

我国期货市场价格与PPI关系的实证研究

——基于LASSO模型*

中国人寿保险股份有限公司 李鸣
长江期货股份有限公司 彭博

一、PPI的重要性及方法探究

（一）PPI的重要性

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。其中，生产者物价指数（Producer Price Index，以下简称PPI）代表工业生产者出厂价格指数，是反映一定时期内全部工业产品出厂价格总水平的变动趋势和程度的相对数，包括工业企业售给本企业以外所有单位的各种产品和直接售给居民用于生活消费的产品。PPI变动反映了原材料价格和生产成本的变化。作为连接生产端与消费端的中介指标，PPI不仅能直接影响企业的盈利状态和定价策略，还能提前反映消费者物价指数（CPI）的走势。因此，准确预测PPI对于理解未来经济走势、预期或者制定相应的宏观政策具有至关重要的作用。

（二）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，如期货市场实时价格，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，也能提升国内经济体系的适应能力和韧性。

二、文献综述

学界对期货市场与PPI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（如南华期货指数）居多。刘健、丁嘉伦等运用单位根检验、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，证明了南华工业品指数与PPI之间存在长期均衡关系，且南华工业品指数对PPI的走势影响显著，南华工业品指数对PPI存在先导作用，并至少领先PPI半年以上。高天辰、高辉亦运用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

*本作品在“加快建成世界一流交易所”征文活动中荣获二等奖，完稿时间为2024年5月14日。

有从预测角度给出具体途径的论述。

业界研究报告中对该问题研究覆盖有较为偏向实操的内容。陈光磊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。

（一）传统上，PPI预测主要方法

1.工业品价格预测法

通过确定权重和调权方式构造模型，直接利用工业品价格数据（如流通领域重要生产资料价格、关键行业商品价格）合成PPI，具有较高的精度和即期性，但权重分配难以严谨。

2.生产资料价格指数预测法

通过生产资料指数、大宗商品指数、CRB现货指数等来预测PPI。方法简便、操作直观但构成关系不明确、难以调参。同时或因存在严重的共线性问题而出现模型过拟合。

3.期货指数预测法

通过期货价格指数（如南华期货商品指数、南华期货工业品指数等）来预测PPI，具有价格前瞻性，但预测精度不足。

4.领先指标预测法

通过确定具有领先性的经济指标（如CRB工业原料指数等）与PPI之间的最优领先期来进行预测。相关性和领先期的不稳定性限制了其精度。

（二）当前方法所面临的问题

1.权重确定困难

在实际操作中，很难精确计算不同工业品

在PPI中的权重。

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价格数据、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的收盘价月均值，即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，再以传统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，进一步降低模型的多重共线性。即通过分析单个期货品种的价格变动，结合期货市场对现货市场的价格发现功能及其领先期，增强预测的前瞻性、调参灵活性，降低共线性影响，从而给出对实时PPI数据的预测值，具有实践价值。

三、基于期货价格的PPI预测模型

（一）单一期货品种价格与PPI的相关性分析

期货价格不仅反映大宗商品供需关系，同时也隐含了市场对未来价格及宏观经济环境的预期。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,本研究就PPI与当前交易所已上市的工业产业链相关的期货品种进行相关性分析。对于单个期货品种,检验其当期价格与PPI当月同比的相关性,并将价格数据滞后最多12期来再次检验相关性。有相当一部分品种都至少在选取的滞后期或是当期与PPI的相关系数接近或超过0.5,即呈现明显相关性。其中,锌和甲醇的即期价格与PPI当月同比相关性较各自滞后期而言相关性最高,相关性系数均为0.64。原油在滞后5期相关系数最高,表明原油价格拟合程度最好的时滞长度为5个月。石油沥青在滞后9期呈现出最高0.64的相关性,表明石油沥青价格变动影响传导至PPI变动的最好时滞长度为9个月。以上相关性分析结果表明用期货价格预测PPI符合数据逻辑。

(二) 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,以单个商品的主力期货合约作为组合综合指数的构成部分,各个品种的权重配置主要考虑了各板块在国民经济中的影响力程度及其在期货市场中的地位水平等原则。在国内期货市场研究中因其简便性和有效性常被普遍用于预测PPI。由于PPI包含的产品定价不仅会很大程度上受到原材料价格波动的影响,同时还有其他相对稳定的其他成本(如人力成本、管理成本等)的影响,因此更高频的南华综合指数对PPI走势有一定领先性,历史数据显示南华指数对PPI具有1-3个月的领先性,但波动幅度高于PPI。

如图1所示,2008年至2015年间,PPI的变化走势大体与南华指数的走势相一致,某种程度上略微滞后于南华期货指数的变化,这体现了期货市场对PPI确实有领先的预测作用。PPI的波动与南华期货指数大致相符,相关系数大约为0.3。2016年起,PPI曲线呈现大起大落的状态,而大宗商品价格却在波动运行中保持了增长趋势。以南华工业品指数为基准,可以看到,在2020年4月左右,PPI曾短暂与之重合并得到支撑从而反弹,此后鲜有与南华指数走势相同的区间。历史走势显示两者开始出现明显偏离,相关系数降至仅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。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。

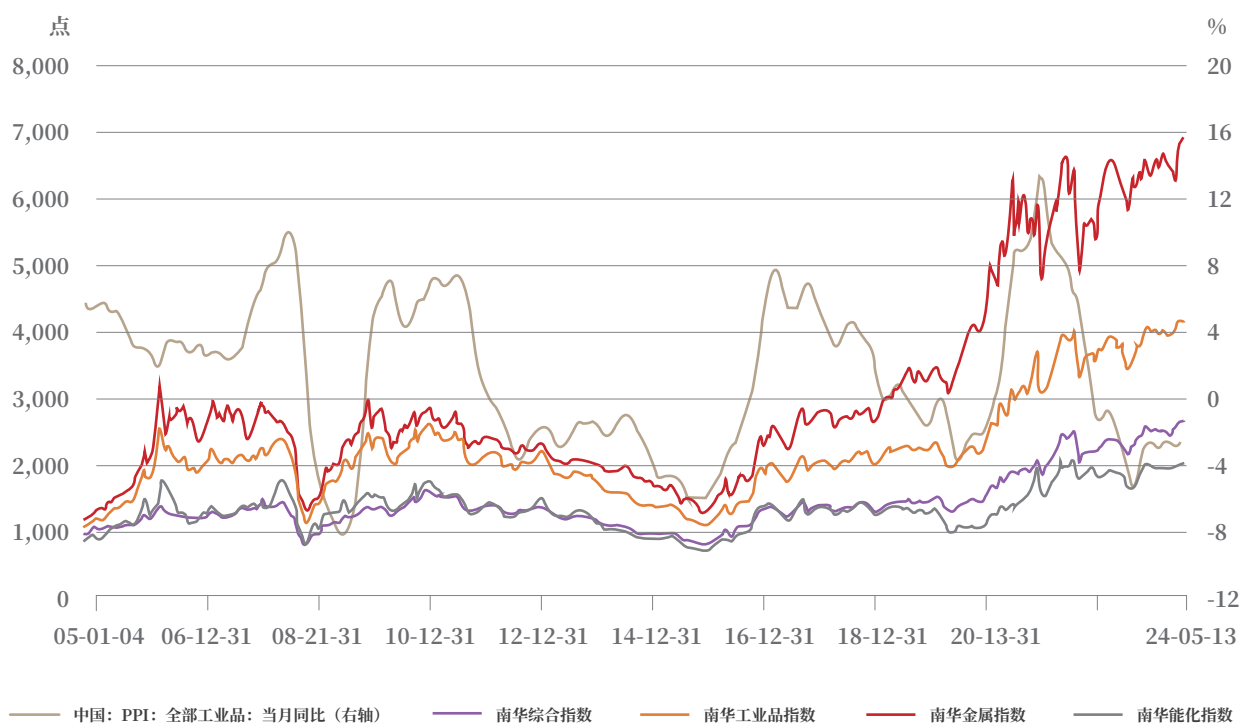
从宏观经济角度来看,2016年至2019年末,PPI变动主要受国内供给侧结构性改革和特朗普1.0的双重影响。一方面,中国在传统行业和重工业领域,如钢铁、煤炭、水泥等的去产能政策成功抑制了产能过剩问题,创造供给缺口拉动价格上涨,推动PPI上涨。另一方面,特朗普总统任期内开启逆全球化时代,出口受阻,外需减弱,抑制PPI上涨。国际油价持续走低的传导作用下,部分原材料成本持续下跌也对PPI形成了一定的抑制作用。新冠疫情冲击之下,需求下降,供给收缩,供需缺口叠加全球量化宽松,引发期货市场剧烈波动。2020年初至2021年末,全球供应链中断导致供给收缩,大宗商品价格上涨。国内前期防疫成效显著供应链率先复苏叠加全球央行天量放水带来的需求回升,推动PPI增速达到历史高位。疫情中后期通胀预期加强,中下游企业受

到需求回落、成本多方面冲击，通过“利润侵蚀工资”到“收入-消费”路径抑制内需。海外群体免疫效果显现，海外供应链复苏利空外需。PPI增速整体呈现快速下滑的趋势。

据统计数据可知，中国出口商品主要集中在机电音像设备、家具玩具等杂项、金属及其制品、纺织制品、化学及相关工业产品以及车

辆等运输设备上。所涉及的原材料端期货品种价格在对应该时段内呈现明显的需求推动型上行趋势，而其他非相关期货品种波动相对平缓或分化。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。

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



数据来源：Wind

图 1：2005 年至今，南华商品指数与 PPI 走势对比及相关系数

四、运用LASSO模型从期货收盘价预测PPI

(一) 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。

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

(二) 数据预处理

由于2015年11月供给侧结构性改革开启，PPI变动与南华期货指数走势相关性开始降低，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。考虑到我国期货市场仍在不断发展完善中，剔除上市时间不足6年、交易不活跃的能化、金属、工业品期货品种，本研究选择了23个期货品种的主力收盘价月均值数据与PPI当月同比数据进行拟合，数据来源为Wind数据库。另外，基于相关性矩阵数据发现，尽管纯碱、纸浆、尿素、乙二醇上市时长不符合条件，但自2019年末疫

情推动国内工业产业链再次升级，这四个品种与PPI当月同比值的相关系数绝对值出现明显上行。故仍将以上四个品种加入建模自变量中。

因此，本研究总共设计训练三个模型。为了体现滞后项的解释性，第一模型使用23个期货品种为自变量，不添加滞后项。第二模型在第一模型的基础之上添加每个自变量滞后1-12期的滞后项以观察期货价格变动对一年以内PPI变动的时滞效应。而在第三模型中，将数据时间跨度调整为2019年12月至2024年3月，并将纯碱、纸浆、尿素、乙二醇这四个品种加入训练数据中，同时包含各自的12期滞后项。

对非交易日数据，采取向后填充方法来补充缺失值。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。

(三) 实验过程及结果

LASSO Coefficient非0即表示变量显著，数值大小与重要程度无关。

1. 第一模型

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%。与相关性矩阵相比，LASSO剔除了一部分共线性问题，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，如表1中的第一模型（1）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，模型解释力度降至59.7%，如表1中的第一模型（2）。

表 1：第一模型回归结果

变量	第一模型 (1) PPI	第一模型 (2) PPI
热轧卷板	0.0018*** (2.999)	0.0017*** (3.604)
锡	-9.074e-06 (-0.894)	—
锌	0.0002 (0.780)	—
铅	-0.0001 (-0.471)	—
天然橡胶	0.0005*** (3.063)	0.0005*** (3.168)
甲醇	0.0049*** (3.309)	0.0051*** (5.048)
Intercept	-24.0596*** (-7.209)	-24.5092*** (-10.345)
N	99	99
R-squared	0.601	0.597

注：*** 表示在 1% 的水平上显著，括号内为 t 值。

2.第二模型

在第二模型中，为了更全面地观测期货价格变动对PPI的传导效果，每个自变量添加12期滞后项。LASSO回归从234个自变量中挑选出21个自变量，模型解释力度为95.7%。观察相关性矩阵不难发现，这些变量中仍存在较明显的共线性。为剔除多余变量，使用OLS检验并剔除不显著的变量，最终保留‘石油沥青lag12’，‘热轧卷板lag2’，‘锡lag1’三个变量，如表2第二模型的R值所示，模型解释力度为93.4%，即沥青价格传导12个月，热轧卷板价格传导2个月、锡价格传导1个月对PPI当月同比变动保留了大部分解释力。对比第一、二模型，显然未添加滞后变量的模型对PPI的

解释力度较差。

3.第三模型

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，模型解释力度为96.9%。再次使用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lag12’三个变量，如表2第三模型的R值所示，模型解释力度高达96.4%。即螺纹钢价格传导1个月，热轧卷板价格传导5个月、纯碱价格传导12个月对PPI当月同比变动保留了大部分解释力。

表 2：第二、三模型回归结果

变量	第二模型 PPI	第三模型 PPI
石油沥青 lag12	-0.0058*** (-27.125)	—
热轧卷板 lag2	0.0026*** (8.892)	—
锡 lag1	2.488e-05*** (6.770)	—
螺纹钢 lag1	—	0.0044*** (9.362)
热轧卷板 lag5	—	0.0031*** (10.166)
纯碱 lag12	—	-0.0031*** (-7.322)
Intercept	4.4805*** (4.090)	-23.4200*** (-8.808)
N	87	40
R-squared	0.934	0.964

注：*** 表示在 1% 的水平上显著，括号内为 t 值。

（四）历史回测与预测效果

由图2可知，第二、三模型历史回测中预测值与实际值拟合效果较好。使用实时大宗商品价格预测PPI可以更好反映出供给需求冲击，以及延伸推断出不同产业板块的滞后效应。应用第二、第三个模型分别对四月PPI当

月同比值进行预测，第二个模型得到的预测结果为-2.06%。第三个模型运用得到的预测结果为-2.44%。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 2：第二模型（左）、第三模型（右）实证拟合效果

（五）实际应用

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。

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比较大，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。

（六）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。

五、总结与展望

（一）总结

本论文探讨了利用期货市场价格数据预测PPI的可行性和有效性。应用LASSO回归模型从具体期货品种的收盘价月均值中筛选出影响PPI的关键变量。本研究不仅增强了对PPI指标变动的动态预测能力，也为期货市场与宏观经济指标之间的联动提供了新的视角。

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，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。

（二）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、贸易政策调整及地缘政治因素影响，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。

（责任编辑：张甜迪）

作者简介：

李鸣，国华人寿保险股份有限公司投资管理部宏观大类资产配置研究员。

彭博，长江期货股份有限公司研究咨询部研究员。