

研 究 工 作 报 告

我国期货市场价格与 PPI 关系的实证研究

—— 基于 LASSO 模型

汇报人员：李鸣（Z0017029）、彭博

长江期货股份有限公司

鄂证监期货字[2014]1 号

2024 年 5 月 14 日

摘要

在宏观经济分析领域，期货市场的价格波动反映了市场参与者对未来经济环境的预期。本研究报告侧重于研究如何利用期货市场交易信息来预测生产者价格指数（Producer Price Index, PPI），该指数是衡量经济通胀和生产成本的重要指标。通过收集与工业相关的期货品种的收盘价及其滞后数据，本文尝试构建基于期货价格的 PPI 预测模型。

在模型构建过程中，本研究采用了 LASSO 回归分析方法，这一方法能在存在多重共线性的情况下，有效地处理变量选择问题。通过 LASSO 方法，本课题筛选出了对 PPI 具有显著预测能力的变量。基于这些变量构建的最终模型，预测准确度分别达到了 85% 和 92%，在历史数据上的回测表现出色。

这一研究结果不仅验证了期货市场价格用于宏观经济预测的有效性，而且市场分析师提供了一种前瞻性因子以更好地理解 and 预测经济走势，为政策制定者制定宏观经济政策提供了重要的依据。

关键词：期货市场价格；宏观经济预测；生产者价格指数（PPI）；LASSO 模型；预测回测准确度

目 录

第一章 PPI 的重要性及方法探究	1
1.1 PPI 的重要参考意义	1
1.2 PPI 的预测局限性	1
第二章 文献综述	2
第三章 基于期货价格的预测 PPI 模型	4
3.1 可行性论证	4
3.2 期货品种价格替代指数预测的必要性	5
第四章 运用 LASSO 模型从期货收盘价预测 PPI	7
4.1 LASSO 回归	7
4.2 数据预处理	7
4.3 实验过程及结果	8
4.3 预测结果	11
4.4 实际应用	12
4.5 结论	12
第五章 总结与展望	14
5.1 总结	14
5.2 展望	14
参考文献	15
研究报告声明	16

第一章 PPI 的重要性及方法探究

1.1 PPI 的重要参考意义

在宏观经济分析中，价格和通胀指标是评估经济健康水平的关键因素。生产者物价指数（**Producer Price Index**，以下简称 **PPI**）是衡量国内生产者在生产过程中所接收商品和服务价格变动的重要指标。**PPI** 反映了原材料价格和生产成本的变化，直接影响企业的盈利状态和定价策略，进而影响整个经济的通货膨胀水平和宏观经济状况。因此，准确预测 **PPI** 对于理解未来经济走势、制定相应的价格稳定和货币政策具有至关重要的作用。此外，**PPI** 还能提前反映消费者物价指数（**CPI**）的走势，作为连接生产端与消费端的中介指标，其预测的准确性对经济决策者和市场参与者都有重大意义。

1.2 PPI 的预测局限性

PPI 的计算通常由国家统计局或其他官方机构负责，涉及的数据类型广泛并且难以直接获取。**PPI** 调查目录五年一次的修订频次及基本原则更新使得动态跟踪工业产品市场价格的变动情况有一定偏差，这加大了独立预测实时 **PPI** 的重要性。如果能通过更透明、易于获取的数据源，如期货市场价格，来预测 **PPI**，将提高经济分析和决策的独立性。这种预测能力的提升，不仅有助于市场参与者更有效地响应经济变化，也增强了国内经济体系的适应能力和韧性。

第二章 文献综述

学界对期货市场与 PPI 的关系的研究众多，研究方法主要以运用时间序列方法分析为主，包括 ADF 单位根检验，Granger 因果关系检验，协整关系检验等等，选取的期货市场研究对象以期货价格指数为多，多为南华期货指数。刘健、丁嘉伦^[1]等运用单位根检验、VAR 模型、ECM 模型等方法对 2004-2016 年南华工业品期货价格指数收盘价与 PPI 进行实证分析，证明了 NHG 与 PPI 之间存在长期均衡关系，且 NHG 对 PPI 的走势影响显著，NHG 对 PPI 存在先导作用，并至少领先 PPI 半年以上。高天辰、高辉^[2]亦运用 Granger 因果关系检验、协整关系检验、误差修正模型及冲击反应函数等方法，研究了期货市场变量（价格指数、总成交额、总持仓量）对 16 个宏观经济变量的作用与影响，其中就包括期货价格指数对 PPI 的影响。然而，少有从预测角度给出具体途径的论述。

业界研报中对该问题研究覆盖有较为偏向实操的内容。陈光磊^[3]运用了时间序列分析方法，确定了滞后阶数，并使用 VAR 模型对未来 PPI 进行预测。证实 CRB 现货综合或 RJ/CRB 的变化在 3.5 个月后对 PPI 的影响达到最大，南华综合和 RJ/CRB 变化分别在 5 和 6 个月后对 PPI 的影响达到最大。何宁^[4]则构建了 LASSO 模型，使用生产资料价格预测 PPI，取得了不错的效果。

传统上，PPI 预测主要采用以下几种方法：

1. **工业品价格预测法**：这是一种直接利用工业品价格数据合成 PPI 的方法。其优点在于具有较高的精度和可靠性，但同时面临着确定不同工业品对 PPI 影响权重的问题。

2. **生产资料价格指数预测法**：类似于工业品价格法，这种方法通过生产资料价格指数来预测 PPI。虽然操作直观，但存在严重的共线性问题，可能导致模型过拟合。

3. **期货指数预测法**：通过期货价格指数来预测 PPI。期货价格反映了市场对未来价格的预期，具有一定的前瞻性，但该方法在调参和数据即时性方面仍存在不足。

4. **领先指标预测法**：此法选取具有领先性的经济指标，如 CRB 工业原料指数，通过确定这些指标与 PPI 之间的最优领先期来进行预

测。然而，相关性和领先期的不稳定性限制了其精度。

尽管上述方法各有优势，但也面临多个挑战：

1. **权重确定困难**：需要精确计算不同工业品在 PPI 中的权重，这在实际操作中很难做到精确。

2. **共线性和过拟合**：特别是在使用线性回归模型时，多种生产资料价格之间的高度相关性会导致模型参数估计不稳定，从而引起过拟合。

3. **缺乏前瞻性**：大部分方法依赖于即期数据，这意味着只有在数据发布后才能进行 PPI 的预测，缺乏对未来趋势的及时反应。

鉴于传统方法的局限性，本研究提出采用基于期货市场的数据、无需等待当期生产资料价格发布即可对实时 PPI 预测的新方法。通过使用期货市场公开易得的具体品种收盘价月均值预、使用具体品种而非编制好的期货指数预测，找到支撑变量，有效剔除相关性较大的无用数据，使预测更精简有效；使用 LASSO 从源数据及其滞后数据中挑选有效变量，并以传统 OLS 作为参照二次精选变量，在最大程度保留模型解释性的前提下，进一步降低了模型的多重共线性。即通过分析单个期货品种的价格变动，并结合期货市场对现货市场的价格发现功能和其领先期，增强预测的前瞻性、调参灵活性，降低共线性影响，给出对实时 PPI 数据的预测值，具有实践价值。

第三章 基于期货价格的 PPI 预测模型

传统 PPI 预测方法虽有其应用价值，但在权重确定、模型稳定性以及前瞻性方面存在不足。相比之下，基于期货市场的预测方法不仅能提供更加灵活的参数调整，还能通过其天然的数据优势，为宏观经济分析带来更为准确和及时的洞察。本章节起，将进一步探讨如何具体实施基于期货价格的 PPI 预测模型，并分析其在实际应用中的表现和潜在的改进方向。

3.1 可行性论证

期货价格通常被视为大宗商品供需关系的反映，同时也融合了市场对未来价格的预期以及宏观经济环境的影响。PPI 则是衡量一定时期内生产资料价格变动的重要指标。理论上，期货价格与 PPI 之间存在一定的关联性，因为两者都与生产资料的价格变动紧密相关。

为了验证这一假设，本研究进行了相关性分析。下图是 PPI 与本研究选定的与工业相关期货品种的相关性表格。对于单个期货品种，检验其当期相关性，并将其数据滞后最多 12 期来再次检验相关性。多数品种都至少在选取的滞后期或是当期与 PPI 呈现很强（接近或超过 0.5）的相关性。如原油在滞后 5 期呈现出 0.48 的相关性，锌在当期呈现 0.64 的相关性，石油沥青在滞后 9 期呈现出 0.64 的相关性，甲醇在当期呈现 0.64 的相关性等。这表明，用期货价格预测 PPI 符合数据逻辑。

表 1：选定各期货品种与 PPI 价格的相关性

	铜	原油	焦炭	焦煤	铝	螺纹钢	热轧卷板	锡	铁矿石	镍	锌	铅	天然橡胶	硅铁	石油沥青	燃料油	甲醇	棕榈油
相关性	0.31	0.43	0.46	0.32	0.28	0.45	0.41	0.24	-0.09	0.35	0.64	0.44	-0.15	0.47	0.39	0.38	0.64	0.17
滞后1期	0.30	0.43	0.43	0.29	0.25	0.41	0.36	0.23	-0.11	0.35	0.59	0.39	-0.25	0.45	0.41	0.33	0.60	0.13
滞后2期	0.29	0.44	0.41	0.26	0.24	0.38	0.31	0.23	-0.13	0.35	0.52	0.35	-0.32	0.42	0.45	0.28	0.56	0.10
滞后3期	0.29	0.45	0.40	0.23	0.23	0.34	0.26	0.23	-0.13	0.35	0.47	0.31	-0.37	0.39	0.49	0.25	0.52	0.08
滞后4期	0.29	0.47	0.38	0.22	0.23	0.31	0.23	0.24	-0.11	0.36	0.42	0.29	-0.38	0.36	0.54	0.23	0.49	0.07
滞后5期	0.28	0.48	0.36	0.21	0.22	0.27	0.19	0.24	-0.07	0.36	0.36	0.27	-0.38	0.31	0.57	0.21	0.43	0.06
滞后6期	0.26	0.47	0.34	0.19	0.20	0.23	0.14	0.23	-0.03	0.36	0.28	0.23	-0.40	0.24	0.59	0.18	0.36	0.04
滞后7期	0.24	0.45	0.31	0.18	0.18	0.19	0.09	0.21	0.01	0.35	0.20	0.17	-0.41	0.19	0.60	0.13	0.30	0.02
滞后8期	0.20	0.44	0.27	0.16	0.16	0.14	0.04	0.18	0.04	0.34	0.10	0.09	-0.41	0.15	0.63	0.08	0.23	0.00
滞后9期	0.17	0.41	0.22	0.14	0.13	0.08	-0.02	0.14	0.05	0.31	0.00	0.02	-0.43	0.10	0.64	0.02	0.15	-0.03
滞后10期	0.12	0.38	0.15	0.10	0.09	0.00	-0.10	0.10	0.06	0.28	-0.10	-0.05	-0.43	0.05	0.63	-0.05	0.06	-0.06
滞后11期	0.08	0.32	0.09	0.07	0.05	-0.08	-0.17	0.07	0.07	0.24	-0.20	-0.10	-0.42	-0.01	0.60	-0.11	-0.04	-0.08
滞后12期	0.03	0.27	0.01	0.02	0.01	-0.15	-0.23	0.02	0.07	0.21	-0.28	-0.13	-0.39	-0.07	0.54	-0.16	-0.13	-0.10

数据来源：长江期货，同花顺

3.2 期货品种价格替代指数预测的必要性

南华期货指数通过选择流动性良好且能够代表市场趋势的三大交易所上市商品（包括农产品、金属及能源化工类）主力合约价格来编制指数，在期货市场研究中因其简便性和有效性常被用于预测PPI。考虑到PPI受到原材料价格以及非物质成本（如人力成本）的影响，其变化趋势与南华期货指数有一定相关性。历史数据显示南华指数对PPI具有1-3个月的领先性。

如图1所示。2008年至2015年间，PPI的波动与南华期货指数大致相符，相关系数普遍在0.3左右。但2016年起，这一系数降至仅为0.23，工业品指数和金属指数与PPI的相关系数更是降至0.1以下。历史走势显示两者开始出现明显偏离。2020年二三季度，PPI短暂与南华工业品指数走势重合并齐升，但于2021年三季度再次分化。从宏观经济角度来看，2016年至2019年末，PPI变动主要受国内供给侧结构性改革和中美贸易关系的双重影响。2020年起，PPI变动则主要与国内外供应链错步疏通和逆全球化对进出口贸易格局的影响相关。依靠单一的南华期货指数进行PPI预测的方法已不再适用，需要更为精细化的分析方法。

因此，使用单一商品的期货收盘价而非综合期货指数进行PPI预测，可以更准确地反映特定原材料或产品价格变动对PPI的直接影响。通过细分市场数据，可以提高模型预测的精确性和实用性，更好地理解 and 预测宏观经济指标。

图1:2005-至今，南华商品指数与PPI走势及相关性





数据来源：长江期货，同花顺

PPI 与南华指数相关系数	商品指数	能化指数	工业品指数	金属指数
2004 年-2015 年	0.28	0.37	0.29	0.29
2016 年 1 月至今	0.23	0.18	0.08	0.07

第四章 运用 LASSO 模型从期货收盘价预测 PPI

在经济预测的实践中，准确预测生产者价格指数（PPI）对于理解和预测经济周期具有重要意义。PPI 的波动与多种宏观经济因素紧密相关，尤其是原材料价格的变动，这些价格变动往往首先在期货市场中得到体现。因此，精确把握期货市场的价格变化对于预测 PPI 具有重要价值。

本研究创新性地选用 LASSO 模型筛选变量以减轻共线性问题，基于相关期货品种收盘价预测 PPI 月同比变化的逻辑，训练出两个模型，分别达到了 85%、92% 的预测精度。

4.1 LASSO 回归

在宏观经济预测中，传统的 OLS（最小二乘法）虽广泛应用，但在面对高维数据时往往会遇到变量选择和共线性问题。

LASSO 回归（Least Absolute Shrinkage and Selection Operator），由 Robert Tibshirani 于 1996 年提出一种有偏的压缩估计方法，通过在损失函数中添加 L1 惩罚项，能有效进行变量压缩及稀疏选择，从而减少不必要的模型复杂度和过拟合的风险。LASSO 模型在约束条件下求解，可以将一些系数精确地缩减到零，实现变量的自动选择和复杂性调控。具体而言，若参数 $s=0$ ，解得系数均为 0；若 s 趋于无穷大，此时求解即为 OLS 估计；当 s 取到一个不太大的非负常数，LASSO 需要在约束条件下求解，形成模型的稀疏解（某些系数为零），多数情况下能较好解决过拟合的问题。

$$\text{Lasso: } \min \sum_{i=1}^n (y_i - \sum_{j=1}^p \hat{\beta}_j x_{ij})^2, \quad \sum_{j=1}^p |\hat{\beta}_j| \leq s$$

4.2 数据预处理

本研究认为，2015 年 11 月供给侧结构性改革的开启推动中国经济结构进入新阶段，故选取 2016 年 1 月 1 日为数据起始时间。而由于成文时 2024 年 4 月 PPI 数据尚未公布，故选取 2024 年 3 月 31 日为数据截止时间。对于期货变量，本研究选择的数据是 27 个期货品种活跃交易的收盘价数据月均值和 PPI 当月同比数据。考虑到我国期货市场仍

在不断发展完善中，时间序列较短的不相关期货品种予以剔除，纯碱、纸浆、尿素、乙二醇四个品种虽然时间序列较短但与国内工业产业链升级明显相关，本研究认为由于其格外重要，应当包含到建模选择中。

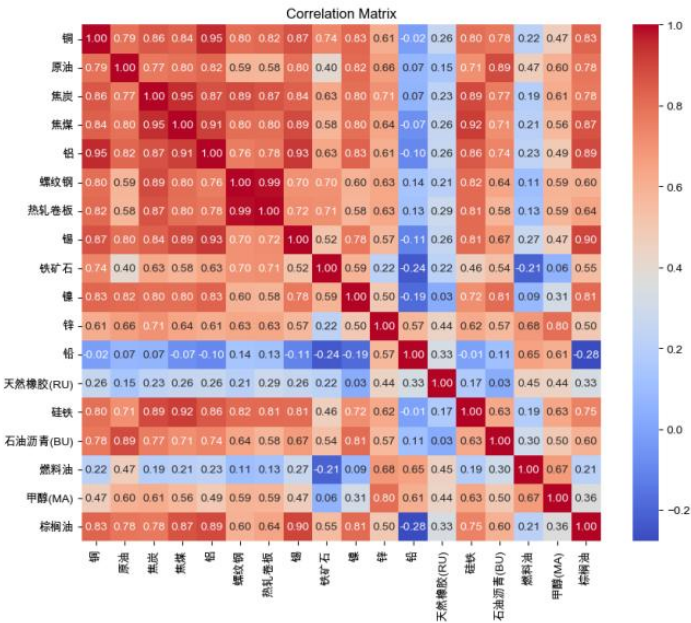
为了体现滞后项的解释性，本研究额外训练一个未加滞后项的模型用以对比。在另两个模型中，将每一个品种收盘价均滞后 1-12 期以拟合当期期货价格变动对一年以内 PPI 变动的的影响。故本研究共训练了三个模型。在第一、二个模型中，2016 月 1 月至 2024 年 3 月为数据始末时间，不包含这四个品种。同时第一个不包含滞后项而第二个包含 12 期滞后项。而在第三个模型中，将时间跨度缩短到 2019 年 12 月至 2024 年 3 月，并将这四个品种添加回训练数据中，同时包含 12 期滞后项。

对非交易日数据，采取向后填充方法来补全缺失值。将数据集进行分割，80%为训练集，20%为测试集。设置 random state=42，以固定随机分配结果。

4.3 实验过程及结果

4.3.1 第一模型

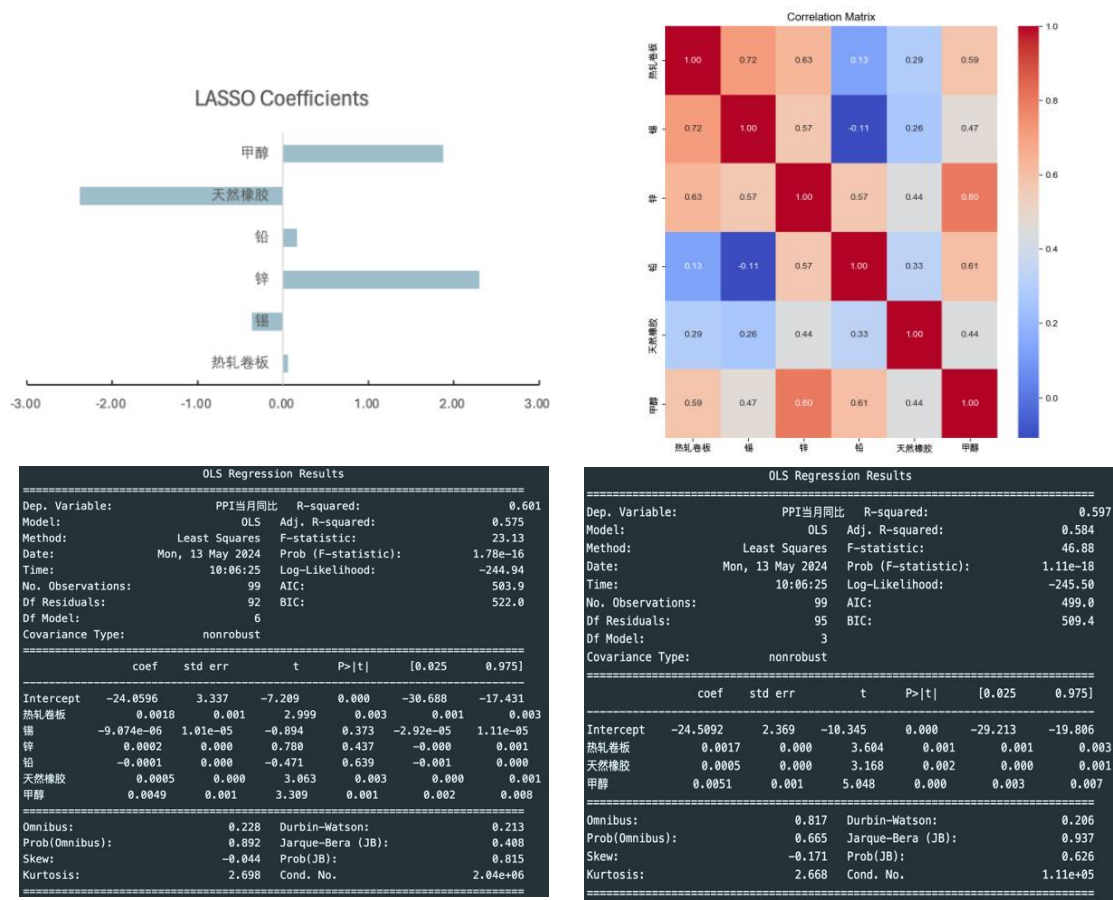
表 2：自变量相关性矩阵



数据来源：长江期货，同花顺

在第一模型中，LASSO 回归从 23 个自变量中提取出 6 个关键变量：热轧卷板、锡、锌、铅、天然橡胶、甲醇，模型解释力度为 60.1%。LASSO Coefficient 非 0 即表示变量显著，数值大小与重要程度无关。与相关性矩阵相比，LASSO 剔除了一部分的共线性问题，譬如在相关程度较高的铜、锡、原油、焦炭、焦煤和铝中仅选择了锡，同时保留了包含较多信息与其他变量相关性较弱的天然橡胶和铅。再一次用传统的 OLS 模型检验变量选择的正确性，发现其中仍存在有显著共线性关系的变量，导致 OLS 模型的系数不显著，新的相关性矩阵也证明了这一点。为此，在保证系数显著的情况下，剔除锡、锌、铅，模型的解释度也来到了 59.7%。显然，未添加滞后变量的模型对 PPI 的解释力度较差。

图 2：模型一实证结果

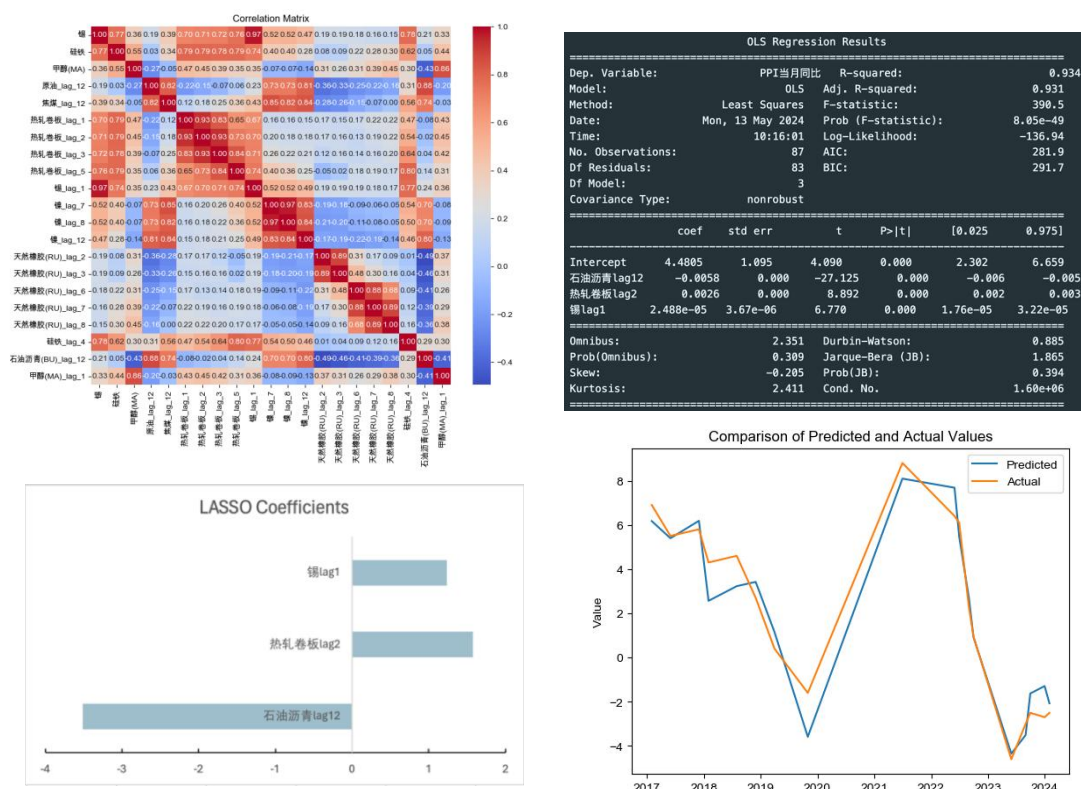


数据来源：长江期货，同花顺

4.3.2 第二模型

在第二模型中，为每个品种的期货添加 12 期滞后项，从而更全面地捕捉期货价格变动对 PPI 的影响，LASSO 回归从 234 个自变量中挑选出 21 个自变量。模型的解释度为 95.7%。但观察相关性矩阵就可以发现，这些变量中仍存在较明显的共线性。为剔除多余变量，再次用 OLS 检验并剔除不显著的变量，最终保留了三个变量：'石油沥青 lag12'，'热轧卷板 lag2'，'锡 lag1'，模型解释度来到 93.4%，保留了大部分解释力。

图 3：模型二实证结果



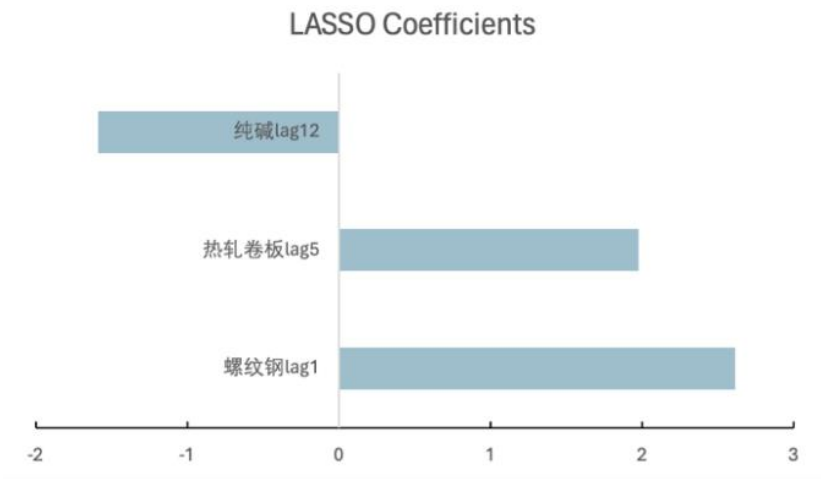
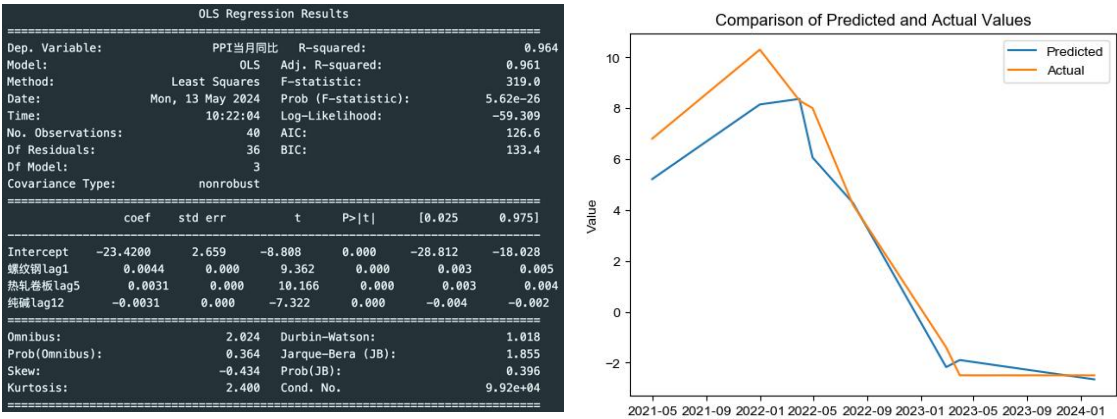
数据来源：长江期货，同花顺

4.3.2 第三模型

考虑到特定工业品种的重要性，在第三个模型中，加入纯碱、纸浆、尿素、乙二醇四个品种，并添加 12 期滞后项。LASSO 回归从 286 个自变量中剔除不显著的因子，保留了 16 个变量，模型解释力度 R^2 为 96.9%。使用 OLS 在尽可能保持模型解释力的前提下筛选掉多余的变量，最终留下螺纹钢 lag1，热轧卷板 lag5，纯碱 lag12，共三个变量，

对 PPI 解释度高达 96.4%。

图4：模型三实证结果



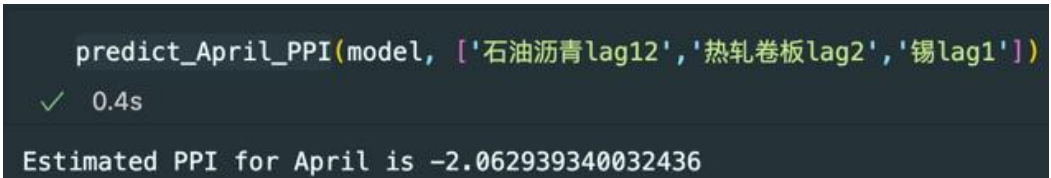
数据来源：长江期货，同花顺

4.3 预测结果

本研究应用表现较好的第二、第三个模型分别对四月 PPI 当月同比值进行预测，第二个模型运用石油沥青 lag12，热轧卷板 lag2，锡 lag1 三个变量进行预测，得到的预测结果为-2.06%。

第三个模型运用螺纹钢 lag1，热轧卷板 lag5，纯碱 lag12 三个变量进行预测，得到的预测结果为-2.44%。

图 5：模型二预测结果



数据来源：长江期货，同花顺

2024 年 5 月 11 日国家统计局发布 2024 年 4 月 PPI 数据,为-2.5%。与模型预测值基本吻合，可以体现模型应用期货收盘价及其滞后数据对 PPI 当月同比数据预测的有效性。

图 6：模型三预测结果



```
predict_April_PPI(model, ['螺纹钢lag1', '热轧卷板lag5', '纯碱lag12'])
✓ 0.7s
Estimated PPI for April is -2.44341507028148
```

数据来源：长江期货，同花顺

4.4 实际应用

4.4.1 政策制定支持

本研究的 LASSO 模型可为政府和决策者提供有力的数据支持，帮助他们理解和预测工业品价格变动对经济指数如 PPI 的影响，从而在制定财政和货币政策时更有针对性。例如，通过模型预测的 PPI 上涨可能暗示未来的通货膨胀压力，政策制定者可以提前调整利率或实施限制性措施。

4.4.2 投资策略优化

投资者和市场分析师可以通过各期货品种价格变动对 PPI 变动的贡献占比及 PPI 变动趋势来优化大宗商品配置组合。例如，若模型显示某一期货品种的价格变动占较大，结合国家政策导向及经济周期阶段，对该品种上调策略关注度并提前拟定风险管理方案。

4.5 结论

通过结合 LASSO 回归和 OLS 方法，本研究成功构建了一个提升精度的宏观经济预测模型，不仅准确地识别出影响 PPI 的关键期货品种，还验证了模型在不同时间段内的适用性和稳定性。这一模型的实

际应用展示了大数据和机器学习技术在宏观经济分析领域的巨大潜力，为政策制定和市场投资提供了重要的决策支持工具。

第五章 总结与展望

5.1 总结

本研究论文《我国期货市场价格与 PPI 关系的实证研究——基于 LASSO 模型》探讨了利用期货市场数据预测 PPI 的可行性和有效性。通过对期货市场中具体品种的收盘价月均值的深入分析，以及应用 LASSO 模型从数据中筛选出影响 PPI 的关键变量，本研究不仅增强了对 PPI 指标变动的预测能力，也为期货市场与宏观经济指标之间的关系提供了新的见解。

本研究的主要亮点在于：

1. **预测的时效性：**通过使用期货市场的日频价格数据，先于官方生产资料价格发布的情况下预测 PPI，大大提高时效性。
2. **数据获取的便利性：**期货市场的数据公开透明，易于获取，这为实时数据分析和即时经济判断提供了可能。
3. **预测模型的精简和有效性：**通过选取具体期货品种而非期货指数，本研究提升了模型的针对性和准确性。此外，LASSO 模型的使用有效地剔除了多余的、相关性高的变量，使模型更为精简和高效。
4. **降低多重共线性：**使用 LASSO 选择关键变量，并结合传统的 OLS 方法进行二次筛选，最大限度地降低了模型的多重共线性问题，同时保留了模型的高解释性。

5.2 展望

尽管本研究取得了一定成果，但未来的研究仍有广阔空间。首先，可将模型应用于不同国家和地区的 PPI 预测，考察其普适性和适应性。其次，随着数据挖掘技术的发展，可集成更多先进的机器学习技术，如深度学习模型，提升预测准确度和鲁棒性。此外，深入挖掘期货市场数据，例如考虑市场微观结构和交易行为，可能揭示更多影响宏观经济的内在机制。鉴于全球经济环境的变化，如全球化退潮、贸易政策调整及环境因素影响，未来研究应考虑这些外部变量对宏观经济指标，尤其是 PPI 的潜在影响。通过构建更复杂和全面的模型，更准确地把握经济走向，为政策制定提供科学依据。

参考文献

- [1] 刘健；丁嘉伦；张玉英．我国工业品期货价格指数与 PPI 关系的实证研究——基于 VAR 模型和 ECM 模型．金融理论与实践[J] **2017**, 75-81.
- [2] 高天辰；高辉．中国期货市场宏观经济影响的实证研究．中国证券期货[J] **2024**, 24-36.
- [3] 大宗商品价格对 PPI 的传导 宏源证券 陈光磊 2014.6.9
- [4] 基于机器学习的 PPI 预测模型——量化知宏观系列之一 华安证券 何宁 2022.3.8

研究报告声明

笔者在长江期货股份有限公司研究咨询部从事研究工作期间所提交的研究报告《我国期货市场价格与 PPI 关系的实证研究——基于 LASSO 模型》，除已明确标注和致谢的地方外，所有的观点、文字、图表及数据等均为笔者自己的研究成果。他人研究对本研究报告的启发和贡献均已作了明确的说明和致谢。

作者签名：

2024 年 5 月 14 日