

从期货市场研判宏观经济 -从期货收盘价到PPI

摘要：商品在期货市场上的价格能够反映市场对于生产生活资料价格的一致预期，从而对宏观经济指标有一定的预测性。本文从PPI入手，试图通过工业相关期货品种收盘价和滞后数据预测PPI，利用LASSO筛选可以拟合PPI的变量，最终获得的模型预测回测准确度高达85%/92%，具有丰富的参考价值。

一、为何要预测PPI

（一）PPI的重要参考意义

价格和通胀是宏观分析研究的重要组成部分。PPI（生产者物价指数）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指标。它主要反映了原材料价格和生产成本的变动情况，是评估通货膨胀水平和宏观经济状况的关键经济指标。预测PPI有助于我们预测未来的经济走势，在价格稳定和货币政策的制定方面为我们提供参考，还可以提前反映CPI的走势，是连接生产端与消费端的中介指标，具有十分重要的意义。故在此我们选择PPI作为我们的研究对象。

（二）PPI计算方法的不透明性，数据难获得性

PPI的计算通常由国家统计局或相应的官方机构进行。这一过程涉及到的数据十分广泛而难以获得。同时PPI的计算公式和方法存在一定的不透明性，这使得独立预测PPI变得尤为重要。如果能够通过其他更透明、易于获取的数据（如期货价格）来预测PPI，那么我们将能够更独立的进行经济分析和决策，减少对官方数据发布的依赖。这种预测能力的提升，可以帮助市场更有效地响应经济变化，提高整个经济体的适应能力和韧性。

二、以往采用的预测PPI方式及存在问题

	工业品价格预测法	生产资料价格指数预测法	期货指数预测法	领先指标预测法
数据	流通领域重要生产资料价格、关键行业商品价格	生产资料指数、大宗商品指数、CRB现货指数等	南华期货商品指数、南华期货工业品指数等	期货、生产资料指数
计量方法	明晰工业品价格和PPI的构成关系，确定权重和调权方式，构造模型	线性回归	线性回归	确定领先期后，滞后相应期数后进行线性回归
预测期限	即期	即期	即期	视领先期而定，可即期可远期
预测精度	很高	较高	较低	较低
优点	精度高	方法简便，预测精度高	方法简便，期货数据有价格前瞻性	方法简便，有一定领先期，期货数据有价格前瞻性
缺点	需要自行确定权重，价格间共线性问题严重，过度依赖即期数据的发布，缺乏前瞻性	构成关系不明确，难以调参，依赖即期数据发布，缺乏前瞻性	构成关系不明确，依赖即期数据发布，精度低	构成关系不明确，领先期不稳定，相关性不稳定，精度低

PPI 预测常见的类型有四种：工业品价格预测法、生产资料价格指数预测法、期货指数预测法和领先指标预测法。

最常见的做法是使用工业品价格合成 PPI，这种做法精度高，可信度和可调试性强。但问题存在于以下三个方面：（1）需要合适的方法确定不同工业品对PPI的影响权重（2）若直接使用线性回归，由于生产资料价格种类繁多，且相互关联，模型会有很严重的共线性问题，导致过拟合。（3）全部使用即期的生产资料价格数据，要在生产资料价格发布后才可以得到预测，缺乏前瞻性。

生产资料和期货价格指数预测法颇有相似性，均使用价格指数对PPI回

归，两种方法的问题也比较类似，指数内不同品种的构成关系不明确且难以调参，且主要使用即期数据进行预测，缺乏前瞻性。但使用期货价格比使用生产资料价格对未来的PPI预测更具参考意义。

领先指标预测法，往往也使用期货或生产资料价格指数，通过挖掘符合经济运行规律的具有领先性的指标，确定最优领先期后进行回归预测，例如CRB工业原料指数领先 PPI 2-3个月，但是由于相关性和领先期不稳定，往往精度较低，同时，依然是运用指数，构成关系不明确，难以调参。

因此在此我们判断，最优解应该是使用单个期货品种的价格，并结合领先期对PPI进行预测，同时要结合对指标的筛选以尽可能减轻共线性。这样既可以兼顾调参的灵活性，又可以保证预测的前瞻性，还可以防止模型的过拟合问题，提高预测的精度。

三、期货价格预测PPI的可行性

(一) 从相关性论证可行性

期货价格往往反映了大宗商品的供需关系，市场对于大宗商品的价格预期以及整体的宏观环境，而PPI是生产者物价指数，反映了生产资料在一定时期内的价格变动。故从学理上，用期货价格预测PPI是符合我们的认知的。这一点从数据上也可以证明。下图是PPI与我们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，我们检验其当期相关性，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是符合数据逻辑的。

	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

图：选定各期货品种与PPI价格的相关性表格

(二) 使用具体期货品种而非指数预测的必要性

市面上不乏使用南华期货指数预测PPI的案例，这种方法曾经简便有效，有着不错的拟合和预测效果。

南华指数选择三大交易所上市品种中比较有代表性且具有较好流动性的商品来编制指数，包括农产品、金属以及能源化工类，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个商品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。

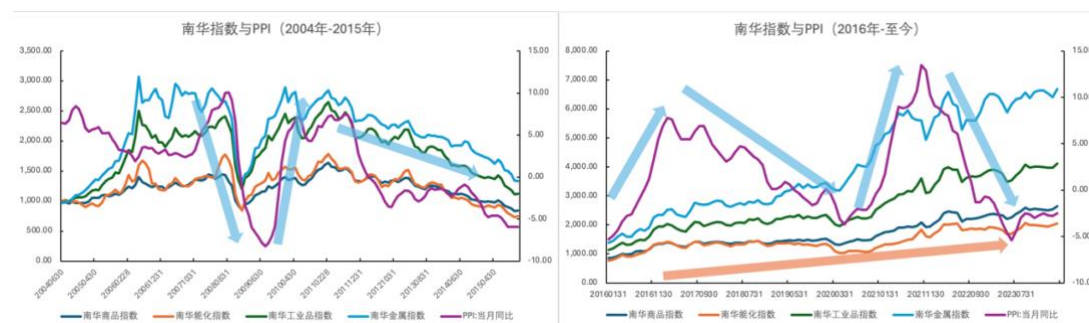
PPI 代表工业生产者出厂价格指数，是反映一定时期内全部工业产品出厂价格总水平的变动趋势和程度的相对数，包括工业企业售给本企业以外所有单位的各种产品和直接售给居民用于生活消费的产品。

由于PPI包含的产品会很大程度上受到原材料价格的影响，但其构成不止大宗商品，其所涵盖的更广泛的商品中还隐含了人力成本等相对稳定的价格因素影响，因此更高频的南华综合指数对PPI走势有一定领先性，从历史看大约领先1-3个月，但PPI的波动幅度小于南华指数。

然而，我们发现，在2016年供给侧改革前后，期货指数与CPI之间镜像变化的逻辑产生了深刻的变化。这导致这种方法过分简单化，不再符合我们预测CPI精准度的需求。下图分别是2004-2015年和2016年至今南华期货指数（商品指数、能化指数、工业品指数以及金属指数）与PPI月同比数据变化的对比图。由于南华期货指数从2004年开始发布，故选取的时间区间从2004年开始。可以看出，除去2004-2008年南华指数刚发布，期货市场亦刚起步尚未企稳的区间，从2008年到2015年，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于PPI确实有领先的预测作用。然而，2016年之后，南华期货指数的运行突然与PPI的运行趋势产生较大的偏离。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。

对此，我们可以从宏观分析的角度解释PPI的异常变化：2016年到疫情前，PPI变化主要受供给侧结构性改革与外部形势带来的供需影响强弱交替的影响。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，使产能供给不断减少，拉动价格上涨，进一步拉动PPI上涨。另一方面，国内需求趋于疲软，同时国际经济环境不确定性增加，全球经济增长放缓，国际市场需求减弱，出口受阻。这对中国的出口导向型产业造成了冲击，进一步抑制了PPI的上涨。部分原材料价格持续下跌，如国际油价的走低等，也对PPI形成了一定的抑制作用。疫情前期（2020年初-2021年末）PPI几乎马不停蹄的上涨，增速来到历史高位，这主要是由于疫情过程中供应链中断带来的生产成本上涨以及全球大宗商品价格上涨带来的带动作用。另外，国内前期的防疫成功带来了一波需求回升，加之央行实行宽松的货币政策，本时期PPI高歌猛进。进入疫情后期，由于变种毒株传播迅猛，国内防疫政策受到挑战，企业供给复苏不均的同时需求也被大力抑制，国际市场需求下降，对中国出口的影响加剧，PPI增速整体呈现快速下滑的趋势。

尽管我们可以从基本面分析的角度解释PPI的变动，但不可否认的是，从数据层面，2016年后，南华期货指数很难再较为准确的刻画PPI变动了。这一点，从PPI与各个南华期货指数的相关性也可以看出。在2015年以前，PPI与南华期货指数的相关系数普遍在0.3左右，而2016年及以后，相关系数最高就只有0.23了，工业品指数和金属指数与PPI的相关系数甚至降到了0.1以下。



相关系数	PPI与南华商品指数	PPI与南华能化指数	PPI与南华工业品指数	PPI与南华金属指数
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

四、运用LASSO模型从期货收盘价预测PPI

综合以上论述，我们创新性的选用LASSO模型筛选变量以减轻共线性问题，从具体期货品种的收盘价预测PPI的月同比变化。训练了两个模型，分别达到了85%/92%的预测精度。

（一）LASSO回归

线性回归中，经典的OLS最小二乘法无法筛选变量，容易产生共线性的问题，导致模型过拟合，预测效果变差。LASSO(Least absolute shrinkage and selection operator)是 Robert Tibshirani提出的一种有偏的压缩估计方法。在 OLS回归的系数上增加一个惩罚项，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。

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

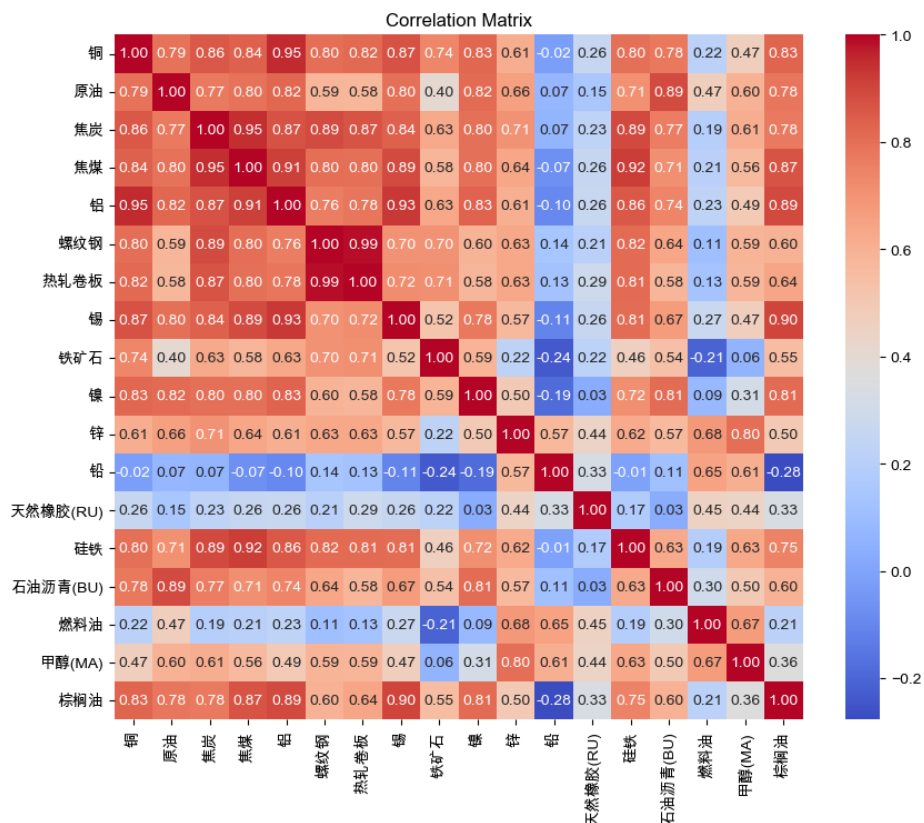
（二）数据预处理：

我们认为，2015年11月开始的供给侧结构性改革昭示着中国经济一个新阶段的开始，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，我们选择的数据是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。因为我国期货市场仍在不断发展完善中，许多期货品种近年刚刚上市，时间序列较短，缺乏参考意义。但纯碱、纸浆、尿素、乙二醇四个品种虽然时间序列较短，我们认为由于其格外重要，应当包含到建模选择中。

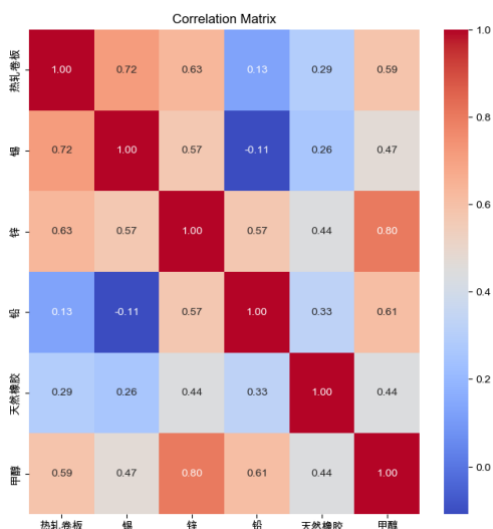
同时，为了体现滞后项的解释性，我们额外训练了一个未加滞后项的模型。在另两个模型中，我们将每一个品种收盘价均滞后1-12期以拟合当期期货价格变动对一年以内PPI变动的的影响。故我们共训练了三个模型。在第一、二个模型中，2016月1月至2024年3月为数据始末时间，不包含这四个品种。同时第一个不包含滞后项而第二个包含12期滞后项。而在第三个模型中，我们将时间跨度缩短到2019年12月至2024年3月，并将这四个品种添加回训练数据中，同时包含12期滞后项。

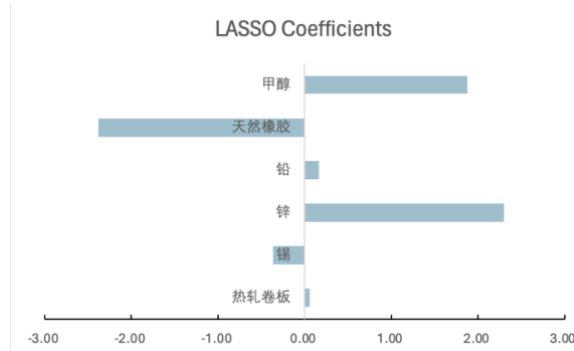
对非交易日数据，我们采取了向后填充方法来补全缺失值。将数据集进行分割，80%为训练集，20%为测试集。设置random state=42，以固定随机分配结果。

（三）实验过程及结果：



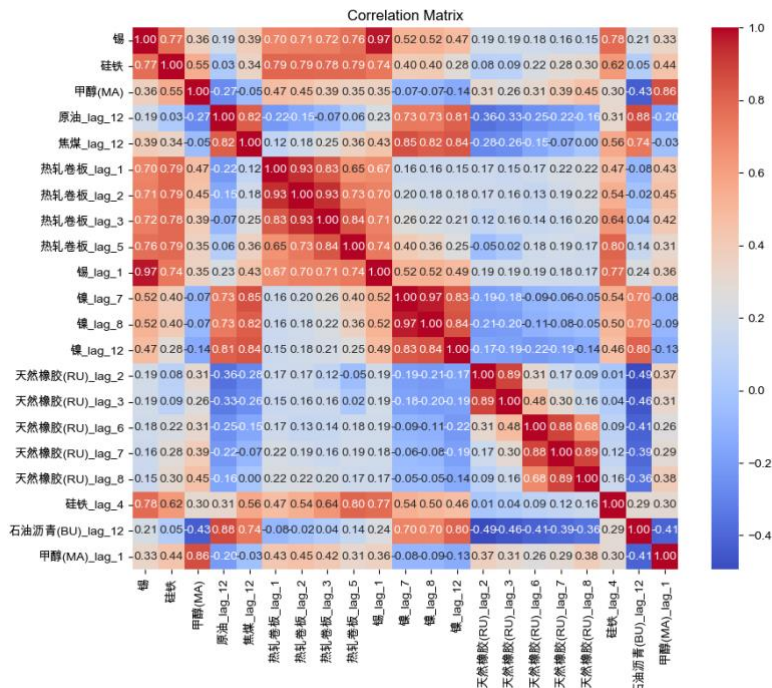
在第一个模型中LASSO回归从23个自变量中提取出6个关键变量：热轧卷板、锡、锌、铅、天然橡胶、甲醇，模型解释力度为60.1%。LASSO Coefficient非0即表示变量显著，数值大小与重要程度无关。与相关性矩阵相比，LASSO为我们剔除了一部分的共线性问题，譬如在相关程度较高的铜、锡、原油、焦炭、焦煤和铝中仅选择了锡，同时保留了包含较多信息与其他变量相关性较弱的天然橡胶和铅。我们再一次用传统的OLS模型检验变量选择的正确性，发现其中仍存在有显著共线性关系的变量，导致OLS模型的系数不显著，新的相关性矩阵也证明了这一点。为此，我们在保证系数显著的情况下，剔除锡、锌、铅，模型的解释度也来到了59.7%。显然，未添加滞后变量的模型，对PPI的解释力度较差。



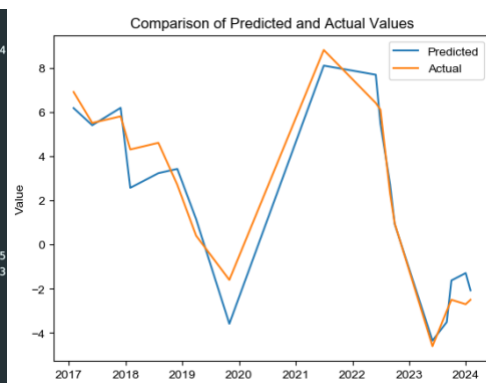


OLS Regression Results							OLS Regression Results						
Dep. Variable:	PPI当月同比	R-squared:	0.601	Dep. Variable:	PPI当月同比	R-squared:	0.597	Dep. Variable:	PPI当月同比	R-squared:	0.597	Dep. Variable:	PPI当月同比
Model:	OLS	Adj. R-squared:	0.575	Model:	OLS	Adj. R-squared:	0.584	Model:	OLS	Adj. R-squared:	0.584	Model:	OLS
Method:	Least Squares	F-statistic:	23.13	Method:	Least Squares	F-statistic:	46.88	Method:	Least Squares	F-statistic:	46.88	Method:	Least Squares
Date:	Mon, 13 May 2024	Prob (F-statistic):	1.78e-16	Date:	Mon, 13 May 2024	Prob (F-statistic):	1.11e-18	Date:	Mon, 13 May 2024	Prob (F-statistic):	1.11e-18	Date:	Mon, 13 May 2024
Time:	10:06:25	Log-Likelihood:	-244.94	Time:	10:06:25	Log-Likelihood:	-245.50	Time:	10:06:25	Log-Likelihood:	-245.50	Time:	10:06:25
No. Observations:	99	AIC:	503.9	No. Observations:	99	AIC:	499.0	No. Observations:	99	AIC:	499.0	No. Observations:	99
Df Residuals:	92	BIC:	522.0	Df Residuals:	95	BIC:	509.4	Df Residuals:	95	BIC:	509.4	Df Residuals:	95
Df Model:	6			Df Model:	3			Df Model:	3			Df Model:	3
Covariance Type:	nonrobust			Covariance Type:	nonrobust			Covariance Type:	nonrobust			Covariance Type:	nonrobust
	coef	std err	t	P> t	[0.025	0.975]		coef	std err	t	P> t	[0.025	0.975]
Intercept	-24.0596	3.337	-7.209	0.000	-30.688	-17.431	Intercept	-24.5892	2.369	-10.345	0.000	-29.213	-19.806
热轧卷板	0.0018	0.001	2.999	0.003	0.001	0.003	热轧卷板	0.0017	0.000	3.684	0.001	0.001	0.003
锡	-9.074e-06	1.01e-05	-0.894	0.373	-2.92e-05	1.11e-05	天然橡胶	0.0005	0.000	3.168	0.002	0.000	0.001
锌	0.0002	0.000	0.780	0.437	-0.000	0.001	甲醇	0.0051	0.001	5.048	0.000	0.003	0.007
铅	-0.0001	0.000	-0.471	0.639	-0.001	0.000							
天然橡胶	0.0005	0.000	3.063	0.003	0.000	0.001							
甲醇	0.0049	0.001	3.389	0.001	0.002	0.008							
Omnibus:	0.228	Durbin-Watson:	0.213	Omnibus:	0.817	Durbin-Watson:	0.206						
Prob(Omnibus):	0.892	Jarque-Bera (JB):	0.408	Prob(Omnibus):	0.665	Jarque-Bera (JB):	0.937						
Skew:	-0.044	Prob(JB):	0.815	Skew:	-0.171	Prob(JB):	0.626						
Kurtosis:	2.698	Cond. No.	2.04e+06	Kurtosis:	2.668	Cond. No.	1.11e+05						

在第二个模型中，我们为每个品种的期货添加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，我们再次用OLS检验并剔除不显著的变量，最终保留了四个变量：'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。



OLS Regression Results						
Dep. Variable:	PPI当月同比	R-squared:	0.934			
Model:	OLS	Adj. R-squared:	0.931			
Method:	Least Squares	F-statistic:	390.5			
Date:	Mon, 13 May 2024	Prob (F-statistic):	8.05e-49			
Time:	10:16:01	Log-Likelihood:	-136.94			
No. Observations:	87	AIC:	281.9			
Df Residuals:	83	BIC:	291.7			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.4805	1.095	4.090	0.000	2.302	6.659
石油沥青lag12	-0.0058	0.000	-27.125	0.000	-0.006	-0.005
热轧卷板lag2	0.0026	0.000	8.892	0.000	0.002	0.003
锡lag1	2.488e-05	3.67e-06	6.770	0.000	1.76e-05	3.22e-05
Omnibus:	2.351	Durbin-Watson:	0.885			
Prob(Omnibus):	0.309	Jarque-Bera (JB):	1.865			
Skew:	-0.205	Prob(JB):	0.394			
Kurtosis:	2.411	Cond. No.	1.68e+06			



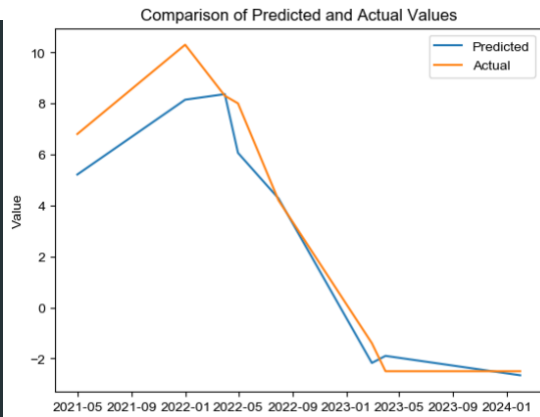
在第三个模型中，我们加入纯碱、纸浆、尿素、乙二醇四个品种，并添加12期滞后项。LASSO回归从286个自变量中剔除不显著的因子，保留了16个变量，模型解释力度 R^2 为96.9%。同样的，我们使用OLS在尽可能保持模型解释力的前提下筛选掉多余的变量，最终留下石油沥青lag12，螺纹钢lag1，热轧卷板lag5，纯碱lag12，乙二醇lag1，乙二醇lag3，共六个变量，对PPI解释度高达96.4%。

OLS Regression Results

Dep. Variable:	PPI当月同比	R-squared:	0.964
Model:	OLS	Adj. R-squared:	0.961
Method:	Least Squares	F-statistic:	319.0
Date:	Mon, 13 May 2024	Prob (F-statistic):	5.62e-26
Time:	10:22:04	Log-Likelihood:	-59.309
No. Observations:	40	AIC:	126.6
Df Residuals:	36	BIC:	133.4
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-23.4200	2.659	-8.800	0.000	-28.812	-18.028
螺纹钢lag1	0.0044	0.000	9.362	0.000	0.003	0.005
热轧卷板lag5	0.0031	0.000	10.166	0.000	0.003	0.004
纯碱lag12	-0.0031	0.000	-7.322	0.000	-0.004	-0.002

Omnibus:	2.024	Durbin-Watson:	1.018
Prob(Omnibus):	0.364	Jarque-Bera (JB):	1.855
Skew:	-0.434	Prob(JB):	0.396
Kurtosis:	2.400	Cond. No.	9.92e+04



（四）预测结果
我们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%。

```
predict_April_PPI(model, ['石油沥青lag12', '热轧卷板lag2', '锡lag1'])
✓ 0.4s
Estimated PPI for April is -2.062939340032436
```

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%

```
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
```

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，可以提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。

五、展望与思考-研究的价值
本文从数据层面证明了从期货市场研判宏观经济指标的可行性。

优势：

1. 使用期货市场的领先数据预测PPI，无需等待当期生产资料价格发布即可预测，具有前瞻性。
2. 使用期货市场具体品种收盘价月均值预测，数据公开易得。
3. 创新性的使用具体品种而非编制好的期货指数预测，找到数个支撑变量，有效剔除相关性较大的无用数据，使预测更精简有效。
4. 创新性的使用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。