

# 商品期货 CTA 专题报告（六）

证券研究报告

2018 年 02 月 09 日

## 基本面分析框架下的黑色系商品库存预测

### 库存变化是供需平衡表分析的灵魂

作为商品基本面分析灵魂的供需平衡表显示：库存是供需相对强弱的体现。真实库存难以统计，研究具有代表性的库存指标变化更有意义。本文采用自上而下的方法，从宏观经济基本面入手，结合微观行业和品种特征对黑色系商品的库存变化这一流量概念展开研究，并构建库存同比增速预测模型。

### 代表性库存指标及宏微观潜在影响因素梳理

结合指标含义、数据长度及可比性要求，我们筛选了黑色系产业链上下游 6 个品种的代表性库存指标，在宏观层面选取了投资、消费、进出口、通胀、货币环境及经济景气度等 6 个类别 53 个可能影响库存变化的关键指标，并从上游原材料、中游替代品及品种本身和下游行业景气度三个部分梳理了单个商品的核心指标。

### 库存同比先行指标筛选及最优阶数确定

为实现预测作用，我们需要挖掘的是领先于库存同比变化的先行指标。通过不同先行阶数下的宏微观指标与各商品库存同比增速的相关性检验，我们筛选出相关性较高且先行阶数在 2-12 阶的指标作为库存同比增速的先行指标。

### 基于 LASSO 回归的库存预测模型

为减弱数据样本量小及初选先行指标间的多重共线性可能引起 OLS 估计偏差大精度小的问题，我们利用 LASSO 回归，基于 10-折交叉验证法确定惩罚系数，对库存同比增速预测模型进行变量选择和参数估计。回归结果显示，除动力煤外，各品种的指标压缩程度均大于 50%。从样本内预测方向准确率来看，铁矿石、焦煤、焦炭及热卷四个品种的准确率高达 90%以上，样本内拟合效果较好。

### 预测模型在样本内外的表现具有一致性

从 6 次样本外滚动测试结果来看，各品种库存同比均具有较稳定的先行指标，样本内拟合效果好的铁矿石和热卷库存预测模型在样本外依然表现优异，预测方向准确率可达 100%。模型外推一期可预测受宏观因素共振影响 2018 年 2 月铁矿石库存同比将继续出现较大幅度上升，而热卷库存同比变化不大。

风险提示：模型基于历史数据，存在失效风险

### 作者

吴先兴 分析师  
SAC 执业证书编号：S1110516120001  
wuxianxing@tfzq.com  
18616029821

### 相关报告

- 1 《金融工程：商品期货 CTA 专题报告（五）我国商品期货分类及异质性基本面分析概述 2018-01-31》  
2018-01-31
- 2 《金融工程：商品期货 CTA 专题报告（四）库存基本面与动量技术面共振的商品期货投资策略 2018-01-05》  
2018-01-05
- 3 《金融工程：商品期货 CTA 专题报告（三）策略的趋势过滤 2017-03-22》  
2017-03-22
- 4 《金融工程：商品期货 CTA 专题报告（二）日内趋势策略初探 2017-03-10》  
2017-03-10
- 5 《金融工程：商品期货 CTA 专题报告（一）量化 CTA 策略概述 2017-02-14》  
2017-02-14



## 内容目录

<b>商品供需平衡表的灵魂：库存</b>	<b>4</b>
库存与库存变化	4
库存变化研究视角：宏观与微观	4
<b>黑色系商品库存潜在影响因素分析</b>	<b>5</b>
各品种代表性库存指标	5
宏观经济基本面指标	7
微观品种基本面指标	7
<b>库存先行指标挖掘</b>	<b>9</b>
数据处理	9
先行指标筛选及最优先行阶数检验	9
<b>基于 LASSO 回归的库存预测模型</b>	<b>10</b>
LASSO 回归模型	10
样本内回归	11
外推一期预测	15
稳健性检验：滚动预测效果	16
<b>总结与展望</b>	<b>18</b>

## 图表目录

图 1：商品库存变化内生影响逻辑梳理	4
图 2：动力煤代表性库存同比增速	6
图 3：黑色系产业链上游商品库存同比增速	6
图 4：黑色系产业链下游商品库存同比增速	7
图 5：黑色系产业链上下游构成	8
图 6：铁矿石库存同比预测值与真实值	12
图 7：动力煤库存同比预测值与真实值	12
图 8：焦煤库存同比预测值与真实值	12
图 9：焦炭库存同比预测值与真实值	12
图 10：螺纹钢库存同比预测值与真实值	12
图 11：热卷库存同比预测值与真实值	12
图 12：滚动预测稳健性检验逻辑	16
表 1：黑色系商品代表性库存指标	6
表 2：宏观经济基本面原始指标分类	7
表 3：各品种供需基本面分析	8
表 4：微观品种基本面指标分类	8
表 5：黑色系期货品种对应现货价格指标	9
表 6：LASSO 回归效果分析	11
表 7：LASSO 回归前后指标间多重共线性检验结果对比	11

表 8：基于 LASSO 回归的铁矿石库存同比影响因素 .....	13
表 9：基于 LASSO 回归的动力煤库存同比影响因素 .....	13
表 10：基于 LASSO 回归的焦煤库存同比影响因素 .....	14
表 11：基于 LASSO 回归的焦炭库存同比影响因素 .....	14
表 12：基于 LASSO 回归的螺纹钢库存同比影响因素 .....	15
表 13：基于 LASSO 回归的热卷库存同比影响因素 .....	15
表 14：基于 LASSO 回归的 2018 年 2 月库存同比增速预测 .....	16
表 15：各品种样本外滚动预测效果 .....	17
表 16：各商品样本外测试的稳定先行指标 .....	17

我们在上一篇系列报告中搭建了五大类商品的基本面分析框架，本文将更细致地从**宏观经济环境及微观品种异质性两个维度**分析黑色系商品的供需影响因子，并试图预测基本面最为重要的指标之一：库存。

## 商品供需平衡表的灵魂：库存

商品基本面分析主要研究供给与需求，根据供需相对强弱判断价格走势。因此，构建供需平衡表（Balance Table）、寻关键变量是基本面分析的第一步也是最关键的一步。

商品的供需平衡表显示：供给+期初库存=需求+期末库存，即当期库存=供给-需求。如果说供需平衡表是商品基本面分析的灵魂，那库存可视为供需平衡表的灵魂。我们希望对库存这一重要指标进行深入研究。

## 库存与库存变化

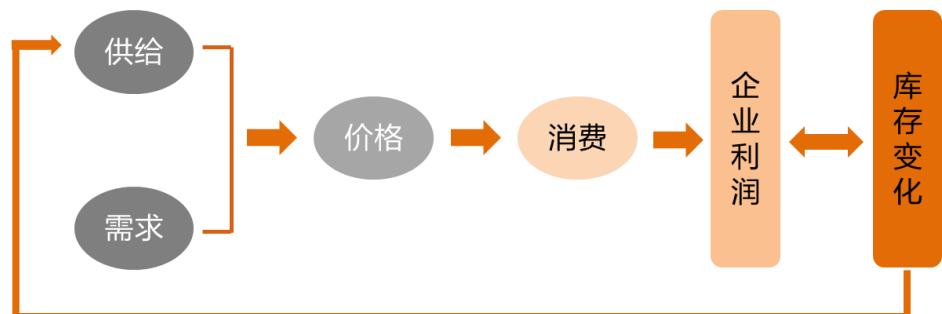
根据定义，库存是指为最终销售而准备的原材料、中间品以及产成品，可理解为由上游供应商、中游贸易商及下游消费产成品三个部分构成。库存本身是一个存量概念，研究其绝对水平意义不大，一方面大部分商品库存具有周期性特征，另一方面库存难以完全统计，单纯依赖局部库存绝对量进行分析容易产生误导性。而**库存变化即库存投资，或者说企业的库存去补行为**，这一流量概念，才是影响价格变化的核心。

## 库存变化研究视角：宏观与微观

从供需平衡表公式来看，研究库存变化需要判断供给与需求的相对变化，供给的量化较为容易，而需求量化的难度和偏误都要大得多。在做需求分析之前，我们首先需要区别“需求”与“消费”这两个概念。对于价格来说，需求是“因”，而消费为“果”，两者不可混为一谈。

除商品自身产业链因素，周期性商品的供需及库存变化受宏观经济环境影响也很大，特别是近几年供给侧改革、环保限产等政策的实施，从外部制约了企业的库存行为。黑色系商品具有强周期性，本文将采用自上而下的方法，从宏观经济基本面入手，结合微观行业和品种特征对黑色系商品的库存变化展开研究。

图 1：商品库存变化内生影响逻辑梳理



资料来源：天风证券研究所

## 黑色系商品库存潜在影响因素分析

根据上一篇系列报告中所做的分类，我们研究的黑色系商品包括：上游的铁矿石、动力煤、焦煤、焦炭，中游的硅铁、锰硅，以及下游的螺纹钢、热轧卷板。本文将对各个品种的产业链上下游作进一步分析，寻找代表性库存指标，并从宏观微观两个维度寻找可能影响库存变化的先行指标。

### 各品种代表性库存指标

结合各大黑色商品期货分析师的研究成果，我们首先梳理了各品种具有代表性的库存指标。这一代表性体现在：

1. **统计口径一致且全面。**分析 Wind 数据库中各商品的库存指标，我们发现许多指标都做了口径修改，尤其是港口库存，统计的港口数不同将导致数据不可比。因此，我们筛选的指标统计口径需尽量保持一致，最好范围足够广，能接近真实库存；
2. **趋势主导性。**考虑到真实库存难以统计，而我们希望预测的是库存变化，非真实库存水平，只要选取的库存指标能够反映该商品的库存趋势即可。
3. **数据长度合理。**为满足从观测到预测的有效性，至少应有 5 年的数据。为剔除季节性影响，我们研究的是库存同比增速，即要求原始库存指标基期在 2012 年 12 月之前。

基于以上“代表性”要求，我们对各品种逐一进行研究，发现 Wind 数据库中不存在硅铁和锰硅库存指标，只有硅铁产量和锰矿库存数据，且数据量不足，我们无法对这两个品种的库存做进一步研究，因此本文仅研究上游和下游共 6 个品种的库存变化。

下面对上游品种库存指标筛选过程做具体说明。

**铁矿石：**进口是关键因素，港口库存代表性强，主要进口国为澳大利亚和巴西。分析港口库存，我们发现 41 港口、45 港口库存数据量不足 5 年，因此，我们选择了满足要求的国内港口库存量（36 港口，口径不变）。

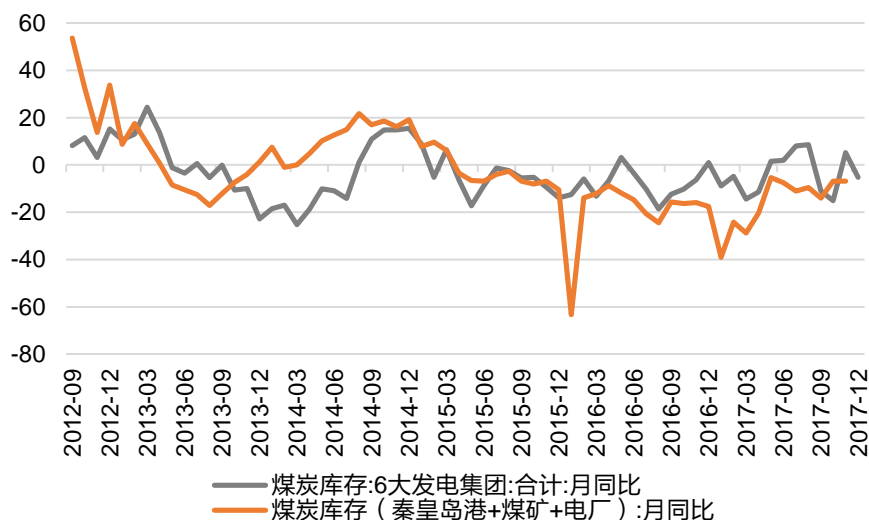
**动力煤：**据相关统计，动力煤产量大约占我国煤炭总产量的 80%，而在我国的动力煤消费结构中，有 65% 以上是用于火力发电。因此下游 6 大发电集团库存是最重要的变量之一，除此之外上游的煤矿、其余重点港口的库存也需要重点关注。但我们发现，这些指标数据更新时间不一，不宜直接加总。从图 2 来看，6 大发电集团库存同比增速与较全面的库存同比增速走势近似，因此，我们将 6 大发电集团库存作为动力煤库存的代表性指标。

**焦煤：**煤矿和港口的焦煤库存较少，且无法区分，因此，对于焦煤品种来说，我们仅观察炼焦煤库存。而炼焦煤在焦化厂的库存数据自 2016 年下半年才开始统计，数据量不足，这里我们使用口径一致且数据基期较早的三港口（京唐、日照、连云港）库存合计。

**焦炭：**我国焦炭出口较多，重点港口为天津港、连云港、日照港、青岛港，而青岛港数据量不足且占比小，因此基于重要性、统计口径和数据长度考虑，我们仅以三港口合计作为代表性库存指标。

下游螺纹钢和铁矿石均采用主要城市库存指标，筛选后的各品种代表性库存指标汇总如表 1 所示。

图 2：动力煤代表性库存同比增速



资料来源: Wind, 天风证券研究所

表 1：黑色系商品代表性库存指标

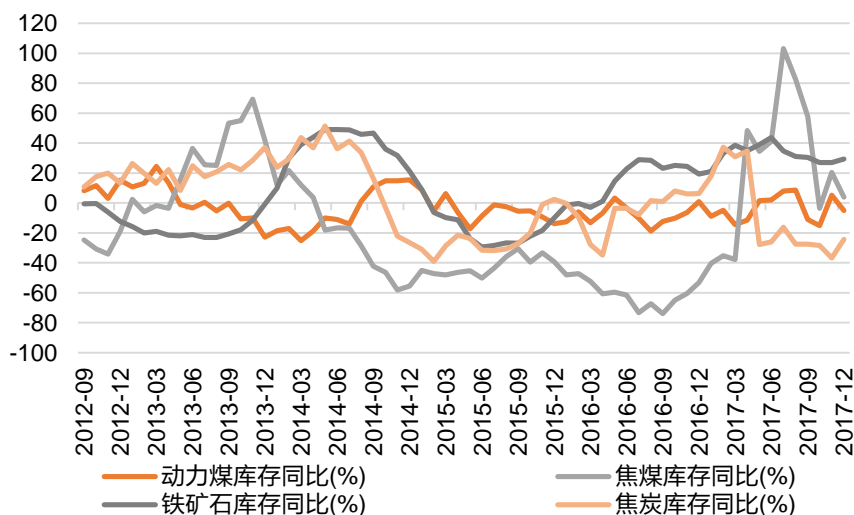
品种	指标名称	数据基期	更新频率
铁矿石	国内铁矿石港口库存量	2006/7/25	周
动力煤	煤炭库存: 六大发电集团: 合计	2009/10/1	日
焦煤	炼焦煤库存: 三港口合计	2011/9/24	周
焦炭	焦炭库存: 天津港+连云港+日照港	2011/4/2	周
螺纹钢	库存: 螺纹钢(含上海全部仓库)	2010/5/15	周
热轧卷板	库存: 热卷(板)	2006/3/11	周

资料来源: Wind, 天风证券研究所

考虑到宏观指标数据更新频率以月频为主, 为保持一致, 我们使用所有库存指标的月平均同比数据。从各指标数据基期来看, 焦炭库存数据起始日最晚, 同比后的数据基期为 2012 年 9 月, 即我们的样本期为: 2012 年 9 月-2017 年 12 月。

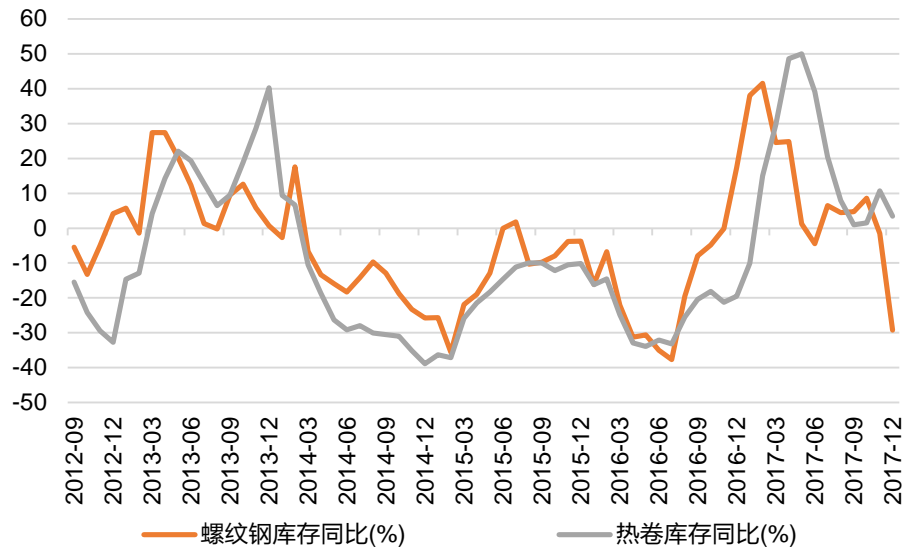
图 3、图 4 展示了上下游品种样本期内库存同比增速, 可以发现下游螺纹钢与热卷的库存变化较为一致。

图 3：黑色系产业链上游商品库存同比增速



资料来源: Wind, 天风证券研究所

图 4：黑色系产业链下游商品库存同比增速



资料来源：Wind，天风证券研究所

## 宏观经济基本面指标

下面寻找可能影响黑色系商品库存的基本面指标。首先，在宏观层面，我们筛选了投资、消费、进出口、通胀、货币环境及经济景气度等 6 个类别共 53 个关键指标，筛选条件为：

1. **经济含义。**反映经济环境的指标很多，我们选取了宏观研究和钢铁、煤炭行业研究中常用的具有代表性的部分；
2. **选择同比数据。**除本身季节性不明显的经济景气度指数、PMI、Shibor、国债到期收益率，其余指标均选择当月同比或累计同比数据以剔除季节性并更好地体现发展增速。
3. **数据长度合理。**为满足从观测到预测的有效性，至少应有 5 年的数据。

表 2：宏观经济基本面原始指标分类

类别	主要指标
投资	工业增加值、固定资产投资、发电量、工业企业产成品库存、产销率
消费	社会消费品零售总额、汽车销量
进出口	波罗的海干散货指数 (BDI)
通胀	CPI、PPI
货币环境	M1、M2、Shibor、国债到期收益率
经济景气度	PMI、克强指数、宏观经济景气指数、消费者指数、投资者信心指数、生产资料库存指数等

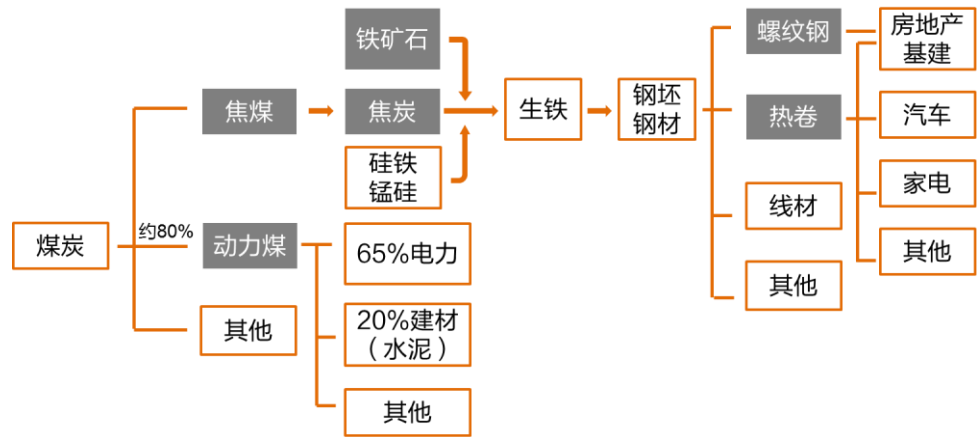
资料来源：Wind，天风证券研究所

## 微观品种基本面指标

在微观层面，结合整个黑色系产业链上下游和各品种库存相关性来看，我们可以将 6 个品种分成铁矿石、煤炭类（包括动力煤、焦煤、焦炭）和钢材类（螺纹钢、热卷）三类进行微观基本面指标筛选。



图 5：黑色系产业链上下游构成



资料来源：Wind, 天风证券研究所

表 3：各品种供需基本面分析

品种	说明	上游	下游	主要价格影响因素
上游 铁矿石	主产国是澳大利亚和中国。我国铁矿石产量无法满足国内钢铁生产需求，存在进口依赖。	矿山	生铁、钢材等	供给：矿产资源、煤炭产能、运输成本、国家政策（供给侧改革、环保限产等）； 需求：宏观经济形势、国际市场、替代品价格、国家政策（房地产、钢铁行业等）
上游 动力煤	需求拉动型市场。近年进口增出口降。秦皇岛港是我国重要的对外贸易口岸，是目前世界最大的煤炭输出港和散货港。	矿井、煤炭	火力发电（65%）、建材用煤、生产水泥玻璃和石灰（20%）等	
上游 焦煤	商品化率高，国内贸易量大，中西部产供，东部消费，运输以铁路为主。	煤矿	焦炭、钢铁等	
上游 焦炭	我国是生产和出口大国，影响力世界第一。	炼焦煤	钢铁（约 85%）、化学制品等	
下游 螺纹钢	完全竞争市场。我国产量最大钢材品种之一，大部分集中在钢厂附近区域销售，区域性、季节性特点突出。	生铁、硅铁、锰硅	房地产、桥梁、道路等土建工程建设	
下游 热卷	重要钢材品种，产业集中度高。我国是全球最大的生产国、消费国和出口国。	板坯（生铁、硅铁、锰硅）	船舶、汽车、桥梁、建筑、家电、压力容器等	

资料来源：上期所，郑商所，大商所，天风证券研究所

表 4：微观品种基本面指标分类

产业链位置	影响来源	指标类别
上游	原材料	产量、库存、价格、进出口、运输成本
中游	替代品	产量、库存、价格
	品种本身	进出口、现货收益率
下游	行业景气度	开工率、产量、消费量、价格、利润

资料来源：Wind, 天风证券研究所



表 5：黑色系期货品种对应现货价格指标

产业链位置	品种	对应现货品种	现价数据基期
上游	铁矿石	青岛港,PB 粉矿,61.5%,澳产	2011/9/8
	动力煤	秦皇岛港,动力末煤,Q5500,山西	2008/12/24
	焦煤	吕梁,主焦,A10.5,V20-24,S1,G75,Y12-15,Mt8	2012/6/7
	焦炭	唐山,二级,A13.5,S0.7	2005/4/19
中游	硅铁	青海,75A	2011/4/27
	锰硅	广西,FeMn68Si18	2012/8/15
下游	螺纹钢	上海,HRB400,20mm	2007/1/8
	热轧卷板	上海,Q345,5.75mm	2012/8/15

资料来源：Wind, 天风证券研究所

从单个品种微观基本面来看，商品库存的影响来源主要是上游原材料、中游替代品及品种本身和下游行业景气度三个部分，指标主要有产量、消费量、价格、进出口以及下游企业的利润和开工率等。其中品种本身的现货收益率是各期货品种对应现货的月度收益率。基于此，我们共筛选了 23 个铁矿石微观层面指标，30 个煤炭类微观层面指标以及 46 个钢材类微观层面指标。下面分类别进行数据建模。

## 库存先行指标挖掘

为了预测库存同比增速，我们首先需要挖掘领先的有效指标，对于滞后指标及领先效果不显著的指标予以剔除。

## 数据处理

在先行指标筛选前，我们对宏观经济指标和微观品种指标数据质量进行分析，发现大部分宏观指标 1 月份数据缺失现象严重，这可能是由于受春节因素影响，因此我们将所有指标 1 月份数据均予以剔除。除去 1 月份，部分指标还存在缺失情况，对于这些指标我们的处理方法是：

1. 若缺失率超过 20%，则将该指标剔除样本；
2. 若缺失率小于 20%，则用该指标过去的均值填补，且不考虑 2010 年以前的数据。如 2014 年 6 月的数据存在缺失，则使用 2010 年 2 月-2014 年 5 月的均值填补。两点说明：第一，我们不用全样本均值或者其他回归填补法是为了避免“盗用”未来数据，因为站在当下，能利用的只有过去的信息；第二，我们仅考虑 2010 年之后的数据，是因为近期信息相对更重要。

至此，我们的数据样本为 2012 年 9 月-2017 年 12 月中除去每年 1 月份，即共 59 个月的完整数据。

## 先行指标筛选及最优先行阶数检验

数据处理完毕后，我们开始检验宏观微观指标与各商品库存同比增速的领先滞后关系。我们对先行指标的定义为：

存在  $L$ ，满足  $t-L$  期指标  $X$  与  $t$  期库存同比增速  $Y$  的相关系数显著大于  $\alpha$ ，则称  $X$  为  $Y$  的  $L$  阶先行指标。

其中， $L$  为某个区间内的正整数，考虑到数据频率及样本量限制，我们取  $L$  最大不超过 12。

考虑到大多数库存指标以周为频率进行更新，相对于大部分宏观自变量指标均能提前获取数据，这就要求先行指标具有更高的先行阶数方可实现对库存同比增速的预测。如我们要预测  $t+1$  期库存同比增速，我们可利用的只有  $t-1$  期或更早期的数据，也就是说我们筛选的先行指标应至少领先库存同比增速 2 阶，即  $L$  最小为 2。

基于以上分析，我们在筛选先行指标时将涉及的参数有最小先行阶数  $L_{\min}$ 、最大先行阶数  $L_{\max}$  和相关系数临界值  $\alpha$ 。本文取参数  $L_{\min} = 2$ ， $L_{\max} = 12$ ， $\alpha = 0.4$  对各指标进行检验。若指标  $X$  在领先 2-12 阶这个区间内与库存同比增速的最大相关系数显著大于 0.4，则认为该指标是该品种库存同比增速的先行指标，且取到最大相关系数所对应的阶数为最优先行阶数。

对各品种库存的宏观潜在影响指标进行上述检验后，我们可得到一系列先行指标及其对应最优先行阶数，并剔除滞后或不显著领先的指标，统计后发现，铁矿石、动力煤、焦煤、焦炭、螺纹钢、热卷库存同比增速的先行指标数分别为 26、9、48、30、25、42 个。由于指标众多，不做一一罗列，我们将在后续模型最终求解结果中给出最核心的影响因素。

## 基于 LASSO 回归的库存预测模型

基于以上筛选结果，我们可利用先行指标及其对应的最优先行阶数构建多元回归模型来预测库存同比增速。分析各指标间关系后，我们发现：

1. 先行指标个数多，而样本数据量少，OLS 回归可能无法求解；
2. 大部分品种的先行指标中存在许多经济含义相近、相关性较高的指标（如工业企业利润总额的累计同比与当月同比），即自变量之间可能存在多重共线性，导致 OLS 估计量方差较大、精度较低。

### LASSO 回归模型

LASSO (Least Absolute Shrinkage and Selection Operator) 最早由 Tibshirani (1996) 提出，该方法主要应用于线性模型，其本质是在残差平方和上添加惩罚函数，在估计参数时，系数被压缩，部分甚至被压缩至 0 来实现模型选择。LASSO 算法在大规模变量模型中兼具岭回归和子集选择的优点，能同时实现变量选择和参数估计，可有效减弱多重共线性的影响，这正好符合我们的需求。

基于以上分析，本文用 LASSO 回归模型来进一步分析库存同比增速的主要影响指标，并试图利用回归结果实现对库存同比增速的预测。

已知数据集  $(X^i, y_i)$ ，其中  $i = 1, 2, 3, \dots, N$ ， $N$  为样本总量， $X^i$  和  $y_i$  分别是第  $i$  个观测点对应的自变量和因变量，本文的自变量为初步筛选的库存同比先行指标，因变量为库存同比增速。考虑建立如下线性回归模型：

$$y_i = \alpha_0 + \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

LASSO 回归的系数估计为：

$$\hat{\beta}^{lasso} = \arg \min_{\alpha_0, \beta} \left\{ \frac{1}{2N} \sum_{i=1}^N \left( y_i - \alpha_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

其中， $\alpha_0$  为截距值； $\lambda$  为正则化参数，又称惩罚系数、控制压缩程度。 $\lambda$  越大，对回归系数的压缩程度越大，得到的模型自由度越小，模型越简洁；反之， $\lambda$  越小，对回归系数的压缩程度减小，更多的变量被选入模型中，模型复杂度增加。LASSO 方法通过调整惩罚项  $\lambda$  得到自由度更小的模型。

对于  $\lambda$  的设定，我们采用 K-折交叉验证法（K-fold Cross Validation）。该方法的原理是将原始数据随机分成 K 个部分，选择一个作为测试样本，剩下的 K-1 个作为训练样本。实验重复做 K 次，保证 K 个部分的数据均被测试，最后对 K 个实验结果取平均。

为保证模型精度和简洁度，我们在考虑残差平方和较小的同时选择尽可能大的  $\lambda$  以实现更好的变量压缩效果。

## 样本内回归

我们设定 K = 10 进行交叉验证，并在最小 MSE（均方误差）一倍标准差范围内选择最大的  $\lambda$  作为惩罚系数。由于交叉验证的分组具有随机性，我们得到的回归结果也存在一定差异，但从多次测试来看，我们发现 LASSO 回归具有以下一致结论：

1. **变量压缩程度高。**除动力煤外，其余品种模型的变量压缩率均大于 50%，大大降低了模型复杂度。
2. **模型拟合效果好。**由于我们判断的是库存同比增速，正负方向比绝对数值更为重要，因此，我们将预测值与真实值方向一致的比率作为粗略估计的准确率。从样本内预测方向准确率来看，铁矿石、焦煤、焦炭及热卷四个品种的准确率能高达 90% 以上。从各品种库存同比预测值与真实值对比图中我们可以更直观地看到，无论从预测方向还是绝对值的角度，这四个品种的拟合效果确实不错。

表 6-13，图 6-11 呈现的是其中一组测试结果，以此代表 LASSO 回归的一般结论。从表 7 回归前后各指标间的多重共线性检验结果来看，除焦煤输入先行指标个数多于回归月份数（样本总月份数与先行指标最大滞后期的差）无法求解外，其余品种各指标的 VIF 值均有大幅下降，可见 LASSO 确实能起到压缩变量、减弱模型多重共线性的作用。

表 6：LASSO 回归效果分析

品种	回归月份数	输入先行 指标数	LASSO 回归 筛选变量数	样本内预测 方向准确率
铁矿石	49	26	11	<b>91.84%</b>
动力煤	47	9	6	80.85%
焦煤	47	48	11	<b>95.74%</b>
焦炭	48	30	15	<b>93.75%</b>
螺纹钢	48	25	8	85.42%
热卷	47	42	7	<b>93.62%</b>

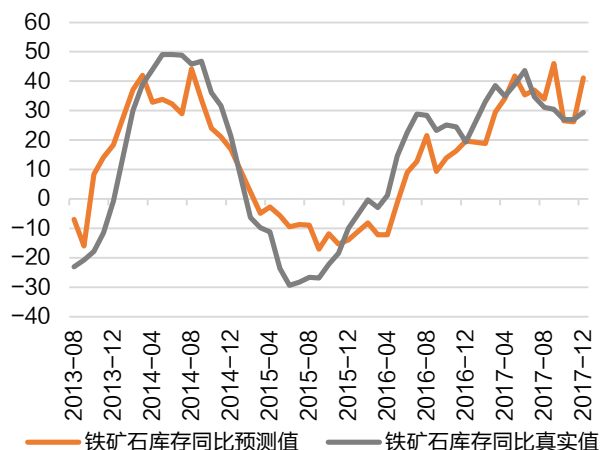
资料来源：天风证券研究所

表 7：LASSO 回归前后指标间多重共线性检验结果对比

品种	指标数		VIF > 10 的指标数		VIF 均值	
	LASSO 前	LASSO 后	LASSO 前	LASSO 后	LASSO 前	LASSO 后
铁矿石	26	11	21	3	8041.17	<b>6.92</b>
动力煤	9	6	0	0	2.56	<b>1.60</b>
焦煤	48	11	不可解	0	不可解	<b>3.15</b>
焦炭	30	15	20	1	149.14	<b>4.92</b>
螺纹钢	25	8	15	0	24.89	<b>2.24</b>
热卷	42	7	41	1	44090.47	<b>5.34</b>

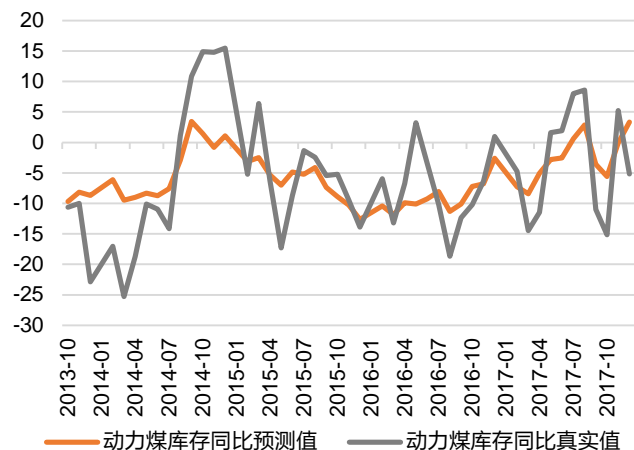
资料来源：天风证券研究所

图 6：铁矿石库存同比预测值与真实值



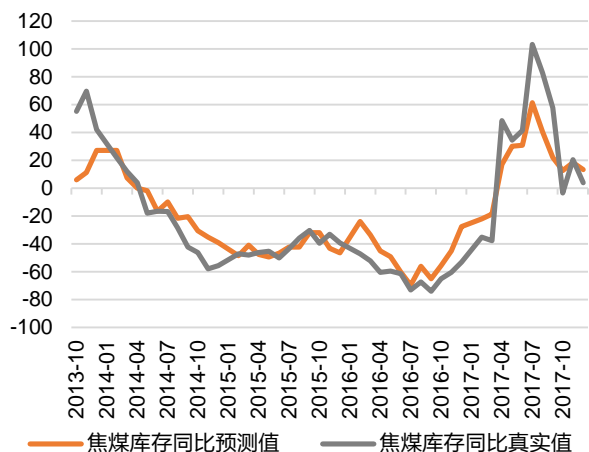
资料来源：Wind, 天风证券研究所

图 7：动力煤库存同比预测值与真实值



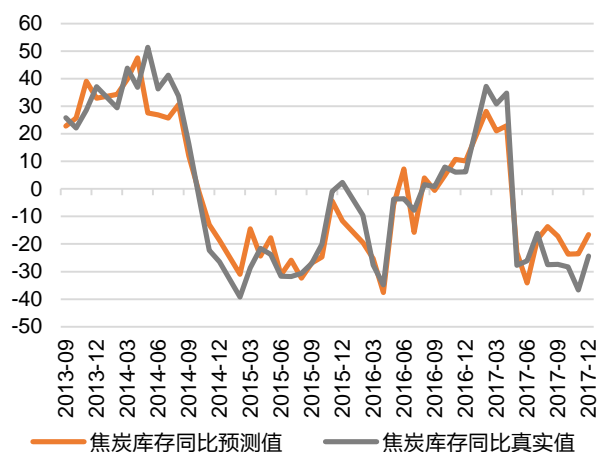
资料来源：Wind, 天风证券研究所

图 8：焦煤库存同比预测值与真实值



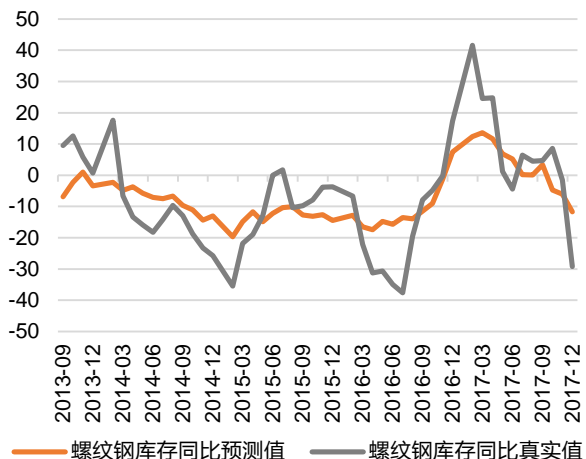
资料来源：Wind, 天风证券研究所

图 9：焦炭库存同比预测值与真实值



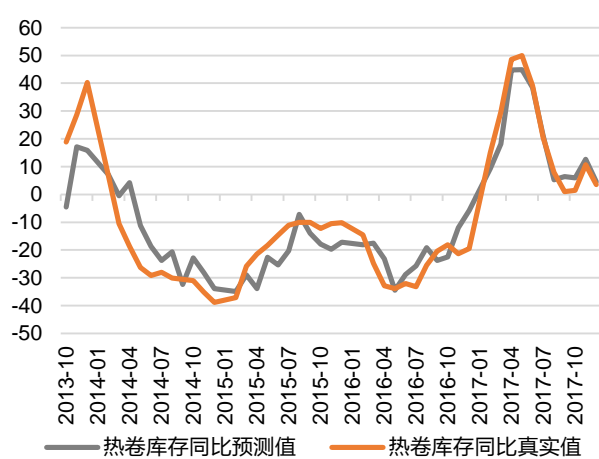
资料来源：Wind, 天风证券研究所

图 10：螺纹钢库存同比预测值与真实值



资料来源：Wind, 天风证券研究所

图 11：热卷库存同比预测值与真实值



资料来源：Wind, 天风证券研究所

分品种回归结果来看，各商品库存变化都同时受宏观经济环境与产业链上下游的影响。综合来看，我们可以归纳得出如下结论：

1. 宏观层面的核心先行指标主要有波罗的海干散货指数（BDI）、产销率、发电量等。其中 BDI 指数是散装原物料（如钢材、煤、矿砂等工业原材料）的运费指数，并被认为是经济先行指标，与黑色系商品尤其是进出口频繁的品种关系紧密。这一理论分析也得到了实证的支撑。根据回归结果，BDI 指数是铁矿石库存同比的 5 阶先行指标、动力煤的 11 阶先行指标、热卷的 2 阶先行指标；
2. 终端消费行业汽车销量是重要的先行指标，但影响方向可能存在差异，这可能与我们的样本期过短有关；
3. 铁矿石库存同比滞后于产业链上游原矿产量、下游废钢、钢材库存变化；
4. 动力煤库存同比滞后于产业链上游煤炭调度、中游现货收益率、下游焦炭表观消费量变化；
5. 焦煤库存同比滞后于产业链中上游煤炭产量、下游房地产景气度、粗钢产量和表观消费量变化等；
6. 焦炭库存同比滞后于产业链上游煤炭进口量、中游焦炭产量、下游水泥行业产量、房地产景气度、钢厂开工率变化等；
7. 螺纹钢库存同比滞后于产业链上游原材料（铁矿石、焦炭）价格、中游热卷收益率、下游房屋开工和家电冰箱产量变化等；
8. 热卷与螺纹钢相关性高，其库存同比影响逻辑与螺纹钢近似。

表 8：基于 LASSO 回归的铁矿石库存同比影响因素

指标类别	先行指标	先行阶数	影响方向
宏观	PMI	2	-
宏观	PPI：生产资料：采掘工业：当月同比	2	+
宏观	克强指数：累计值	2	-
宏观	波罗的海干散货指数（BDI）：同比	5	+
宏观	产量：发电量：当月同比	6	+
宏观	M1：同比	9	+
上游	产量：铁矿石原矿量：当月值	3	+
下游	废钢平均库存可用天数：国内大中型钢厂	4	+
下游	销量：汽车：当月同比	7	+
下游	销量：汽车：累计同比	7	-
下游	库存：主要钢材品种：合计	10	+

资料来源：Wind, 天风证券研究所

表 9：基于 LASSO 回归的动力煤库存同比影响因素

指标类别	先行指标	先行阶数	影响方向
宏观	工业企业：产成品存货累计同比	3	+
宏观	波罗的海干散货指数（BDI）：同比	11	+
上游	煤炭调度：秦皇岛港：锚地船舶数	11	+
中游	动力煤现货收益率	9	+
下游	表观消费量：焦炭：累计值	2	+
下游	销量：汽车：当月同比	12	+

资料来源：Wind, 天风证券研究所

表 10：基于 LASSO 回归的焦煤库存同比影响因素

指标类别	先行指标	先行阶数	影响方向
宏观	产销率：当月值：同比	10	+
上游	产量：原煤：当月同比	2	+
上游	煤炭调度：秦皇岛港：预到船舶数	2	+
中游	产量：炼焦煤（炼焦精煤）：当月值	5	+
中游	进口数量：动力煤：当月值	10	+
中游	动力煤现货收益率	10	+
下游	表观消费量：粗钢：当月值	2	+
下游	产量：焦炭：当月同比	3	+
下游	产量：粗钢：当月同比	3	+
下游	30 大中城市：商品房成交面积	5	+
下游	销量：汽车：累计同比	7	+

资料来源：Wind, 天风证券研究所

表 11：基于 LASSO 回归的焦炭库存同比影响因素

指标类别	先行指标	先行阶数	影响方向
宏观	宏观经济景气指数：预警指数	2	+
宏观	产量：发电量：累计同比	2	-
宏观	产量：发电量：当月同比	3	+
宏观	CPI：当月同比	3	+
宏观	M2：同比	10	+
上游	进口数量：动力煤：当月值	2	+
上游	进口数量：炼焦煤：当月值	10	+
上游	煤炭销量：全国：当月同比	11	+
中游	产量：焦炭：当月同比	11	+
下游	销量：汽车：当月同比	2	+
下游	销量：汽车：累计同比	2	+
下游	30 大中城市：商品房成交面积	5	+
下游	产量：水泥：当月同比	5	+
下游	唐山钢厂：高炉开工率	6	+
下游	表观消费量：钢材：当月值	6	-

资料来源：Wind, 天风证券研究所



表 12：基于 LASSO 回归的螺纹钢库存同比影响因素

指标类别	先行指标	先行阶数	影响方向
宏观	固定资产投资完成额：制造业：黑色金属冶炼及压延加工业：累计同比	8	+
宏观	产销率：当月值：同比	9	+
宏观	产销率：累计值：同比	9	+
上游	铁矿石价格指数：62%Fe:CFR 中国北方	2	+
上游	天津港：平仓价（含税）：一级冶金焦：山西产	2	+
中游	热卷现货收益率	11	+
下游	产量：家用电冰箱：当月同比	2	+
下游	房屋新开工面积：累计同比	8	+

资料来源：Wind, 天风证券研究所

表 13：基于 LASSO 回归的热卷库存同比影响因素

指标类别	先行指标	先行阶数	影响方向
宏观	波罗的海干散货指数（BDI）：同比	2	+
宏观	产销率：当月值：同比	5	+
上游	铁矿石价格指数：62%Fe:CFR 中国北方	3	+
上游	天津港：平仓价（含税）：一级冶金焦：山西产	3	+
中游	库存：主要钢材品种：合计	2	+
中游	库存：螺纹钢（含上海全部仓库）	2	+
下游	产量：家用电冰箱：当月同比	6	+

资料来源：Wind, 天风证券研究所

## 外推一期预测

基于 2012 年 9 月-2017 年 12 月的样本期数据 LASSO 回归结果,我们得到了各品种库存同比的有效先行指标及影响强弱,根据样本内参数估计结果,我们可以试图对下期库存同比增速进行预测。由于各先行指标至少领先两期,几乎不存在前一月数据未更新的情况,可以实现月初对月末数据的预测。1 月份不在我们的研究样本内,我们以样本内回归结果预测 2018 年 2 月份数据,以期提供一个**方向上的判断**。

以铁矿石为例,根据 LASSO 压缩后的先行指标,要预测 2018 年 2 月库存同比增速,需要领先 2 阶(即 2017 年 11 月)的 PMI 数据、领先 5 阶(即 2017 年 8 月)的 BDI 指数,以此类推。将各先行指标与 LASSO 回归结果相结合即可实现对 2018 年 2 月库存同比增速的预测(见表 14)。根据 10 次 10-折交叉验证(10-fold CV)结果来看,除焦炭和热卷预测值之间标准差相对较大,其余品种 10 次交叉验证预测值较稳定。从平均预测值来看,铁矿石库存同比增速将继续出现较大幅度上升,热卷库存同比变化不大,其余品种库存大概率将呈现同比下降趋势。进一步分析铁矿石库存同比增速高位的原因,我们发现:下游先



行指标钢材 17 年 9 月库存同比和废钢 17 年 3 月库存同比较强劲的增速，叠加 17 年 4 月 M1 同比的高位，17 年 7 月发电量同比的上行以及 17 年 8 月 BDI 指数同比回升，经济回暖利好钢企迹象明显，宏观因素共振是铁矿石可能补库的主要诱因。

表 14：基于 LASSO 回归的 2018 年 2 月库存同比增速预测

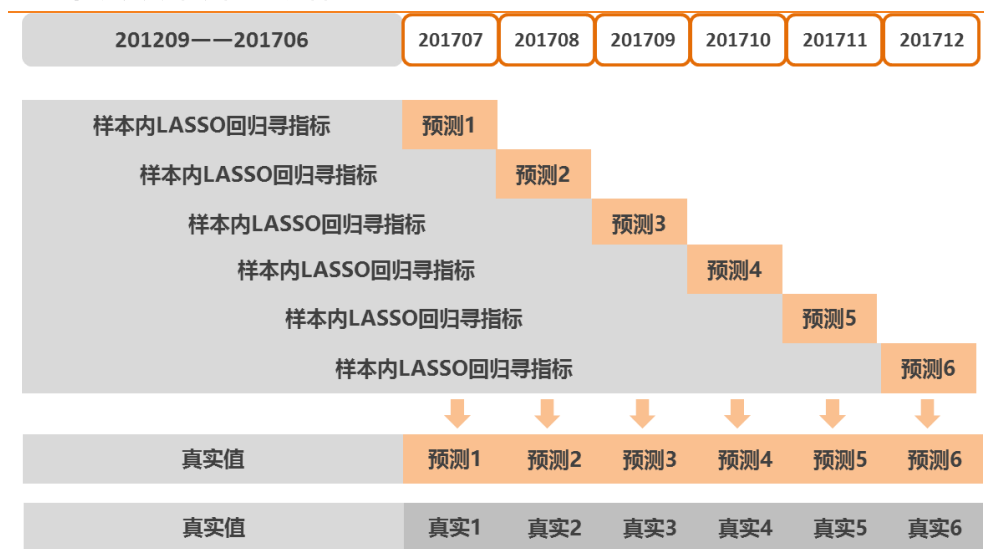
品种	预测 2018 年 2 月库存同比增速 (10 次 10-fold CV 平均结果)	预测标准差
铁矿石	49.65% (↑)	0.04
动力煤	-2.18% (↓)	0.09
焦煤	-6.42% (↓)	0.17
焦炭	-22.61% (↓)	3.10
螺纹钢	-12.14% (↓)	0.24
热卷	2.73% (↑)	2.98

资料来源：天风证券研究所

### 稳健性检验：滚动预测效果

从以上整个样本期模型拟合效果来看，基于交叉验证的 LASSO 回归能较好地筛选黑色系商品库存同比的核心先行指标。但样本外表现如何呢？为验证模型的稳健性，我们将样本期最后 6 个月作为检验样本，逐月预测库存同比增速并于其真实值进行对比。操作逻辑如图 12 所示。

图 12：滚动预测稳健性检验逻辑



资料来源：天风证券研究所

表 15 汇总了各品种 2017 年 7 月-12 月的库存同比滚动预测值与真实值，结合样本内拟合效果，我们发现：

从方向准确率来看，样本内拟合效果好的铁矿石和热卷的库存预测模型在样本外表现依然非常优异，准确率达 100%，而样本内拟合效果欠佳的动力煤、螺纹钢，在样本外表现则依然不尽如人意。据此，我们推测，基于 LASSO 回归的库存预测模型在样本内外的表现具有一致性。

从 6 次样本外测试结果来看，各品种库存同比均具有影响较为稳定的先行指标。表 15 统计了 6 次都被纳入回归模型的先行指标，我们认为观察这些指标对于预测库存同比变化更具指向性意义。而部分可能影响较大的指标（如盈利钢厂比率、全国高炉开工率、矿山开工率等）由于数据样本量限制未被纳入模型，这将在一定程度上降低预测结果的精度。此外，在筛选先行指标时，为了避免过拟合，我们将相关系数临界值统一设定为 0.4，而这对单个品种来说并非最优，因此预测的绝对数值大小仅作参考。

表 15：各品种样本外滚动预测效果

	铁矿石		动力煤		焦煤		焦炭		螺纹钢		热卷	
	LASSO_ predict	True	LASSO_ predict	True	LASSO_ predict	True	LASSO_ predict	True	LASSO_ predict	True	LASSO_ predict	True
201707	63.5	34.7	-1.3	8.0	8.5	103.2	17.4	-16.2	-1.9	6.4	11.4	20.5
201708	57.9	31.1	2.8	8.6	28.2	82.7	4.0	-27.5	-0.5	4.4	9.9	8.0
201709	69.4	30.5	8.9	-11.0	69.1	57.7	0.4	-27.4	-15.2	4.7	6.3	1.0
201710	24.1	27.0	-3.4	-15.1	43.2	-3.6	-8.8	-28.4	-13.8	8.6	5.7	1.5
201711	27.1	27.0	-0.1	5.2	25.0	20.4	-4.6	-36.7	-3.5	-1.5	14.7	10.7
201712	58.5	29.4	5.9	-5.2	18.3	3.8	-17.3	-24.4	-5.2	-29.2	5.2	3.5
准确率	100%		33%		83%		50%		33%		100%	

资料来源：Wind, 天风证券研究所

表 16：各商品样本外测试的稳定先行指标

铁矿石		动力煤		焦煤	
稳定先行指标	先行阶数	稳定先行指标	先行阶数	稳定先行指标	先行阶数
M1：同比	10	动力煤现货收益率	9	煤炭调度：秦皇岛港：预到船舶数	2
库存：主要钢材品种：合计	10	波罗的海干散货指数（BDI）： 同比	11	产量：粗钢：当月同比	3
销量：汽车：当月同比	7	销量：汽车：当月同比	12	30 大中城市：商品房成交面积	5
产量：发电量：当月同比	6			产销率：当月值：同比	10
产量：铁矿石原矿量：当月值	3				
波罗的海干散货指数 （BDI）：同比	5				
销量：汽车：累计同比	7				
焦炭		螺纹钢		热卷	
稳定先行指标	先行阶数	稳定先行指标	先行阶数	稳定先行指标	先行阶数
进口数量：炼焦煤：当月值	10	固定资产投资完成额：制造业：黑色金属冶炼及压延加工业：累计同比	7	波罗的海干散货指数（BDI）：同比	2
CPI：当月同比	4	天津港：平仓价（含税）：一级冶金焦：山西产	2	产量：家用电冰箱：当月同比	6
产量：发电量：当月同比	3	热卷现货收益率	10	产销率：当月值：同比	5
产量：水泥：当月同比	5	产量：家用电冰箱：当月同比	2		
M2：同比	10				

资料来源：Wind, 天风证券研究所

## 总结与展望

本文首先从宏观经济环境及微观品种异质性两个维度全面梳理了黑色系商品库存同比的潜在影响指标，其次通过领先滞后相关性分析筛选先行指标并确定最优先行阶数，进而构建 LASSO 模型，利用 10-折交叉验证法确定惩罚系数，对库存同比增速预测模型进行变量选择和参数估计，并预测了 2018 年 2 月各品种库存同比增速可能的方向。最后，我们利用样本外滚动预测对模型进行稳健性检验。主要结论有：

1. 样本内回归结果显示，除动力煤外，各品种的指标压缩程度均大于 50%，从样本内预测方向准确率来看，铁矿石、焦煤、焦炭及热卷四个品种的准确率高达 90%以上，拟合效果较好；
2. 6 次样本外滚动测试结果显示，各品种库存同比均具有较为稳定的先行指标，样本内拟合效果好的铁矿石和热卷的库存预测模型在样本外依然表现优异，方向预测准确率可达 100%。模型外推一期可预测 2018 年 2 月铁矿石库存同比将继续出现较大幅度上升，而热卷库存同比变化不大。

最后需要说明的是每次交叉验证下的 LASSO 估计结果将存在一定差异。这一方面是由于模型本身的原因，我们对模型精度和复杂度的权衡决定了惩罚系数  $\lambda$  的设定，另一方面主要受制于经济数据的样本量，结果的稳健性有待考证，且基本面指标众多、数据质量也参差不齐，我们无法做到像宏观或行业分析师那样专业，对于初始指标的筛选可能不尽全面。这一系列因素都将导致预测结果的偏差。而本文的主要目的是希望能提供基本面研究的一种新思路：**抓主要矛盾，寻关键变量。**

感谢实习生何青青对本文的贡献。

### 分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

### 一般声明

除非另有规定，本报告中的所有材料版权均属天风证券股份有限公司（已获中国证监会许可的证券投资咨询业务资格）及其附属机构（以下统称“天风证券”）。未经天风证券事先书面授权，不得以任何方式修改、发送或者复制本报告及其所包含的材料、内容。所有本报告中使用的商标、服务标识及标记均为天风证券的商标、服务标识及标记。

本报告是机密的，仅供我们的客户使用，天风证券不因收件人收到本报告而视其为天风证券的客户。本报告中的信息均来源于我们认为可靠的已公开资料，但天风证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，天风证券及/或其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，天风证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。

天风证券的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。天风证券没有将此意见及建议向报告所有接收者进行更新的义务。天风证券的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

### 特别声明

在法律许可的情况下，天风证券可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。因此，投资者应当考虑到天风证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突，投资者请勿将本报告视为投资或其他决定的唯一参考依据。

### 投资评级声明

类别	说明	评级	体系
股票投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	买入	预期股价相对收益 20%以上
		增持	预期股价相对收益 10%-20%
		持有	预期股价相对收益 -10%-10%
		卖出	预期股价相对收益 -10%以下
行业投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	强于大市	预期行业指数涨幅 5%以上
		中性	预期行业指数涨幅 -5%-5%
		弱于大市	预期行业指数涨幅 -5%以下

### 天风证券研究

北京	武汉	上海	深圳
北京市西城区佟麟阁路 36 号	湖北武汉市武昌区中南路 99	上海市浦东新区兰花路 333	深圳市福田区益田路 4068 号
邮编：100031	号保利广场 A 座 37 楼	号 333 世纪大厦 20 楼	卓越时代广场 36 楼
邮箱：research@tfzq.com	邮编：430071	邮编：201204	邮编：518017
	电话：(8627)-87618889	电话：(8621)-68815388	电话：(86755)-82566970
	传真：(8627)-87618863	传真：(8621)-68812910	传真：(86755)-23913441
	邮箱：research@tfzq.com	邮箱：research@tfzq.com	邮箱：research@tfzq.com