中国制造业采购经理指数(purchasing manager's index, PMI),它涵盖生产与流通、制造业与非制造业等领域。PMI的“荣枯分水线”为50,大于50意味着增长,数值越大,增长速度越大;小于50意味着减少,数值越小,减少速度越大;50表示持平。本文选取了5个变量,其中在手订单指数、原材料库存指数都是正向指数,供应商配送时间指数是逆向指数,供应商配送时间值越大,交货时间越慢,表明经济越不活跃。选择的潜在变量见表2。

表2 潜在变量表

变量	x_1	x_2	x_3	x_4	x_5
变量含义	角钢价格	锌锭价格	粗钢产量当期值	钢材产量当期值	钢筋产量当期值
变量	x_6	x_7	x_8	x_9	x_{10}
变量含义	制造业采购经理指数	在手订单指数	主要原材料购进价格指数	原材料库存指数	供应商配送时间指数

注:数据来自我的钢铁网、上海有色网、金投网和国家统计局网站。

(三)特征筛选

根据本文第二节中的建模步骤,在数据归一化的基础上,运用Copula模型计算变量之间的相关性。

1. 举例计算

以投标价格均值(y)与粗钢产量当期值(x_3)为例,分析步骤如下。

(1)判断变量分布。首先确定两变量的边缘分布。两变量的直方图如图3和图4所示,根据图像可初步判定变量 y 和 x_3 不服从标准正态分布及类似于正态分布的单峰分布。

其次,对 y 与 x_3 分别采用K-S检验(Kolmogorov-Smirnov检验)进行正态性检验,原假设 $h=0$ 为变量的分布情况满足正态分布, $h=1$ 为备择假设,检验结果见表4。

根据表中的结果可以看出,两变量的精确 p 值均小于0.01,均拒绝原假设,即分布都非标准正态分布。这时,可以采用近似总体分布函数和核光滑方法估计总体分布。变量 y 和 x_3 的经验分布函数和核分布估计图如图5和图6所示。

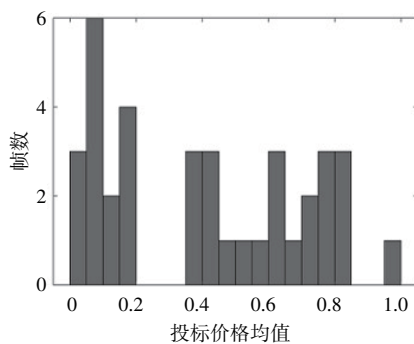


图3 投标价格均值直方图

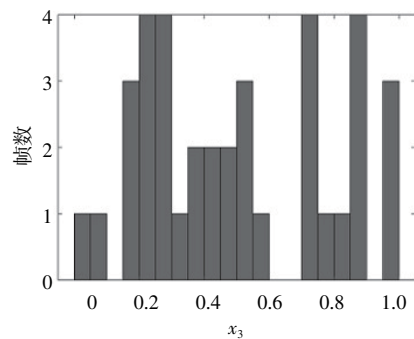


图4 粗钢产量当期值直方图

表4 K-S检验结果

	投标价格均值(y)	粗钢产量当期值(x_3)
h	1	1
p 值	7.85×10^{-9}	2.28×10^{-9}

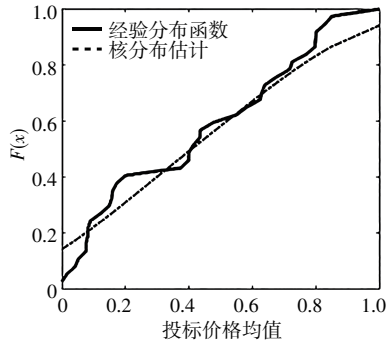


图 5 y 经验分布函数和核分布估计图

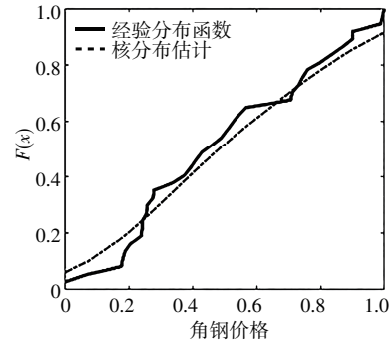


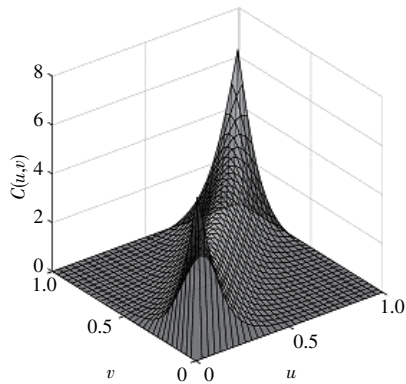
图 6 x_3 经验分布函数和核分布估计图

(2)选择适当的 Copula 函数并计算相关系数。分别对三种 Copula 函数进行参数估计再将参数值带入到表 1 的函数结构中,得到相应的 Copula 函数,结果见表 5,三种 Copula 函数的分布函数和密度函数如图 7 所示。

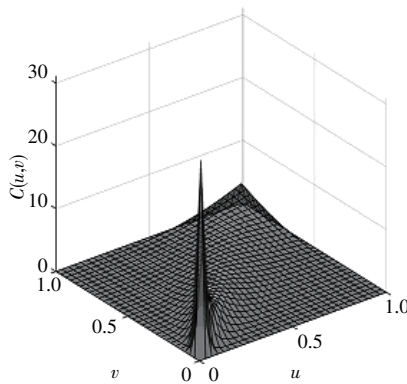
为了选择出较为合适的函数类型,接下来对三种函数的估计结果进行检验。根据原始数据观察得到的经验 Copula 函数,如图 8 所示。计算经验 Copula 函数与上文参数估计计算出的 Copula 函数的欧氏距离,结果见表 6。

表 5 模型估计结果

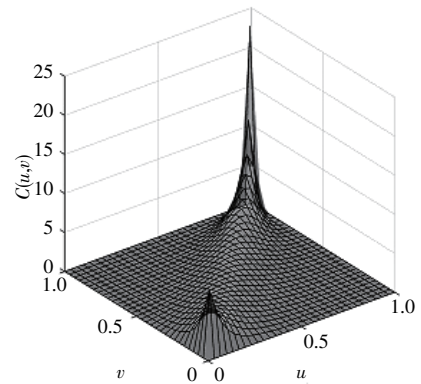
模型	参数	函数
Frank	11.1695	$\frac{1}{11.1695} \ln \left[1 + \frac{(e^{-11.1695u} - 1)(e^{-11.1695v} - 1)}{e^{-11.1695} - 1} \right]$
Clayton	3.9425	$\max \left[\left(u^{-3.9425} + v^{-3.9425} - 1 \right)^{-1/3.9425}, 0 \right]$
Gumbel	3.4805	$\exp \left\{ - \left[\left(-\ln u \right)^{1/3.4805} + \left(-\ln v \right)^{1/3.4805} \right]^{3.4805} \right\}$



(a) Frank Copula 密度函数



(b) Clayton Copula 密度函数



(c) Gumbel Copula 密度函数

图 7 三种 Copula 密度函数图

从结果中可以看出,Gumbel Copula 函数条件下的欧氏距离最小,说明该函数在三种函数中可以较好地描述变量 y 和 x_3 之间的相关结构,并且欧氏距离小于 0.1,模型评价效果较好。根据该变量的 Gumbel Copula 函数,计算 Spearman 秩相关系数,结果为 0.8856,大于 0.85,说明 x_3 和 y 之间具有很强的相关性。

2. 全变量特征筛选

根据本文设计的分析模型和上述计算过程,对剩下的变量分别进行计算,得到投标价格均值(y)与 10 个外部因素的 Copula 模型结果,见表 7。此外,表 7 中还展示了引入 Copula 函数前,直接计算 Spearman 秩相关系数的方法的变量筛选情况。在进行筛选时,引入 Copula 函数前,选择秩相关系数大于 0.5 且 $p < 0.05$ 的变量;同样的,引入 Copula 函数后,选取相关系数大于 0.5 的变量且 $p < 0.05$ 的变量。

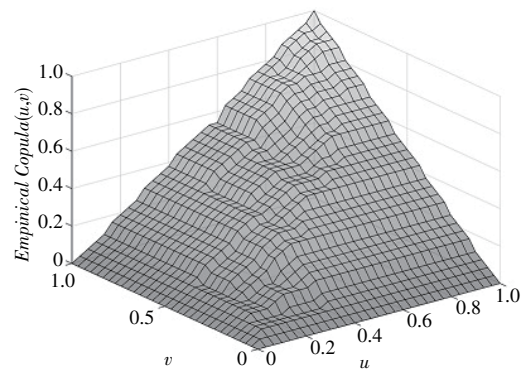


图 8 经验 Copula 分布函数图

表 6 各个模型欧氏距离

模型	Frank	Clayton	Gumbel
欧氏距离	0.0146	0.0221	0.0136

表7 潜在变量Copula函数计算结果

变量	模型	参数	欧氏距离	尾部系数	p	引入后相关系数	变量筛选	引入前相关系数	变量筛选
x_1	Clayton	2.44	0.09	0.75	0.00	0.74	√	0.65	√
x_2	Clayton	1.31	0.13	0.59	0.03	0.56	√	0.36	
x_3	Gumbel	3.48	0.01	0.78	0.00	0.89	√	0.86	√
x_4	Frank	2.62	0.01	—	0.05	0.40		0.33	
x_5	Frank	5.34	0.03	—	0.00	0.67	√	0.6	√
x_6	Clayton	0.00	0.11	0	0.08	1×10^{-6}		-0.29	
x_7	Frank	-1.49	0.07	—	0.42	-0.24		-0.14	
x_8	Frank	-0.92	0.05	—	0.69	-0.15		-0.07	
x_9	Clayton	0.00	0.05	0	0.30	1×10^{-6}		-0.12	
x_{10}	Frank	-1.29	0.05	—	0.20	-0.21		-0.22	

注:“—”为无,“√”为选择该变量。

针对所有的因素都选择出了最适合的Copula函数模型,部分因素最适合的模型是Frank Copula函数,该函数没有尾部系数,所以计算了所有因素的秩相关系数。从结果中可以看出,投标价格均值(y)与角钢价格(x_1)、镀锌价格(x_2)、粗钢产量当期值(x_3)和钢筋产量当期值(x_5)的相关系数大于0.5,具有较强的相关性。这4个因素是与原材料相关的,且为正相关,说明供应商的投标价格高低主要由成本来决定,生产成本越高,投标价格越高。

与PMI指数相关的5个因素与投标价格均值相关性较小。在手订单指数(x_7)、主要原材料购进价格指数(x_8)和供应商配送时间指数(x_{10})的秩相关系数为负数,但是按照常识来看,生产制造企业在手订单数增加时会增加出厂价;主要原材料购进价格上升时,生产成本会上升,出厂价也会随之上升;配送时间越长,产品价格也会增长。这三个因素的相关性似乎与实际情况不符。问题的原因在于PMI指数的度量方式上,PMI相关指数的增加或减少并不能直接与因素的涨跌挂钩,要结合具体数值范围来判断。

(四)预测对比分析

使用第二章中介绍的多元线性预测、随机森林预测和支持向量机预测方法,利用归一化后的投标价格均值(y)和引入Copula函数前后筛选的因素(表7)分别进行预测计算。37组数据中,使用30组数据作为训练集,剩余7组数据作为预测集。同时为了避免随机因素的影响,采用多次随机抽样的方法确定训练集和预测集。为了更直观地进行比较,三种预测方法使用相同的训练集和预测集,总共进行了5次随机抽样。

图9展示的是引入Copula函数前后,三种预测方法的预测情况。根据结果可以看出,引入Copula函数后,三种预测方法的预测值均与实际值更加接近。

此外,还利用MSE计算公式,根据预测结果计算了三种预测方法的预测值与实际值的误差情况,如图10所示。从图中可以看出,在引入Copula函数进行特征筛选后,三种预测方法的MSE整体上均有降低,5次抽样的MSE平均值也均有降低。说明引入Copula函数进行特征筛选后,三种预测方法的预测精度均变高,三种预测方法均可以更好地预测投标价格。

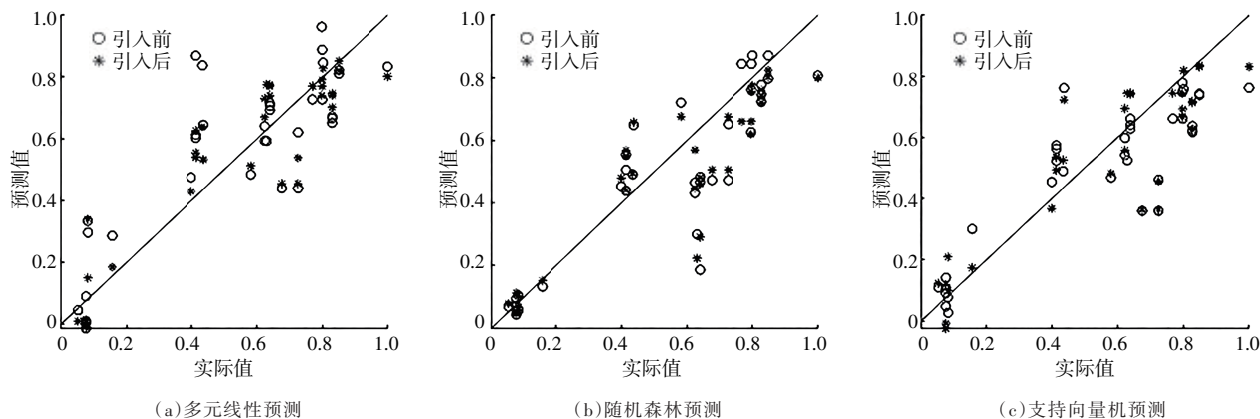


图9 三种预测方法预测情况

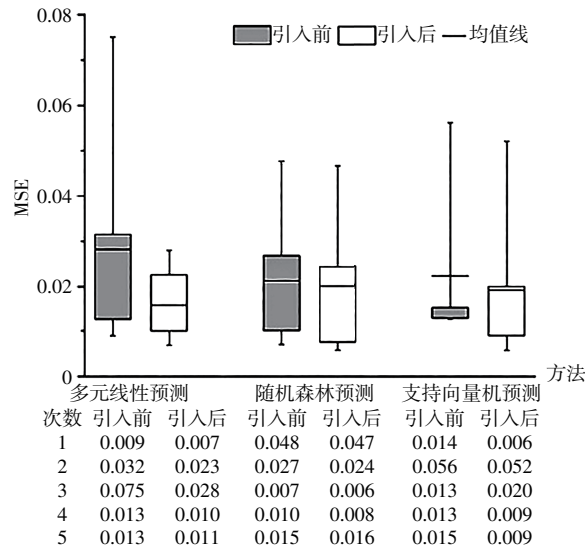


图 10 引入 Copula 前后特征筛选预测误差情况

四、结论与展望

筛选电力物资供应商投标价格影响因素,是预测投标价格及制定采购策略的基础。本文主要针对电力物资采购中的投标价格影响因素筛选进行研究,以期深入探究价格的驱动因素和更好地对投标价格进行预测。引入了 Copula 函数并计算相关系数,对投标价格的影响因素进行特征筛选。筛选出角钢价格(x_1)、镀锌价格(x_2)、粗钢产量当期值(x_3)和钢筋产量当期值(x_5)这 4 个变量。这 4 个变量均为与原材料相关的因素,所以与投标价格与原材料的相关性较高。多次随机选取相同的训练集和预测集,使用多元线性预测、随机森林预测和支持向量机预测三种预测方法,分别计算了引入 Copula 函数进行特征筛选前后的预测结果,并计算了误差。结果显示,引入 Copula 函数进行特征筛选后,三种预测方法的 MSE 有不同程度的减小,预测精度更高。说明通过引入 Copula 函数进行特征筛选,可以更好地进行特征筛选,提高投标预测的准确性,可以为电力物资招标策略的优化提供依据。

对于电力企业来说,根据本文的研究结论,在进行电力物资采购时可以事先收集与原材料产业链相关的关键指标的资料与数据,根据历史数据提前分析影响因素在围投标期中的变化趋势,来预测招标采购时供应商的投标价格及投标价格的变化。根据预测结果,可以预估供应商投标情况,有助于及时调整招标采购策略。此外,合理运用本文提出的模型和思路对供应商投标价格进行特征筛选与预测,可以帮助电力企业制定合理有效的招标采购策略。

在进行特征筛选时,利用 Copula 函数进行特征筛选的方法,限制条件少,预测精度高,在后续的研究和其他特征筛选分析中可以类比使用。

本文在进行投标价格特征筛选分析时,使用的数据是与招投标活动同期的统计数据,在后续的研究中可以考虑外部环境因素对于供应商投标价格影响的滞后性,设立提前期进行计算,这样的分析更具有全面性。

参考文献

- [1] 柴尚蕾,周鹏,2019. 基于非参数 Copula-CVaR 模型的碳金融市场集成风险测度[J]. 中国管理科学, 27(8): 1-13.
- [2] 方若男,骆品亮,2020. 基于随机森林的第三方支付违规风险预警研究[J]. 技术经济, 39(9): 11-21.
- [3] 高雅,2017. 基于 Copula 函数京津冀区域货运量相关性研究[D]. 北京: 华北电力大学(北京).
- [4] 郭利宁,黄运成,2021. 库存对石油期货市场联动关系有影响吗?——基于随机 Copula 模型的分析[J]. 中央财经大学学报(1): 86-95.
- [5] 何胜美,谢家泉,2021. 超高维多类别判别分析的特征筛选方法研究[J]. 数理统计与管理(4): 1-13.
- [6] 李永泉,张双美,杨康,2017. 建筑工程并行项目钢铁最优采购策略研究[J]. 商业研究(2): 143-150.
- [7] 刘剑锋,何文祥,2020. 国际原油市场与中国股票行业板块之间的相关性研究——基于 Copula 模型的方法[J]. 金融理论与实践(6): 79-85.
- [8] 刘玉敏,李洋,赵哲耘,2021. 基于特征选择的 RF-LSTM 模型成分股价格趋势预测[J]. 统计与决策, 37(1): 157-160.

- [9] 王玲, 吴鸿亮, 彭道鑫, 2020. 新一轮输配电价监管周期下的电网投资规模预测——基于优化的支持向量机模型[J]. 技术经济, 39(10): 45-53.
- [10] 徐娟, 汪小京, 刘志学, 2014. 多供应商环境下的最优采购策略[J]. 系统管理学报, 23(6): 883-890.
- [11] 徐永锋, 2018. 不同采购模式下的配套部件采购策略[J]. 系统管理学报, 27(6): 1176-1183.
- [12] 张国祯, 李丹丹, 岳永丽, 等, 2021. 兰州西固区PM_{2.5}和烷醇类分布特征及其与气象因素相关性分析[J]. 兰州理工大学学报, 47(2): 65-71.
- [13] 周利国, 何卓静, 蒙天成, 2019. 基于动态 Copula 的企业集团信用风险传染效应研究[J]. 中国管理科学, 27(2): 71-82.
- [14] 周赛玉, 王长军, 邵栗, 2020. Copula-CVaR下考虑随机成本和需求的单周期订货决策研究[J]. 运筹与管理, 29(4): 138-146.
- [15] AKAMP M, MÜLLER M, 2013. Supplier management in developing countries [J]. Journal of Cleaner Production, 56: 54-62.
- [16] CHANG L, ZHANG L, XU X, 2021. Correlation-oriented complex system structural risk assessment using Copula and belief rule base[J]. Information Sciences, 564: 220-236.
- [17] DURGAM S, BHOSALE A, BHOSALE V, et al, 2020. Support vector regression method for predicting temperatures of heat sources cooled by forced convection in a horizontal channel[J]. Thermal Science and Engineering Progress, 20: 100725.
- [18] GARCIA-JORCANO L, BENITO S, 2020. Studying the properties of the Bitcoin as a diversifying and hedging asset through a copula analysis: Constant and time-varying[J]. Research in International Business and Finance, 54: 101300.
- [19] MUMTAZ A, RAMENDRA P, YONG X, et al, 2020. Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms[J]. Renewable and Sustainable Energy Reviews, 132: 110003.
- [20] SAHANA V, SREEKUMAR P, MONDAL A, et al, 2020. On the rarity of the 2015 drought in India: A country-wide drought atlas using the multivariate standardized drought index and copula-based severity-duration-frequency curves [J]. Journal of Hydrology: Regional Studies, 31: 100727.
- [21] SEPEHR H, HUCHANG L, RUXUE R, et al, 2020. A likelihood-based multi-criteria sustainable supplier selection approach with complex preference information[J]. Information Sciences, 536(C): 135-155.
- [22] SHAFIEI KISOMI M, SOLIMANPUR M, DONIAVI A, 2016. An integrated supply chain configuration model and procurement management under uncertainty: A set-based robust optimization methodology [J]. Applied Mathematical Modelling, 40(17): 7928-7947.
- [23] VIRBICKAITÈ A, FREY C, MACEDO D N, 2020. Bayesian sequential stock return prediction through copulas [J]. The Journal of Economic Asymmetries, 22.
- [24] WEIRAN Y, ENKHJARGAL G, BUYANTUSHIG B, et al, 2018. Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city [J]. Environmental Pollution, 245: 746-753.

Tender Price Prediction of Power Materials Suppliers Based on Copula Function Feature Selection

Liu Da, Liu Yumeng, Xu Xiaomin

(School of Economics and Management, North China Electric Power University, Beijing 102206, China)

Abstract: Analyzing the characteristics of power material suppliers' tender price is the basis for purchase enterprises to estimate tender price and formulate reasonable purchasing strategy. The traditional Spearman rank correlation coefficient feature selection method only considered the numerical order of features, and cannot mine the intrinsic relationship between feature variables and targets, which will lead to poor prediction effect. Copula function was introduced into the forecasting model of material tender price of power enterprises, and the dependent relationship between variables was analyzed through the joint probability distribution of characteristic variables and target variables. Firstly, the marginal distribution of characteristic variables and target variables were determined, and the Copula parameters were estimated. Then, the appropriate Copula function was selected and the correlation coefficient was calculated to screen the characteristics of the suppliers' tender price. Finally, a variety of prediction methods were used to verify whether the prediction accuracy was improved after the introduction of Copula function. The results show that the Copula function has higher prediction accuracy and better effect.

Keywords: Copula function; tender price; feature screening; forecasting