

## 专题报告-基本面量化

### 基于因果模型发现领先与同步关系



报告日期：2022 年 6 月 2 日

章顺 资深分析师(基本面量化)  
从业资格号：F0301166  
投资咨询号：Z0011689  
Tel: 8621-63325888-3902  
Email: [shun.zhang@orientfutures.com](mailto:shun.zhang@orientfutures.com)

#### ★摘要：

与以往的领先和同步关系的研究方法不同，我们使用因果模型发现多元资产之间、多元资产与基本面变量之间的领先和同步关系，该因果模型源自 Jakob Runge 的  $PCMCI^+$ 。鉴于因果模型发现领先和同步关系的效果，我们将  $PCMCI^+$  应用于金融时间序列，样本选用了 25 个资产标的且涵盖股指、国债、外汇和商品大类，还包括来自统计局和央行的 88 个基本面变量，实证结果表明：

第一，Jakob Runge 的预测方案显示 A50 因果发现模型具有预测的效果；

第二，我们设定样本内为 30 个月的数据，依据  $PCMCI^+$  识别因果关系对，通过线性算法测试因果关系对的预测效果，实证结果表明因果模型在样本外的线性预测准确率多数超过 60%且能够获得正收益；

第三，我们对比了样本内 30 个月、36 个月和 42 个月的因果发现结果，因果关系对并未出现较大的变化；

第四，我们固定 30 个月的数据窗口，每间隔 6 个月测试因果发现的结果，动态测试因果模型的预测效果，实证结果显示动态测试相对于静态测试并未明显提升预测效果。

继《多元资产之间的因果分析》之后，本文继续探讨了因果模型的应用，应用的落脚点在领先与同步关系。领先与同步关系一直是金融研究的核心问题之一，领先与同步关系意味着可预测变量的趋势，既可以预测资产标的趋势，也可以预测基本面变量的趋势。基于因果模型发现领先与同步关系，是一种新的尝试，通过因果模型寻找金融市场的关系或许会成为潮流。

**重要事项：**本报告版权归上海东证期货有限公司所有。未获得东证期货书面授权，任何人不得对本报告进行任何形式的发布、复制。本报告的信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，报告中的信息或意见并不构成交易建议，投资者据此做出的任何投资决策与本公司和作者无关。  
**有关分析师承诺，见本报告最后部分。并请阅读报告最后一页的免责声明。**

## 目录

1、 领先与同步——因果模型的视角.....	5
2、 Jakob Runge 的 PCMCI 算法概览.....	5
3、 PCMCI 框架因果发现的原理.....	7
4、 通过因果模型发现领先与同步关系.....	10
4.1、 样本数据.....	10
4.2、 实证的框架.....	12
4.3、 Jakob Runge 的预测方案.....	13
4.4、 基于因果模型测试领先与同步关系.....	15
4.4.1、 变量组合一的领先与同步关系测试.....	16
4.4.2、 变量组合二的领先与同步关系测试.....	17
4.4.3、 变量组合三的领先与同步关系测试.....	18
4.4.4、 变量组合四的领先与同步关系测试.....	19
4.4.5、 变量组合五的领先与同步关系测试.....	20
4.4.6、 变量组合六的领先与同步关系测试.....	21
4.4.7、 变量组合七的领先与同步关系测试.....	22
4.4.8、 变量组合八的领先与同步关系测试.....	22
4.4.9、 多个时间窗口的测试结果差异.....	24
4.4.10、 领先与同步关系动态测试.....	24
5、 结论及展望.....	27
6、 参考文献.....	28

## 图表目录

图表 1: PCMC1 相关算法问题.....	7
图表 2: 滞后骨架阶段伪代码 (作者: Jakob Runge) .....	9
图表 3: 同步骨架阶段伪代码 (作者: Jakob Runge) .....	10
图表 4: 基本面变量概况.....	11
图表 5: 资产标的概况.....	12
图表 6: 领先与同步关系的实证框架.....	13
图表 7: 变量组合一的 Runge 方案预测效果.....	13
图表 8: 变量组合二的 Runge 方案预测效果.....	13
图表 9: 变量组合三的 Runge 方案预测效果.....	14
图表 10: 变量组合四的 Runge 方案预测效果.....	14
图表 11: 变量组合五的 Runge 方案预测效果.....	14
图表 12: 变量组合六的 Runge 方案预测效果.....	14
图表 13: 变量组合七的 Runge 方案预测效果.....	14
图表 14: 变量组合八的 Runge 方案预测效果.....	14
图表 15: 因果模型发现领先与同步关系的效果评价.....	15
图表 16: Jakob Runge 的因果模型实证测试参数.....	16
图表 17: 基本面变量组合一的领先与同步关系测试.....	16
图表 18: RMB 样本外测试累计盈亏 (变量组合一) .....	17
图表 19: SR 样本外测试累计盈亏 (变量组合一) .....	17
图表 20: V 样本外测试累计盈亏 (变量组合一) .....	17
图表 21: USBOND 样本外测试累计盈亏 (变量组合一) .....	17
图表 22: 基本面变量组合二的领先与同步关系测试.....	18
图表 23: CNBOND 样本外测试累计盈亏 (变量组合二) .....	18
图表 24: RMB 样本外测试累计盈亏 (变量组合二) .....	18
图表 25: 基本面变量组合三的领先与同步关系测试.....	19
图表 26: 基本面变量组合四的领先与同步关系测试.....	19
图表 27: SHCU 样本外测试累计盈亏 (变量组合四) .....	20
图表 28: C 样本外测试累计盈亏 (变量组合四) .....	20
图表 29: 基本面变量组合五 (Y41-Y50) 的领先与同步关系测试.....	20
图表 30: 基本面变量组合六 (Y51-Y60) 的领先与同步关系测试.....	21
图表 31: JM_L 样本外测试累计盈亏 (变量组合六) .....	21
图表 32: 基本面变量组合七 (Y61-Y70) 的领先与同步关系测试.....	22
图表 33: 基本面变量组合八 (Y71-Y88) 的领先与同步关系测试.....	23

图表 34: L_C 样本外测试累计盈亏 (变量组合八) .....	23
图表 35: SHNI 样本外测试累计盈亏 (变量组合八) .....	23
图表 36: RB 样本外测试累计盈亏 (变量组合八) .....	24
图表 37: Y87_C 样本外测试累计盈亏 (变量组合八) .....	24
图表 38: 多个窗口测试结果中重复的领先与同步关系 .....	24
图表 39: 领先与同步关系动态测试.....	25
图表 40: 领先与同步关系动态测试在样本外的效果.....	26

## 1、领先与同步——因果模型的视角

领先与同步关系一直是金融研究的核心问题之一，通常通过线性相关性和格兰杰因果检验（GC）等方法实证分析时间序列之间的领先与同步关系。关于线性相关和格兰杰因果检验，《多元资产之间的因果分析》和《非线性相关在多元资产组合中的应用》已经对此进行了说明，这两类方法都存在局限，非线性的方法可能是适宜的视角。

许多框架利用不同的假设解决了从观测数据中发现因果关系的问题。除了基于贝叶斯分数的方法（Chickering, 2002）、经典 Granger 因果关系(Granger, 1969)以及最近的受限结构因果模型(SCM)框架工作（Peters et al., 2017, Spirtes and Zhang, 2016），基于条件独立（CI）的网络学习算法（Spirtes et al., 2000）构成了主要支柱。

在因果充分的情况下，CI 框架的主要代表是 PC 算法（Spirtes and Glymour, 1991）。首先，它的优势在于灵活地利用范围广泛且不断增长的 CI 测试类别，包括线性偏相关(ParCorr)和基于非参数残差的方法（Ramsey, 2014, Runge et al., 2019b）、基于条件互信息的测试（Runge, 2018b）和神经网络（Sen et al., 2017）。其次，PC 算法利用稀疏性使其也适用于大量变量，而基于分数和 SCM 的方法更难以适应非线性高维因果发现。

显然，时间顺序有助于识别滞后链接的因果方向（原因先于结果）。这构成了 GC 的基础，然而，GC 无法处理同时期的链接并遭受维度灾难（Runge et al., 2019b）。在 Moneta 等人发表的《Causal search in structural vector autoregressive models》一文中，PC 算法增强了 GC。然而，诸如非平稳性，尤其是自相关性等属性会使因果发现的可靠性大大降低。

实际上，自相关，一种普遍存在的时间序列属性（例如，温度数据），尤其有害。我们借鉴一种新的基于 CI 的方法  $PCMCI^+$ ，它扩展了  $PCMCI$  方法（Runge et al., 2019b），包括同时期链接的发现。 $PCMCI^+$  基于两个中心思想，基于 PC 算法和 FCI 的时间序列适应性[Entner and Hoyer, 2010, Malinsky and Spirtes, 2018]。首先，针对滞后和同期条件集分别进行边缘去除阶段，滞后阶段使用更少的 CI 测试。其次，更重要的是， $PCMCI^+$  优化了单个 CI 测试的条件集选择，使它们在自相关下得到更好的校准，并通过利用瞬时条件独立性思想提高检测能力。

接下来，我们将从因果模型的视角，实证分析多元资产之间、多元资产与基本面变量之间的领先同步关系。

## 2、Jakob Runge 的 $PCMCI$ 算法概览

$PCMCI$  算法出自《Detecting and quantifying causal associations in large nonlinear time series datasets》（Runge et al., 2019b），Runge 在 2020 年提出了改进后的  $PCMCI^+$  算法。需要注意的是，Runge 的  $PCMCI^+$  算法目前主要的实证案例集中在地球系统科学领域，Runge 也是复杂系统领域的学者。

因果推断正逐渐成为一种成熟的科学方法。与数据驱动的机器学习方法（如概率建模、深度学习）相比，机器学习主要关注预测和分类，因果推断旨在发现和量化底层系统的

因果关系。尽管解释深度学习模型是一个活跃的研究领域，但从深度学习黑匣子中提取特定现象（例如飓风）的原因通常是不容易的。因此，因果推理方法对于补充预测机器学习以提高我们对底层系统的理论理解至关重要。不幸的是，许多因果推理方法仍然只在一小部分方法开发人员中为人所知，并且很少在地球系统科学等应用领域中采用。

举一个因果推理方法在地球系统科学中应用的例子，一个热带气候示例说明西太平洋月度地表气压异常与中太平洋和东太平洋地表气温异常之间的依赖关系。相关性分析和标准双变量格兰杰因果关系得出一个完全连通的图，而多元因果方法(*PCMCI*)更好地识别了沃克环流。一般，格兰杰因果检验基于线性自回归建模，但非线性依赖关系可以用更复杂的时间序列模型甚至信息论模拟传输熵建模。*PCMCI* 系统地识别时间滞后变量之间的共同驱动因素和间接联系。

*PCMCI* 使用重建的因果网络来量化一般因果相互作用。这里提出的使用重构因果网络来量化一般因果交互的框架基于时间序列图和因果路径的概念。*PCMCI*<sup>+</sup> 扩展了

*PCMCI*，*PCMCI*<sup>+</sup> 是为解决高维时滞线性和非线性因果发现而开发的，并考虑了自相关，用于在因果充分的情况下从观察时间序列中发现线性和非线性、滞后和同步的因果关系。广泛的数值实验表明，与其他方法相比，*PCMCI*<sup>+</sup> 具有更高的邻接检测能力，尤其是更多的同步关系定向召回，同时更好地控制了误报。与 PC 算法相比，优化的条件集还可以缩短运行时间。

*PCMCI*<sup>+</sup> 在许多实际应用场景中具有相当大的用途。目前 Runge 并未使用 *PCMCI*<sup>+</sup> 算法测试金融时间序列的领先与同步关系，我们基于地球系统科学与金融复杂性理论的联系，以及复杂系统的相关思想，尝试把 *PCMCI*<sup>+</sup> 算法应用于金融市场。

*PCMCI* 相关算法问题主要是时间序列图、PC 算法和自相关的问题，具体内容如下表所示。

*PCMCI*<sup>+</sup> 的目标是优化 CI 测试中条件组的选择，以提高检测能力，同时保持良好校准的测试。该方法基于两个中心思想，(1) 将骨架边缘去除阶段分离为滞后和同期的调节阶段，CI 测试要少得多；(2) 利用瞬时条件相关 (MCI) 测试在同期调节阶段的思路。首先，PC 的骨架阶段的目标是通过在每次迭代中以变量相邻邻接的子集为条件来消除所有由间接路径和常见原因引起的邻接。



图表 1: PCMCI 相关算法问题

名称	内容	说明
时间序列图	<p>时间序列图可以表示复杂动态系统下的时间依赖结构。考虑一个潜在的离散时间结构因果过程：</p> $X_t = (X_t^1, \dots, X_t^N), X_t^j = f_j(p(X_t^j), \eta_t^j)$ <p>其中，<math>f_j</math> 是任意可测函数，<math>\eta_t^j</math> 表示独立的动态噪声。假设存在一个时间序列图 <math>G</math>，<math>G</math> 中的节点表示不同滞后时间的变量 <math>X_t^j</math>，<math>X_t^j</math> 所依赖的变量集定义了因果父项 <math>p(X_t^j) \subset X_{t+1}^- = (X_t, X_{t-1}, \dots) \setminus \{X_t^j\}</math>，用 <math>p_t^-(X_t^j) = p(X_t^j) \cap X_t^-</math> 表示滞后的父项。如果 <math>X_{t-\tau}^i \in p(X_t^j)</math>，则存在滞后(<math>\tau &gt; 0</math>)或同步(<math>\tau = 0</math>)因果关系 <math>X_{t-\tau}^i \rightarrow X_t^j</math>。</p>	<p>骨架:把有向图 <math>G</math> 的有向边变成无向边；</p> <p>马尔科夫等价:贝叶斯网络 <math>\langle G1, P1 \rangle</math> 和 <math>\langle G2, P2 \rangle</math> 马尔科夫等价,当且仅当 <math>G1</math> 和 <math>G2</math> 具有相同的框架和 <math>V</math> 结构；</p> <p>CPDAG: 设 <math>G = (V, E)</math> 是一个部分有向无环图,若 <math>E</math> 中的有向边都是可逆的,并且 <math>E</math> 中的无向边都是可逆的,则称 <math>G</math> 是一个完全部分有向无环图 (CPDAG)。</p>
PC 算法	<p>PC 算法是一种基于约束的因果结构识别算法,对于较符合假设条件的情况以及给定条件独立性准确情况下,能够得到较为可靠的识别结果。但是,PC 算法并不是稳定的,需要附加合理的假设,在结构识别中过于关注局部信息,具有约束性方法不能识别所有边方向的共性劣势,因此 PC 算法有很大程度的改进空间。PC 算法包括两个部分,第一部分是识别骨架,第二部分是根据马尔科夫等价类扩展骨架得到完全部分有向无环图 (CPDAG)。</p>	
自相关	<p>为了说明自相关的挑战,Runge 展示了一个线性示例。PC 算法首先测试所有无条件独立性 (<math>p = 0</math>),耦合对 (文中的“X5”和“X6”)也是因为 (文中的“X7”和“X8”) 独立于其他变量,并且从彼此的邻接集中移除,这表明 PC 算法如何利用稀疏性并降低估计维度,而不是像在 GC 框架中那样在整个样本拟合完整模型。</p> <p>在实验中,PC 算法具有夸大的误报,误报的另一个原因是校准不良的测试。为了正确模拟零分布,每个单独的 CI 测试都需要考虑自相关,这在复杂的多变量和潜在非线性设置中是困难的。</p>	

资料来源:东证衍生品研究院

### 3、PCMCI 框架因果发现的原理

在地球系统科学这个领域, Jakob Runge 撰写了多篇论文,其中《Discovering contemporaneous and lagged causal Relations in autocorrelated》重点介绍了  $PCMCI^+$  算法及其应用案例。 $PCMCI^+$  的目标是重建图。与适用于时间序列的标准 PC 算法相比,  $PCMCI^+$  基于两个中心思想:首先,骨架阶段对滞后和同期条件集分别进行,滞后阶段使用更少的测试。其次,  $PCMCI^+$  修改了单个条件独立测试的条件集,使其在自相关下得到很好的校准,并通过利用瞬时条件独立 (MCI) 方法提高检测能力。总而言之,  $PCMCI^+$  通过四阶段算法估计时间滞后和同时期的因果联系。

骨架发现阶段：从完全连接的图开始，首先是通过识别哪些节点对于其他节点的某个子集是条件独立的，来估计邻接的骨架  $X_t^i - X_{t-\tau}^j$ 。选择条件的具体方式见论文，这与原始 PC 算法不同。条件独立对之间的邻接被移除。然后，该骨架中的滞后邻接会按时间顺序自动定向。例如，无向链接  $X_{t-2}^i - X_t^j$  只能定向为  $X_{t-2}^i \rightarrow X_t^j$ ，因为因果效应不能及时回溯。

具有滞后条件的骨架发现阶段（与 PCMC1 相同）：对于每个变量  $X_t^j$ ，使用迭代 PC1 算法估计滞后 ( $\tau > 0$ ) 父级 ( $\hat{B}_t^-(X_t^j)$ ) 的超集。

具有同时期条件和瞬时条件独立性 (MCI) 测试的骨架发现阶段：迭代同时期邻接的子集  $S \subset X_t$  并删除所有（滞后和同时期）有序对  $(X_{t-\tau}^i, X_t^j)$  的邻接关系，其中  $X_t^j \in X_t$  和  $X_{t-\tau}^i \in X_t \cup \hat{B}_t^-(X_t^j)$ ，如果 MCI 条件独立成立 ( $\hat{B}_{t-\tau}^-(X_{t-\tau}^i)$  是  $X_{t-\tau}^i$  的父级)：

$$X_{t-\tau}^i \perp X_t^j | S, \hat{B}_t^-(X_t^j), \hat{B}_{t-\tau}^-(X_{t-\tau}^i)。$$

碰撞定向阶段：基于非屏蔽三元组定向同期链路，可选择应用保守或多数规则（也基于 MCI 测试）。基于以下碰撞规则对同时期的邻接  $X_t^i - X_t^j$  进行定向。对于具有  $\tau \geq 0$  的非屏蔽三元组  $X_{t-\tau}^k - X_t^i - X_t^j$ （对于  $\tau > 0$ ，我们总是有  $X_{t-\tau}^k \rightarrow X_t^i$ ）， $X_{t-\tau}^k$  和  $X_t^j$  之间没有邻接，如果  $X_t^i$  不是使  $X_{t-\tau}^k$  和  $X_t^j$  独立，然后将  $X_{t-\tau}^k - X_t^i - X_t^j$  定向为  $X_{t-\tau}^k \rightarrow X_t^i \leftarrow X_t^j$ 。此规则适用于所有非屏蔽三元组。

规则定向阶段：根据论文中的规则 R1-R3 定向剩余的同时期链接。

该算法生成的图包含有向（滞后和同时）链接  $X_{t-\tau}^i \rightarrow X_t^j$ ，但也包含无法根据上述阶段定向的同时期链接  $X_{t-\tau}^i - X_t^j$ 。例如，如果完全同时期三元组  $X_t^k - X_t^i - X_t^j$  中的中间节点  $X_t^i$  是使  $X_t^k$  和  $X_t^j$  独立的条件集的一部分，那么我们无法区分真正的因果关系是  $X_t^k \rightarrow X_t^i \rightarrow X_t^j$  还是  $X_t^k \leftarrow X_t^i \leftarrow X_t^j$  或  $X_t^k \rightarrow X_t^i \leftarrow X_t^j$ 。这些然后形成马尔可夫等价类。



图表 2: 滞后骨架阶段伪代码 (作者: Jakob Runge)

**Algorithm 1 (PCMCI+ lagged skeleton phase)**

Require: Time series dataset  $X = (X_1, \dots, X_N)$ , max. time lag  $\tau_{\max}$ , significance threshold  $\alpha_{\text{pc}}$ , CI test  $\text{CI}(X, Y, Z)$  returning p-value and test statistic value  $I$

- 1: for all  $X_t^j$  in  $X_t$  do
- 2:   Initialize  $\widehat{B}_t^-(X_t^j) = X_t^- = (X_{t-1}, ?, \dots, X_{t-\tau_{\max}})$  and  $I^{\min}(X_{t-\tau}^i, X_t^j) = \infty \forall X_{t-\tau}^i \in \widehat{B}_t^-(X_t^j)$
- 3:   Let  $p = 0$
- 4:   while any  $X_{t-\tau}^i \in \widehat{B}_t^-(X_t^j)$  satisfies  $|\widehat{B}_t^-(X_t^j) \setminus \{X_{t-\tau}^i\}| \geq p$  do
- 5:     for all  $X_{t-\tau}^i$  in  $\widehat{B}_t^-(X_t^j)$  satisfying  $|\widehat{B}_t^-(X_t^j) \setminus \{X_{t-\tau}^i\}| \geq p$  do
- 6:        $S =$  first  $p$  variables in  $\widehat{B}_t^-(X_t^j) \setminus \{X_{t-\tau}^i\}$
- 7:        $(p\text{-value}, I) \leftarrow \text{CI}(X_{t-\tau}^i, X_t^j, S)$
- 8:        $I^{\min}(X_{t-\tau}^i, X_t^j) = \min(|I|, I^{\min}(X_{t-\tau}^i, X_t^j))$
- 9:       if  $p\text{-value} > \alpha_{\text{pc}}$  then mark  $X_{t-\tau}^i$  for removal
- 10:   Remove non-significant entries and sort  $\widehat{B}_t^-(X_t^j)$  by  $I^{\min}(X_{t-\tau}^i, X_t^j)$  from largest to smallest
- 11:   Let  $p = p + 1$
- 12: return  $\widehat{B}_t^-(X_t^j)$  for all  $X_t^j$  in  $X_t$

资料来源: Jakob Runge, 东证衍生品研究院

因果发现最后的输出结果为数组, 具体条目说明如下:

$\text{graph}[i,j,\tau] = '-->'$  在  $\tau > 0$  时表示有向的、滞后的因果链接  $X_{t-\tau}^i \rightarrow X_t^j$ ;

$\text{graph}[i,j,0] = '-->'$  和  $\text{graph}[j,i,0] = '<--'$  表示一个有向的、同时期的因果关系  $X_t^i \rightarrow X_t^j$ ;

$\text{graph}[i,j,0] = 'oo'$  和  $\text{graph}[j,i,0] = 'oo'$  表示无方向的同时邻接  $X_t^i - X_t^j$  表示不能应用对撞机和方向规则 (马尔可夫等价);

$\text{graph}[i,j,0] = 'xx'$  和  $\text{graph}[j,i,0] = 'xx'$  表示  $X_t^i$  和  $X_t^j$  之间存在冲突的同时期邻接, 表明方向性由于方向规则冲突而未确定。

根据输出条目的信息, 我们可以梳理出多元资产之间、多元资产与基本面变量之间的领先和同步关系。

图表 3：同步骨架阶段伪代码（作者：Jakob Runge）

**Algorithm 2 (PCMCI+ / PCMCI+ ◦ contemporaneous skeleton phase / PC full skeleton phase)**

Require: Time series dataset  $X = (X_1, \dots, X_N)$ , max. time lag  $\tau_{\max}$ , significance threshold  $\alpha_{\text{pc}}$ , CI test  $\text{CI}(X, Y, Z)$

PCMCI+ / PCMCI+ ◦ :  $\widehat{B}_t^-(X_t^j)$  for all  $B \in \{-t, \dots, t\}$  for all  $X_t^j$  in  $X_t$

1: PCMCI+ / PCMCI+ ◦ : Form time series graph  $G$  with lagged links from  $\widehat{B}_t^-(X_t^j)$  for all  $X_t^j$  in  $X_t$  and fully connect all contemporaneous variables, i.e., add  $X_t^i \circ - \circ X_t^j$  for all  $X_t^i \neq X_t^j \in X_t$

PC: Form fully connected time series graph  $G$  with lagged and contemporaneous links

2: PCMCI+ / PCMCI+ ◦ : Initialize contemporaneous adjacencies  $A(X_t^j) := \widehat{A}(X_t^j) = \{X_t^i \neq X_t^j \in X_t : X_t^i \circ - \circ X_t^j \text{ in } G\}$

PC: Initialize full adjacencies  $\widehat{A}(X_t^j)$  for all (lagged and contemporaneous) links in  $G$

3: Initialize  $\widehat{B}_t^-(X_t^j) = X_{t-}^j = (X_{t-1}^j, \dots, X_{t-\tau_{\max}}^j)$  and  $I^{\min}(X_{t-}^i, X_t^j) = \infty \forall X_{t-}^i \in \widehat{B}_t^-(X_t^j)$

4: Let  $p = 0$

5: while any adjacent pairs  $(X_{t-}^i, X_t^j)$  for  $\tau \geq 0$  in  $G$  satisfy  $|A(X_t^j) \setminus \{X_{t-}^i\}| \geq p$  do

6:     Select new adjacent pair  $(X_{t-}^i, X_t^j)$  for  $\tau \geq 0$  in  $G$  satisfy  $|A(X_t^j) \setminus \{X_{t-}^i\}| \geq p$

7:     while  $(X_{t-}^i, X_t^j)$  are adjacent in  $G$  and not all  $S \subseteq A(X_t^j) \setminus \{X_{t-}^i\}$  with  $|S| = p$  have been considered do

8:         Choose new  $S \subseteq A(X_t^j) \setminus \{X_{t-}^i\}$  with  $|S| = p$

9:         PCMCI+: Set  $Z = (S, \widehat{B}_t^-(X_t^j) \setminus \{X_{t-}^i\}, \widehat{B}_{t-}^-(X_{t-}^i))$

           PCMCI+ ◦ : Set  $Z = (S, \widehat{B}_t^-(X_t^j) \setminus \{X_{t-}^i\})$

           PC: Set  $Z = S$

10:         (p-value, I)  $\leftarrow \text{CI}(X_{t-}^i, X_t^j, Z)$

11:          $I^{\min}(X_{t-}^i, X_t^j) = \min(|I|, I^{\min}(X_{t-}^i, X_t^j))$

12:         if p-value  $> \alpha_{\text{pc}}$  then

13:             Delete link  $X_{t-}^i \rightarrow X_t^j$  for  $\tau > 0$  (or  $X_t^i \circ - \circ X_t^j$  for  $\tau = 0$ ) from  $G$

14:             Store (unordered) sepset  $(X_{t-}^i, X_t^j) = S$

15:     Let  $p = p + 1$

16:     Re-compute  $A(X_t^j)$  from  $G$  and sort by  $I^{\min}(X_{t-}^i, X_t^j)$  from largest to smallest

17: return  $G$ , sepset

资料来源：Jakob Runge，东证衍生品研究院

## 4、通过因果模型发现领先与同步关系

### 4.1、样本数据

样本数据主要分为行情数据和基本面数据，具体明细如下表所示，其中基本面的指标覆盖宏观和中观，数据来自中国央行和统计局。样本数据中资产标的涵盖股指、国债、商品和外汇，内盘商品指数来自聚微编制的可交易指数，样本标的数量明显多于我们在《多元资产之间的因果分析》和《非线性相关在多元资产组合中的应用》中的样本设定。另

外，由于样本包含基本面指标，而实证部分需要对齐数据，对齐后的起点在 2017 年。

图表 4：基本面变量概况

基本面变量名称	说明
M2 同比、社会融资规模增量:当月值、社融增量人民币贷款、 制造业 PMI 指数(月)、财新中国 PMI 指数(月)、 CPI:当月同比、CPI:不包括食品和能源:当月同比、规模以上工业增加值:累计同比、 规模以上工业增加值:累计同比 (煤炭开采和洗选业、石油和天然气开采业、有色金属矿采选业、 黑色金属矿采选业、纺织业、木材加工、造纸及纸制品业、石油煤炭及其他燃料加工业、化学原料 和化学制品制造业、化学纤维制造业、黑色金属冶炼和压延加工业、有色金属冶炼和压延加工业、 电力热力生产和供应业)、社会消费品零售总额:累计同比、 固定资产投资(不含农户)完成额:累计同比、固定资产投资(不含农户)完成额:制造业:累计同比、 固定资产投资(不含农户)完成额:累计同比 (煤炭开采和洗选业、石油和天然气开采业、黑色金属矿 采选业、有色金属矿采选业、纺织业、化学原料及化学制品制造业、化学纤维制造业、黑色金属冶 炼和压延加工业、有色金属冶炼和压延加工业、信息传输等服务业、房地产业): 原煤:产量:累计同比、石油沥青生产量:累计增长、燃料油:产量:累计同比、原油加工量生产量:累计 增长、碳酸钠(纯碱):产量:累计同比、合成橡胶:产量:累计同比、乙烯:产量:累计同比、初级形态的塑 料(塑料树脂及共聚物):产量:累计同比、水泥:产量:累计同比、 平板玻璃:产量:累计同比、铁合金:产量:累计同比、生铁:产量:累计同比、粗钢:产量:累计同比、钢材: 产量:累计同比、线材(盘条):产量:累计同比、焊接钢管:产量:累计同比、 钢筋:产量:累计同比、冷轧薄板:产量:累计同比、中厚宽钢带:产量:累计同比、 发电量:累计同比、核能发电量:累计同比、成品糖:产量:累计同比、 精制食用植物油:产量:累计同比、鲜冷藏肉:产量:累计同比、饲料:产量:累计同比、 纱:产量:累计同比、布:产量:累计同比、塑料制品:产量:累计同比、 汽车:产量:累计同比、铝材:产量:累计同比、铝合金:产量:累计同比、 十种有色金属:产量:累计同比、精炼铜(铜):产量:累计同比、铜材:产量:累计同比、 原铝(电解铝):产量:累计同比、氧化铝:产量:累计同比、铁矿石原矿:产量:累计同比、 化学纤维:产量:累计同比、合成纤维:产量:累计同比、橡胶轮胎外胎:产量:累计同比、 出口总值:当月同比、进口总值:当月同比、出口总值:累计同比、进口总值:累计同比、 国家财政赤字:累计值、国家一般公共支出:合计:累计同比、 国家一般公共收入:合计:累计同比、PPI:当月同比、PPI:生产资料:当月同比、 PPI:加工:当月同比、PPI:生活资料:当月同比、PPI:耐用消费品:当月同比、PPI:一般日用品:当月同比	基本面变 量代号 Y1-Y88; 样本数据 对齐后的 区间为 2017 年 1 月 至 2022 年 3 月。

资料来源：Finoview，东证衍生品研究院

图表 5：资产标的概况

资产标的	时间区间
COMEX 铜连三 (CU) COMEX 黄金连一 (AU) WTI 连一 (WTI) 布伦特原油连一 (BRENT) 沪深 300 指数 (HS300) 十年期美国国债收益率 (%) (USBOND) 十年期中国国债收益率 (%) (CNBOND) 美元指数 (DOLLAR) 离岸人民币汇率 (RMB) 标准普尔 500 指数 (SP500) 纳斯达克 100 指数 (NASDAQ) 富时中国 A50 指数 (A50) 沪铜超额收益指数 (SHCU) 沪铝超额收益指数 (SHAL) 沪镍超额收益指数 (SHNI) 螺纹钢超额收益指数 (RB) 焦煤超额收益指数 (JM) 铁矿石超额收益指数 (I) 玉米超额收益指数 (C) 豆粕超额收益指数 (M) 白糖超额收益指数 (SR) 棉花超额收益指数 (CF) 塑料超额收益指数 (L) PTA 超额收益指数 (PTA) PVC 超额收益指数 (V) 玻璃超额收益指数 (FG)	资产标的代号 X1-X25; 离岸人民币汇率、超额收益指数时间区间偏短, 样本数据对齐后的区间为 2017 年 1 月至 2022 年 3 月, 其中超额收益指数为繁微平台开发的可交易指数。

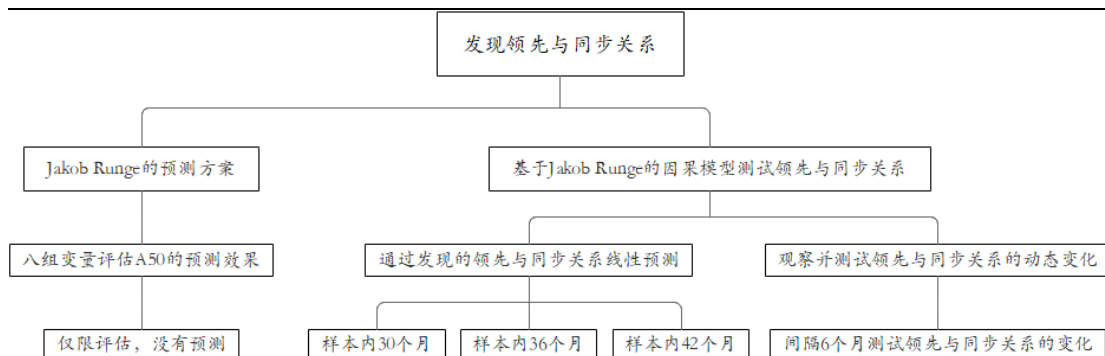
资料来源: Finoview, 东证衍生品研究院

## 4.2、实证的框架

样本数据既包含基本面指标, 也包含资产价格时间序列, 我们在实证阶段利用因果构建的网络是覆盖宏观节点的, 这与之前在《多元资产之间的因果分析》中的设定不太一样, 涵盖宏观节点的因果网络可能在因果发现方面更具优势。

为了测试因果模型发现领先与同步关系的效果, 我们在实证部分做了如下安排: 首先用 Jakob Runge 的预测方案测试一下 A50 的预测效果; 然后, 通过 Jakob Runge 的因果模型测试领先与同步关系, 只是这里的测试方法为线性算法, 具体测试方案分为静态测试和动态测试, 静态测试样本内分为 30、36 和 42 个月三种情形, 动态测试则间隔 6 个月观察预测效果的变化。

图表 6：领先与同步关系的实证框架



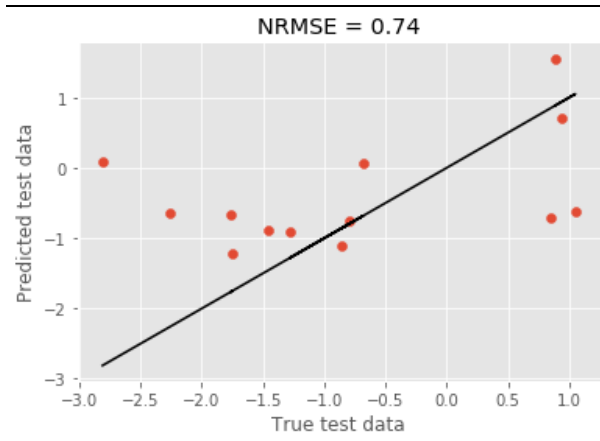
资料来源：Finoview，东证衍生品研究院

### 4.3、Jakob Runge 的预测方案

多元时间序列预测一直是一个具有挑战性的问题，“维数灾难”、“过度拟合”等问题并没有非常好的解决方案。由于预测变量的正确组合通常很重要，理想情况下，应该测试所有可能的预测变量子集的预测能力，但是组合数量的指数增长使得这种方法在计算上令人望而却步。Jakob Runge 在《Optimal model-free prediction from multivariate time series》中提出了一种克服这种强烈限制的预测方案，利用因果预选步骤，该步骤大大减少了自变量的数量，从而使全局最优搜索方案易于处理。该方法提出了一个通用框架来应用最优的无模型方法来选择变量并随后拟合模型以进一步改进预测或学习统计依赖性。Jakob Runge 通过厄尔尼诺南方涛动的气候指数说明了该框架的性能。

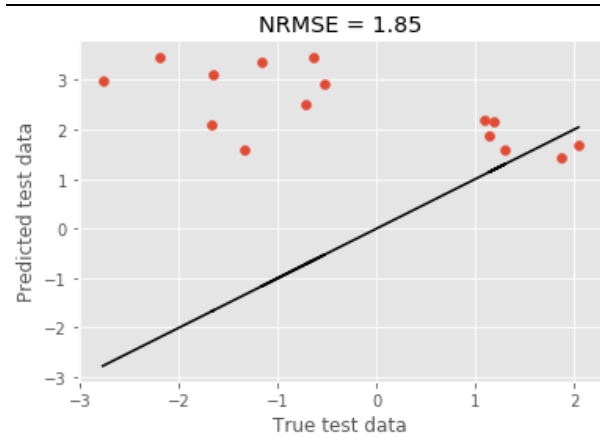
我们以富时中国 A50 指数为例，选取的样本除了所有资产标的外，基本面变量分为八组，分别进行预测的测试。下面的图表中展示的是 Jakob Runge 预测方案的评价结果，NRMSE（标准均方根误差）越小则预测效果越好，具体 Python 输出图表如下所示。

图表 7：变量组合一的 Runge 方案预测效果



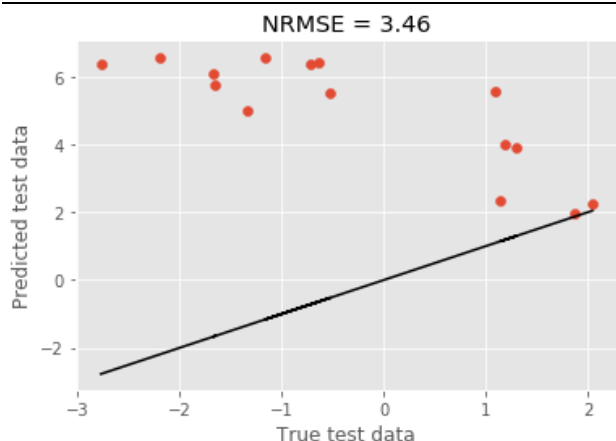
资料来源：Finoview，东证衍生品研究院

图表 8：变量组合二的 Runge 方案预测效果



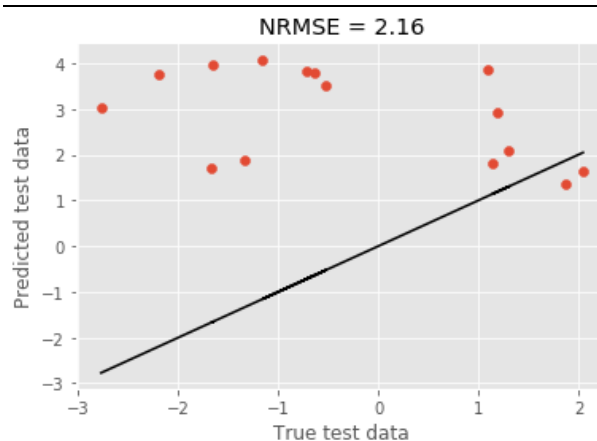
资料来源：Finoview，东证衍生品研究院

图表 9：变量组合三的 Runge 方案预测效果



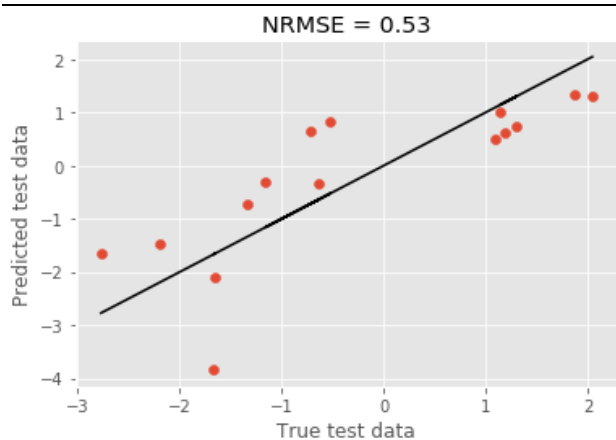
资料来源：Finoview，东证衍生品研究院

图表 10：变量组合四的 Runge 方案预测效果



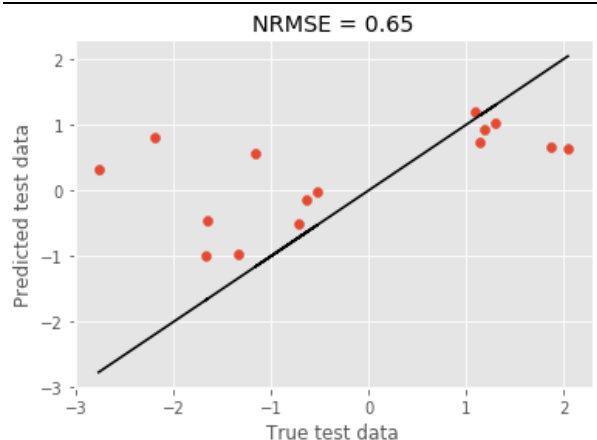
资料来源：Finoview，东证衍生品研究院

图表 11：变量组合五的 Runge 方案预测效果



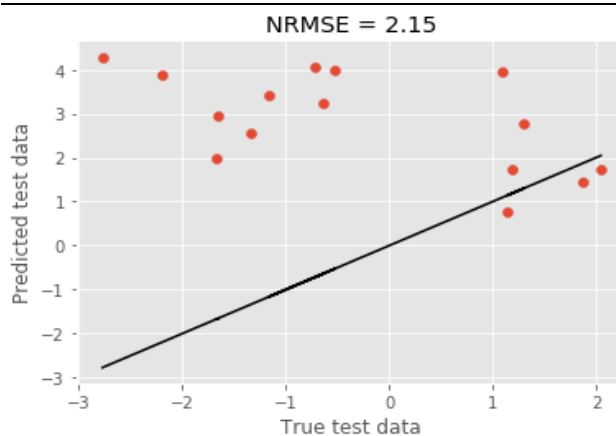
资料来源：Finoview，东证衍生品研究院

图表 12：变量组合六的 Runge 方案预测效果



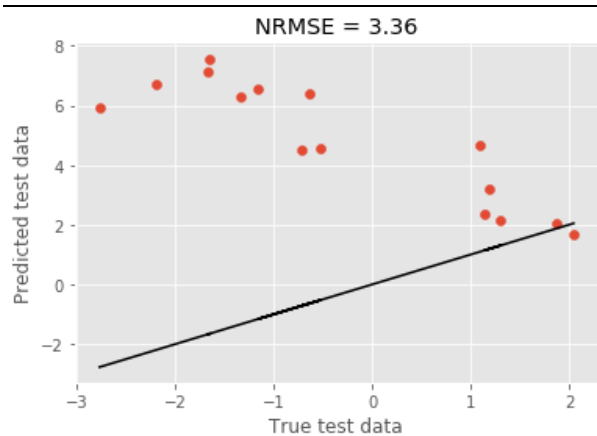
资料来源：Finoview，东证衍生品研究院

图表 13：变量组合七的 Runge 方案预测效果



资料来源：Finoview，东证衍生品研究院

图表 14：变量组合八的 Runge 方案预测效果



资料来源：Finoview，东证衍生品研究院

对比八组样本的 NRMSE（标准均方根误差）值，变量组合一、五、六的值低于 1，预



测效果高于其他样本。Jakob Runge 的预测方案在 A50 上的测试结果表明变量的差异会影响因果模型的预测效果，特别是基本面变量会显著影响 A50 的预测效果。对于 A50 而言，基本面变量对于月度行情的预测是有一定作用的。

Jakob Runge 通过厄尔尼诺南方涛动指数的实证案例检验了因果模型的效果，但预测模块的应用范围是预测时间序列的未来值，评估预测值偏离实际值的程度。针对这个问题，Jakob Runge 也提到要调整该方案，比如只需要预测厄尔尼诺等极端事件的出现。我们在对金融资产进行建模时，通常是预测金融资产的涨跌方向，因此在后文中多元资产的建模中，并未使用 Jakob Runge 的预测模块。

#### 4.4、基于因果模型测试领先与同步关系

Jakob Runge 的因果模型在发现领先和同步关系方面具有一定的效果，关于厄尔尼诺南方涛动指数的实证研究验证了模型的性能。我们借鉴 Jakob Runge 的因果模型意在发现多元资产之间、资产与基本面变量之间的领先与同步关系，具体的资产标的和基本面变量见样本数据。在发现领先与同步关系后，为了测试因果发现的效果，我们通过线性的算法来判别模型的效果，无论是资产还是基本面指标，市场更关注行情或者基本面指标的涨跌方向。

在测试领先与同步关系时，评价因果模型的算法具体设定如下图表所示。

图表 15：因果模型发现领先与同步关系的效果评价

名称	因果发现关系指向：A→B		评价算法
	A	B	
领先关系	资产标的	基本面变量	假设 A 与 B 是正向关系， $\Delta A * \Delta B$ 大于或者等于 0，则模型判断正确； 假设 A 与 B 是负向关系， $\Delta A * \Delta B$ 小于或者等于 0，则模型判断正确。 针对 $\Delta A * \Delta B$ 等于 0 的情形，如果 B 为资产标的，等于 0 时意味着不进行任何操作，当期不计入盈亏。 A 与 B 的关系，既参考相关性，也参考经济常识。
	资产标的	资产标的	
	基本面变量	资产标的	
同步关系	资产标的	基本面变量	
	资产标的	资产标的	
	基本面变量	资产标的	

资料来源：Finoview，东证衍生品研究院

与上一节中关于样本分组的设置一致，将基本面变量按照十个一组分组，最后一组变量个数超过十，共分为八组。Jakob Runge 的因果模型在实证部分的参数设置如下表所示。

Jakob Runge 的  $PCMCI^+$  模型的测试分为三个不同的测试窗口，窗口的设置参考《多元资产的因子风格及配置策略》，测试窗口均为 6 的倍数。另外为了充分观察领先与同步关系的动态变化，设置窗口动态间隔期数为 6。

图表 16: Jakob Runge 的因果模型实证测试参数

模型	样本	测试窗口	滞后期数	窗口动态间隔期数
Jakob Runge 的 <i>PCMCI</i> <sup>+</sup>	变量组合一: 资产标的+基本面变量组合一 变量组合二: 资产标的+基本面变量组合二 变量组合三: 资产标的+基本面变量组合三 变量组合四: 资产标的+基本面变量组合四 变量组合五: 资产标的+基本面变量组合五 变量组合六: 资产标的+基本面变量组合六 变量组合七: 资产标的+基本面变量组合七 变量组合八: 资产标的+基本面变量组合八	30 个月 36 个月 42 个月	0、1、2、3 (0 期代表 同步关系)	6

注意: 不同变量组合会找到相同的关系对, 这种情况下不会重复展示, 仅展示第一次出现的情形。

资料来源: Finoview, 东证衍生品研究院

#### 4.4.1、变量组合一的领先与同步关系测试

寻找领先与同步关系, 是构建预测模型的常见方法。为了进一步测试领先与同步关系在样本外模拟交易的效果, 以月度持仓测试样本外的累计盈亏, 累计盈亏为资产标的的每期绝对收益的累计值。

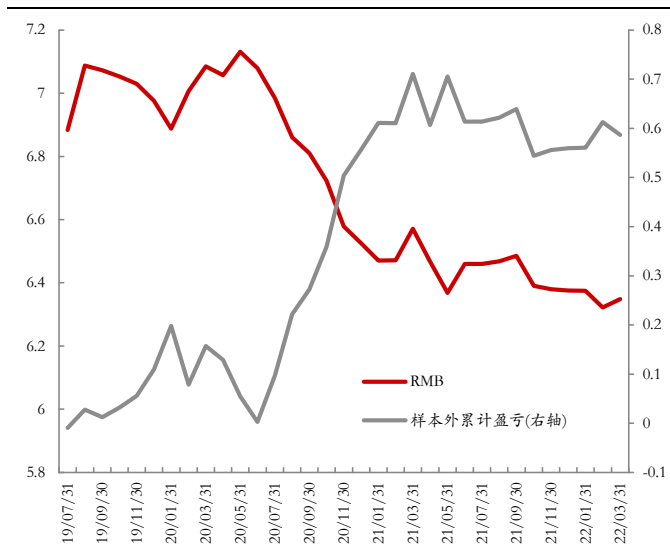
变量组合一在样本内 30 个月数据的情形下, 测试的结果如下表所示, 样本外准确率的计算参考上述图表的评价方法。观察领先与同步关系的测试结果, RMB 与 A50、RB 与 I 的同步关系显著, 这与一般的经济常识一致。需要注意的是, RMB 滞后一期的关系类似动量效应, SR 滞后一期的关系则为反转效应。

图表 17: 基本面变量组合一的领先与同步关系测试

基本面变量组合一	领先与同步关系	滞后期数	样本外准确率
M2 同比	RMB 与 RMB	1	66.7%
社会融资规模增量: 当月值	SR 与 SR	1	54.5%
社融增量人民币贷款	RMB 与 A50	0	75.8%
制造业 PMI 指数(月)	Y9 与 V	2	54.5%
财新中国 PMI 指数(月)	USBOND 与 USBOND	1	63.6%
CPI: 当月同比	RB 与 I	0	78.8%
CPI: 不包括食品和能源: 当月同比	JM 与 I	0	54.5%
规模以上工业增加值: 累计同比			
规模以上工业增加值: 煤炭开采和洗选业: 累计同比			
规模以上工业增加值: 石油和天然气开采业: 累计同比			

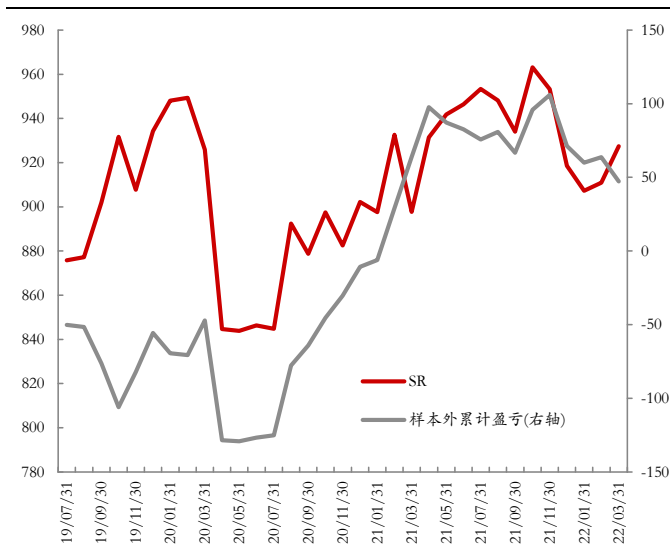
资料来源: Finoview, 东证衍生品研究院

图表 18: RMB 样本外测试累计盈亏 (变量组合一)



资料来源: Finoview, 东证衍生品研究院

图表 19: SR 样本外测试累计盈亏 (变量组合一)



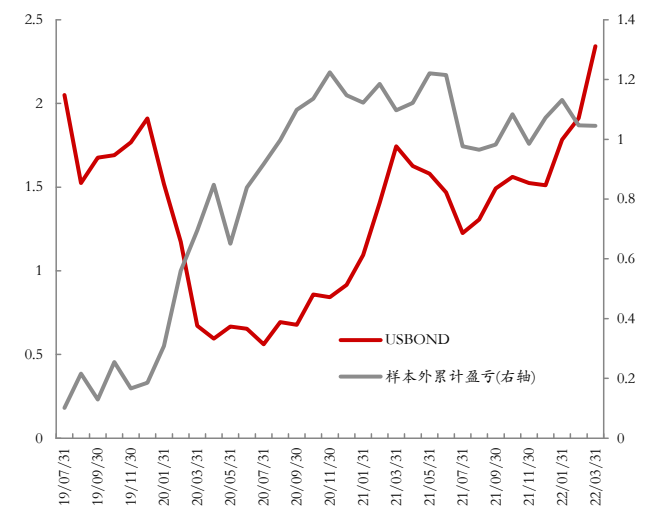
资料来源: Finoview, 东证衍生品研究院

图表 20: V 样本外测试累计盈亏 (变量组合一)



资料来源: Finoview, 东证衍生品研究院

图表 21: USBOND 样本外测试累计盈亏 (变量组合一)



资料来源: Finoview, 东证衍生品研究院

#### 4.4.2、变量组合二的领先与同步关系测试

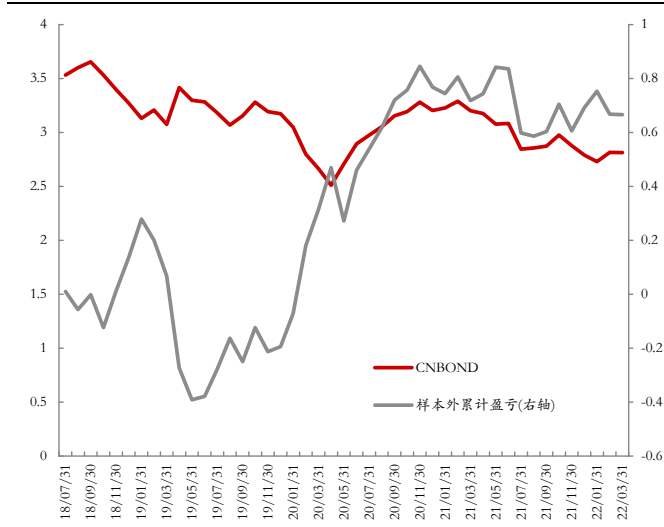
变量组合二在样本内 30 个月数据的情形下, 测试的结果如下表所示。Y11 与 Y17、Y12 与 Y15、Y17 与 Y18 的显著同步关系与经济常识一致, 中债十年国债的动量效应明显, 美元指数与人民币的领先关系暗示美元指数可能是人民币的领先指标。一般情况下, 投资者比较关注基本面变量对于资产价格的影响, 变量组合二的测试结果中展示 A50 与 Y15 的领先关系, 资产价格可能也会影响基本面的情况。

图表 22: 基本面变量组合二的领先与同步关系测试

基本面变量组合二	领先与同步关系	滞后期数	样本外准确率
规模以上工业增加值:有色金属矿采选业:累计同比 规模以上工业增加值:黑色金属矿采选业:累计同比 规模以上工业增加值:纺织业:累计同比 规模以上工业增加值:木材加工及木等制品业:累计同比 规模以上工业增加值:造纸及纸制品业:累计同比 规模以上工业增加值:石油、煤炭及其他燃料加工业:累计同比 规模以上工业增加值:化学原料和化学制品制造业:累计同比 规模以上工业增加值:化学纤维制造业:累计同比 规模以上工业增加值:黑色金属冶炼和压延加工业:累计同比 规模以上工业增加值:有色金属冶炼和压延加工业:累计同比	SP500 与 NASDAQ	0	86.7%
	Y11 与 Y17	0	77.8%
	Y12 与 Y15	0	88.9%
	Y17 与 Y18	0	97.8%
	CNBOND 与 CNBOND	1	60.0%
	DOLLAR 与 RMB	1	57.8%
	A50 与 Y15	1	68.9%

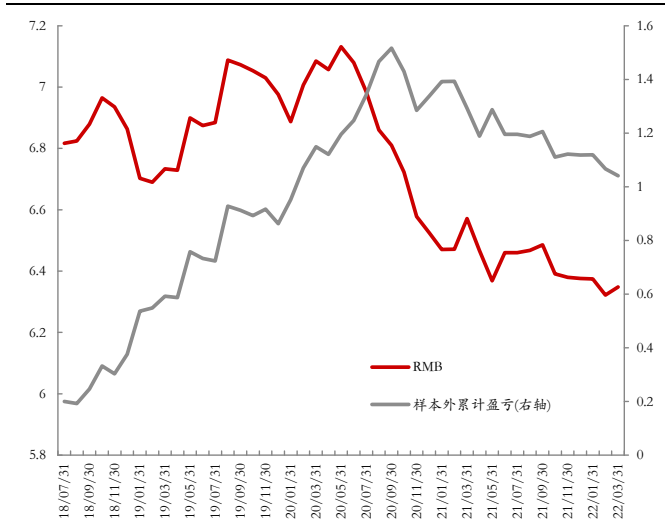
资料来源: Finoview, 东证衍生品研究院

图表 23: CNBOND 样本外测试累计盈亏 (变量组合二)



资料来源: Finoview, 东证衍生品研究院

图表 24: RMB 样本外测试累计盈亏 (变量组合二)



资料来源: Finoview, 东证衍生品研究院

#### 4.4.3、变量组合三的领先与同步关系测试

变量组合三在样本内 30 个月数据的情形下, 测试的结果如下表所示。变量组合三的测试结果中展示的领先与同步关系多数为基本面指标, 这些指标同属固定资产投资类的指标。AU 与 Y26 的滞后期数为 3, 暗示黄金价格会影响油气投资, 国际黄金与油气开采固定资产投资的关系符合经济常识。

图表 25: 基本面变量组合三的领先与同步关系测试

基本面变量组合三	领先与同步关系	滞后期数	样本外准确率
规模以上工业增加值:电力、热力生产和供应业:累计同比 社会消费品零售总额:累计同比	Y24 与 Y23	0	88.9%
固定资产投资(不含农户)完成额:累计同比	Y29 与 Y26	0	60.0%
固定资产投资(不含农户)完成额:制造业:累计同比	AU 与 Y26	3	73.3%
固定资产投资(不含农户)完成额:采矿业:煤炭开采和洗选业:累计同比	Y25 与 Y23	3	75.6%
固定资产投资(不含农户)完成额:采矿业:石油和天然气开采业:累计同比	Y30 与 Y29	3	64.4%
固定资产投资(不含农户)完成额:采矿业:黑色金属矿采选业:累计同比			
固定资产投资(不含农户)完成额:采矿业:有色金属矿采选业:累计同比			
固定资产投资(不含农户)完成额:制造业:纺织业:累计同比			
固定资产投资(不含农户)完成额:制造业:化学原料及化学制品制造业:累计同比			

资料来源: Finoview, 东证衍生品研究院

#### 4.4.4、变量组合四的领先与同步关系测试

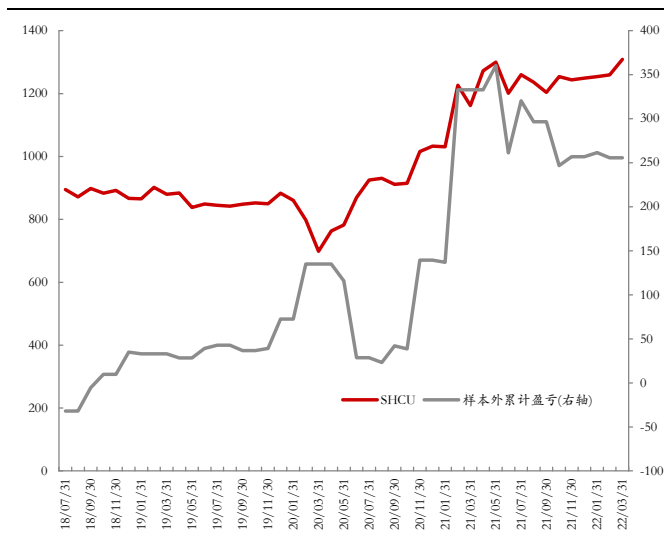
变量组合四在样本内 30 个月数据的情形下, 测试的结果如下表所示。在同步关系中, SHCU 与 CU 是符合一般常识的, AL 与 V、CF 与 SR 的基本面关系相对弱一些。在滞后期数大于 0 的结果中, Y38 与 SHCU、Y37 与 NI、Y38 与 C、Y35 与 USBOND、Y39 与 SR 都是可交易的关系对, 样本外的测试效果显著。

图表 26: 基本面变量组合四的领先与同步关系测试

基本面变量组合四	领先与同步关系	滞后期数	样本外准确率
固定资产投资(不含农户)完成额:制造业:化学纤维制造业:累计同比	SHCU 与 CU	0	86.7%
固定资产投资(不含农户)完成额:制造业:黑色金属冶炼和压延加工业:累计同比	AL 与 V	0	71.1%
固定资产投资(不含农户)完成额:制造业:有色金属冶炼和压延加工业:累计同比	CF 与 SR	0	62.2%
固定资产投资(不含农户)完成额:信息传输、软件和信息技术服务业:累计同比	Y32 与 Y37	0	80.0%
固定资产投资(不含农户)完成额:房地产业:累计同比	DOLLAR 与 Y37	1	68.9%
原煤:产量:累计同比	BRENT 与 Y39	2	57.8%
石油沥青生产量:累计增长	Y38 与 SHCU	2	71.1%
燃料油:产量:累计同比	Y37 与 NI	2	68.9%
原油加工量生产量:累计增长	Y38 与 C	2	82.2%
碳酸钠(纯碱):产量:累计同比	Y35 与 USBOND	3	73.3%
	Y39 与 SR	3	68.9%

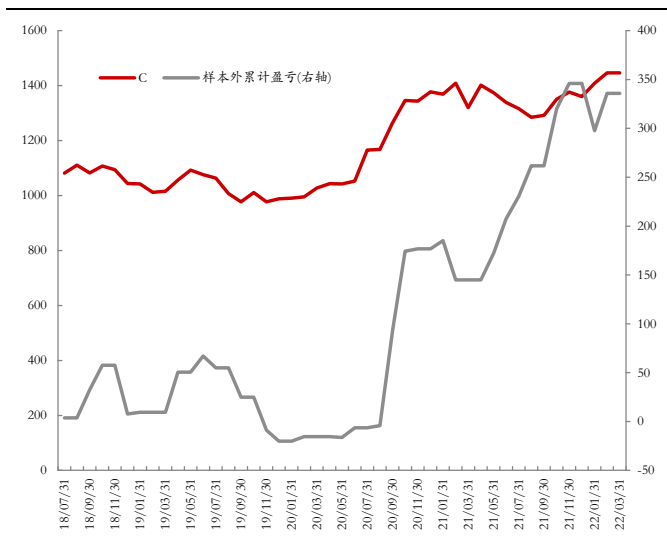
资料来源: Finoview, 东证衍生品研究院

图表 27: SHCU 样本外测试累计盈亏 (变量组合四)



资料来源: Finoview, 东证衍生品研究院

图表 28: C 样本外测试累计盈亏 (变量组合四)



资料来源: Finoview, 东证衍生品研究院

#### 4.4.5、变量组合五的领先与同步关系测试

变量组合五在样本内 30 个月数据的情形下, 测试的结果如下表所示。Y41 与 NI、Y44 与 CF 这两个关系对的测试结果表明可以进行交易的尝试, 但要注意滞后期数较长和基本面变量环比为 0 的情况, 这些问题可能会影响实际的交易结果, 测试结果仅代表理论上的可行性。在变量组合五中包含较多的商品供给数据, 大宗产量指标市场关注的焦点之一, SR 与 Y42、RMB 与 Y47 这两个关系对与一般的逻辑常识并不相符, 测试结果的表现可以作为一种额外的参考。

图表 29: 基本面变量组合五 (Y41-Y50) 的领先与同步关系测试

基本面变量组合五	领先与同步关系	滞后期数	样本外准确率
合成橡胶:产量:累计同比	Y41 与 NI	2	73.3%
乙烯:产量:累计同比	Y44 与 CF	3	73.3%
初级形态的塑料(塑料树脂及共聚物):产量:累计同比	Y44 与 Y41	3	82.2%
水泥:产量:累计同比	SR 与 Y42	3	73.3%
平板玻璃:产量:累计同比	RMB 与 Y47	3	68.9%
铁合金:产量:累计同比			
生铁:产量:累计同比			
粗钢:产量:累计同比			
钢材:产量:累计同比			
线材(盘条):产量:累计同比			

资料来源: Finoview, 东证衍生品研究院



#### 4.4.6、变量组合六的领先与同步关系测试

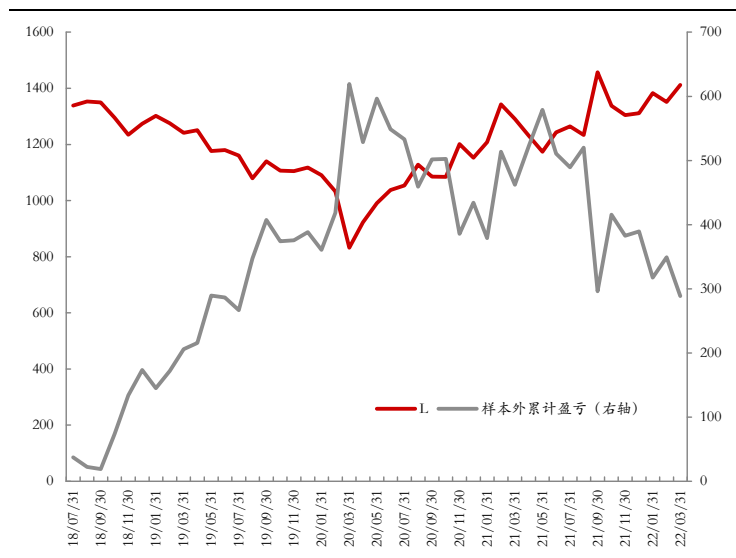
变量组合六在样本内 30 个月数据的情形下，测试的结果如下表所示。模型识别到 JM 与 L 滞后一期的关系对，JM 与 L 具有产业逻辑的支撑，符合一般的经济常识，样本外的累计盈亏也能获得正收益。SR 与 AU 是同步关系对，但白糖与国际黄金的同步关系并符合一般的基本面逻辑。

图表 30：基本面变量组合六（Y51-Y60）的领先与同步关系测试

基本面变量组合六	领先与同步关系	滞后期数	样本外准确率
焊接钢管:产量:累计同比	SR 与 AU	0	62.2%
钢筋:产量:累计同比			
冷轧薄板:产量:累计同比	Y58 与 Y51	0	84.4%
中厚宽钢带:产量:累计同比			
发电量:累计同比	JM 与 L	1	55.6%
核能发电量:累计同比			
成品糖:产量:累计同比	Y54 与 Y51	3	75.6%
精制食用植物油:产量:累计同比			
鲜冷藏肉:产量:累计同比	USBOND 与 Y56	3	82.2%
饲料:产量:累计同比			

资料来源：Finoview，东证衍生品研究院

图表 31：JM\_L 样本外测试累计盈亏（变量组合六）



资料来源：Finoview，东证衍生品研究院

#### 4.4.7、变量组合七的领先与同步关系测试

变量组合七在样本内 30 个月数据的情形下，测试的结果如下表所示。Y65 与 NI 是滞后 3 期的关系对，测试结果表明铝材产量与沪镍存在因果关系，铝和镍在产业层面也存在逻辑支撑。

图表 32：基本面变量组合七（Y61-Y70）的领先与同步关系测试

基本面变量组合七	领先与同步关系	滞后期数	样本外准确率
纱:产量:累计同比	Y63 与 Y65	0	60.0%
布:产量:累计同比			
塑料制品:产量:累计同比	Y65 与 NI	3	66.7%
汽车:产量:累计同比			
铝材:产量:累计同比	Y69 与 CF	3	64.4%
铝合金:产量:累计同比			
十种有色金属:产量:累计同比	Y62 与 Y61	3	71.1%
精炼铜(铜):产量:累计同比	Y62 与 Y65	3	73.3%
铜材:产量:累计同比			
原铝(电解铝):产量:累计同比	Y62 与 Y69	3	84.4%

资料来源：Finoview，东证衍生品研究院

#### 4.4.8、变量组合八的领先与同步关系测试

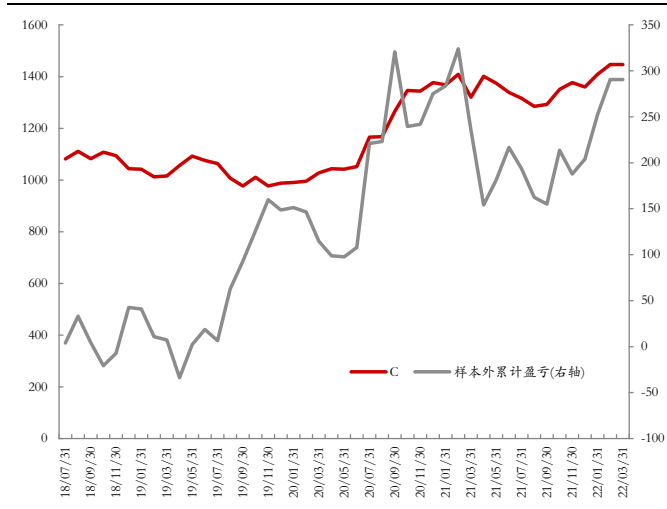
变量组合八在样本内 30 个月数据的情形下，测试的结果如下表所示。RB 与 FG 是同步关系对，螺纹钢和玻璃同属房地产相关的大宗商品，基本面逻辑清晰。L 与 C 是滞后 1 期的关系对，塑料和玉米的基本面逻辑并不强，但测试结果表明玉米累计盈亏在样本外为正收益，塑料与玉米的因果关系值得关注。Y75 与 SHNI 是滞后 2 期的关系对，橡胶轮胎外胎产量与镍在逻辑上存在联系，SHNI 在样本外测试中累计盈亏为正收益。另外，在样本外累计盈亏的测试中表现较好的还有 Y87 与 RB、Y87 与 C 两个关系对。

图表 33: 基本面变量组合八 (Y71-Y88) 的领先与同步关系测试

基本面变量组合八	领先与同步关系	滞后期数	样本外准确率
氧化铝:产量:累计同比	RB 与 FG	0	73.3%
铁矿石原矿:产量:累计同比	L 与 C	1	57.8%
化学纤维:产量:累计同比	Y82 与 Y76	1	80.0%
合成纤维:产量:累计同比	Y75 与 SHNI	2	60.0%
橡胶轮胎外胎:产量:累计同比	Y87 与 CNBOND	3	75.6%
出口总值:当月同比、进口总值:当月同比	Y87 与 RB	3	82.2%
出口总值:累计同比、进口总值:累计同比	Y87 与 C	3	64.4%
国家财政赤字:累计值	Y81 与 Y76	3	66.7%
国家一般公共支出:合计:累计同比			
国家一般公共收入:合计:累计同比			
PPI:当月同比、PPI:生产资料:当月同比			
PPI:加工:当月同比、PPI:生活资料:当月同比			
PPI:耐用消费品:当月同比、PPI:一般日用品:当月同比			

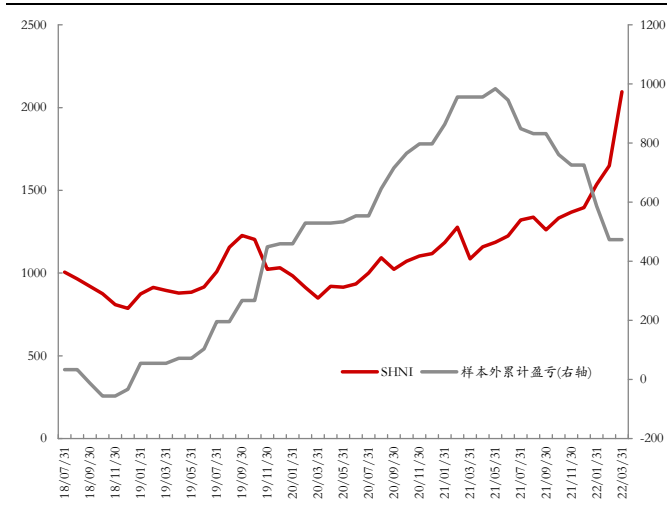
资料来源: Finoview, 东证衍生品研究院

图表 34: L\_C 样本外测试累计盈亏 (变量组合八)



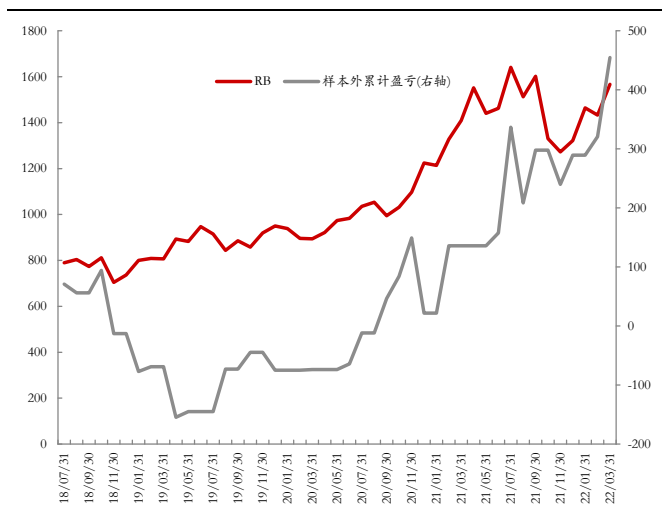
资料来源: Finoview, 东证衍生品研究院

图表 35: SHNI 样本外测试累计盈亏 (变量组合八)



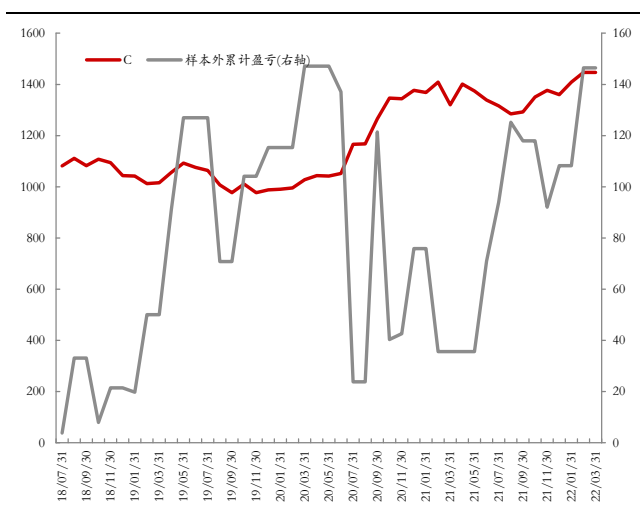
资料来源: Finoview, 东证衍生品研究院

图表 36: RB 样本外测试累计盈亏 (变量组合八)



资料来源: Finoview, 东证衍生品研究院

图表 37: Y87\_C 样本外测试累计盈亏 (变量组合八)



资料来源: Finoview, 东证衍生品研究院

#### 4.4.9、多个时间窗口的测试结果差异

上述针对八个变量组合的测试使用的样本内数据为 30 个月。为了观察样本内数据长度是否影响因果发现的结果,我们将样本内 30 个月、36 个月、42 个月三种不同时间窗口因果发现的结果进行对比,重复的因果关系对如下表所示,扩大样本内的数据长度并没有明显改变因果发现的结果。

图表 38: 多个窗口测试结果中重复的领先与同步关系

测试窗口	滞后期数: 0	滞后期数:1	滞后期数: 3
30 个月 36 个月 42 个月	RMB 与 A50 RB 与 I RB 与 FG Y58 与 Y51	USBOND 与 USBOND、RMB 与 RMB、I 与 I、SR 与 SR、 V 与 V、CNBOND 与 CNBOND、DOLLAR 与 DOLLAR、 DOLLAR 与 RMB、Y24 与 Y24、Y28 与 Y28、 Y34 与 Y34、Y35 与 Y35、Y36 与 Y36、JM 与 L、 Y53 与 Y53、Y64 与 Y64、L 与 C、Y82 与 Y76	AL 与 CNBOND Y25 与 Y23 Y30 与 Y29 Y44 与 CF Y54 与 Y51

注: Y1~Y88 为基本面变量,具体见样本数据说明。

资料来源: Finoview, 东证衍生品研究院

#### 4.4.10、领先与同步关系动态测试

前面的静态测试中,样本内的数据是固定的窗口。为了测试因果关系的动态变化,我们将样本内的数据窗口固定为 30 个月,但是每间隔 6 个月进行测试,因果发现的测试期分为 2017 年 7 月-2019 年 12 月、2018 年 1 月-2020 年 6 月、2018 年 6 月-2020 年 12 月、2019 年 1 月-2021 年 6 月,各期因果发现的结果如下表所示。

图表 39: 领先与同步关系动态测试

30 个月动态测试窗口 (间隔 6 个月)	滞后期数: 0	滞后期数:1	滞后期数:2	滞后期数: 3
2017 年 7 月- 2019 年 12 月	-	SHNI_SHNI、AU_SHNI、 Y4_SHAL、L_L、 Y5_Y5、CNBOND_CNBOND、 RMB_RMB、Y10_Y8	I_USBOND RB_CF SHCU_Y7	SHCU_DOLLAR CNBOND_RB Y8_CF、SHUCU_Y10
2018 年 1 月- 2020 年 6 月	Y3_Y2	AU_AU、JM_JM、C_C、FG_SR、 Y8_V、JM_CNBOND、 SHNI_SHNI、I_I、C_SR、Y8_Y1、 Y8_Y2、Y6_Y6、Y10_Y10	DOLLAR_I	USBOND_NASDAQ HS300_M Y6_CNBOND
2018 年 6 月- 2020 年 12 月	SP500_FG	SHNI_SHNI、AU_AU、C_C Y5_SP500、Y6_Y6	Y9_SR	Y6_CNBOND
2019 年 1 月- 2021 年 6 月	-	USBOND_USBOND CNBOND_CNBOND NASDAQ_NASDAQ SHNI_USBOND、Y6_Y6	-	SHNI_SR

资料来源: Finoview, 东证衍生品研究院

领先与同步关系动态测试结果表明, 间隔 6 个月的因果关系会发生变化。针对这一问题, 我们通过进一步的测试来观察动态测试能否提升因果关系的预测能力。通过因果关系建立的线性预测方案显示, 动态测试并未明显改善因果关系的判断准确率。

综上, 因果模型是发现领先与同步关系的一种新方法, 与传统的相关性分析和格兰杰因果检验有明显的区别, 我们在实证部分对 25 个资产标的和 88 个基本面变量的分析结果显示因果模型能够识别多元资产之间、多元资产与基本面指标之间的领先和同步关系。另外, 因果发现的领先与同步关系有些并没有频繁变化, 可能是变量之间内在存在一些联系, 比如我们经常关注的逻辑, 这里的逻辑既有产业逻辑, 也有宏观逻辑。

图表 40: 领先与同步关系动态测试在样本外的效果

30 个月动态测试窗口（间隔 6 个月）	滞后期数	领先与同步关系	判断准确率
2017 年 7 月-2019 年 12 月	0	-	
	1	CNBOND_CNBOND	63.0%
		RMB_RMB	66.7%
		SHNI_SHNI	63.0%
		Y10_Y8	70.4%
2	I_USBOND	51.9%	
	RB_CF	51.9%	
	SHCU_Y7	59.3%	
3	Y8_CF	55.6%	
	SHCU_Y10	70.4%	
2018 年 1 月-2020 年 6 月	0	Y3_Y2	90.5%
	1	JM_CNBOND	66.7%
		SHNI_SHNI	71.4%
		I_I	57.1%
		C_SR	57.1%
		Y8_Y1	71.4%
		Y8_Y2	57.1%
Y6_Y6		66.7%	
Y10_Y10	85.7%		
2	-		
3	HS300_M	76.2%	
	Y6_CNBOND	52.4%	
2018 年 6 月-2020 年 12 月	0	SP500_FG	66.7%
	1	SHNI_SHNI	73.3%
		Y6_Y6	66.7%
	2	Y9_SR	53.3%
3	Y6_CNBOND	53.3%	
2019 年 1 月-2021 年 6 月	0	-	
	1	USBOND_USBOND	66.7%
		NASDAQ_NASDAQ	55.6%
		SHNI_USBOND	55.6%
		Y6_Y6	77.8%
2	-		
3	SHNI_SR	55.6%	

资料来源: Finoview, 东证衍生品研究院



## 5、结论及展望

与以往的领先和同步关系的研究方法不同，我们使用因果模型发现多元资产之间、多元资产与基本面变量之间的领先和同步关系。该因果模型源自 Jakob Runge 的  $PCMCI^+$ ， $PCMCI^+$  改进了 PC 算法，在地球系统科学领域有实证分析的案例，并未应用于金融时间序列的相关研究。

鉴于因果模型发现领先和同步关系的效果，我们将  $PCMCI^+$  应用于金融时间序列，样本选用了 25 个资产标的且涵盖股指、国债、外汇和商品大类，还包括来自统计局和央行的 88 个基本面变量，实证结果表明：

- 第一，Jakob Runge 的预测方案显示 A50 因果发现模型具有预测的效果；
- 第二，我们设定样本内为 30 个月的数据，依据  $PCMCI^+$  识别因果关系对，通过线性算法测试因果关系对的预测效果，实证结果表明因果模型在样本外的线性预测准确率多数超过 60% 且能够获得正收益；
- 第三，我们对比了样本内 30 个月、36 个月和 42 个月的因果发现结果，因果关系对并未出现较大的变化；
- 第四，我们固定 30 个月的数据窗口，每间隔 6 个月测试因果发现的结果，动态测试因果模型的预测效果，实证结果显示动态测试相对于静态测试并未明显提升预测效果。

继《多元资产之间的因果分析》之后，本文继续探讨了因果模型的应用，应用的落脚点在于领先与同步关系。领先与同步关系一直是金融研究的核心问题之一，领先与同步关系意味着可预测变量的趋势，既可以预测资产标的趋势，也可以预测基本面变量的趋势。基于因果模型发现领先与同步关系，是一种新的尝试，通过因果模型寻找金融市场的关系或许会成为潮流。

## 6、参考文献

- [1]Chickering, D. M. (2002). Learning Equivalence Classes of Bayesian-Network Structures. J. Mach. Learn. Res., 2:445 – 498.
- [2]Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3):424 – 438.
- [3]Spirtes, P., Glymour, C., and Scheines, R. (2000). Causation, Prediction, and Search. MIT Press, Boston, MA.
- [4]Spirtes, P. and Zhang, K.(2016). Causal discovery and inference: concepts and recent methodological advances. Appl. Informatics, 3(1):3.
- [5]Peters, J., Janzing, D., and Schölkopf, B. (2017). Elements of causal inference: foundations and learning algorithms. MIT Press, Cambridge, MA.
- [6]Sen, R., Suresh, A. T., Shanmugam, K., Dimakis, A. G., and Shakkottai, S. (2017). Model-Powered Conditional Independence Test. In Proc. 30th Conf. Adv. Neural Inf. Process. Syst., pages 2955 – 2965.
- [7]Entner, D. and Hoyer, P. O. (2010). On causal discovery from time series data using FCI. In Proc. Fifth Eur. Work. Probabilistic Graph. Model., pages 121 – 128.
- [8]Ramsey, J. D. (2014). A Scalable Conditional Independence Test for Nonlinear, NonGaussian Data.
- [8]Malinsky, D. and Spirtes, P. (2018). Causal structure learning from multi-variate time series in settings with unmeasured confounding. In Proc. of 2018 ACM SIGKDD Work. On Causal Discovery, pages 23 – 47.
- [9]Runge et al. (2019a). Inferring causation from time series in earth system sciences. Nature Comm., 10(1):2553.
- [10]Runge et al. (2012). Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy. Phys. Rev. E, 86(6):061121.
- [11]Runge et al. (2019b). Detecting and quantifying causal associations in large nonlinear time series datasets. Science Advances, eaau4996(5).

### 期货走势评级体系（以收盘价的变动幅度为判断标准）

走势评级	短期（1-3 个月）	中期（3-6 个月）	长期（6-12 个月）
强烈看涨	上涨 15%以上	上涨 15%以上	上涨 15%以上
看涨	上涨 5-15%	上涨 5-15%	上涨 5-15%
震荡	振幅-5%-+5%	振幅-5%-+5%	振幅-5%-+5%
看跌	下跌 5-15%	下跌 5-15%	下跌 5-15%
强烈看跌	下跌 15%以上	下跌 15%以上	下跌 15%以上

### 上海东证期货有限公司

上海东证期货有限公司成立于 2008 年，是一家经中国证券监督管理委员会批准的经营期货业务的综合性公司。东证期货是东方证券股份有限公司全资子公司，注册资本金 23 亿元人民币，员工近 600 人。公司主要从事商品期货经纪、金融期货经纪、期货投资咨询、资产管理、基金销售等业务，拥有上海期货交易所、大连商品交易所、郑州商品交易所和上海国际能源交易中心会员资格，是中国金融期货交易所全面结算会员。公司拥有东证润和资本管理有限公司，上海东祺投资管理有限公司和东证期货国际（新加坡）私人有限公司三家全资子公司。

东证期货以上海为总部所在地，在大连、长沙、北京、上海、郑州、太原、常州、广州、青岛、宁波、深圳、杭州、西安、厦门、成都、东营、天津、哈尔滨、南宁、重庆、苏州、南通、泉州、汕头、沈阳、无锡、济南等地共设有 33 家营业部，并在北京、上海、广州、深圳多个经济发达地区拥有 134 个证券 IB 分支网点，未来东证期货将形成立足上海、辐射全国的经营网络。

自 2008 年成立以来，东证期货秉承稳健经营、创新发展的宗旨，坚持市场化、国际化、集团化的发展道路，打造以衍生品风险管理为核心，具有研究和技术两大核心竞争力，为客户提供综合财富管理平台的一流衍生品服务商。

## 分析师承诺

### 章顺

本人具有中国期货业协会授予的期货执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接接收到任何形式的报酬。

## 免责声明

本报告由上海东证期货有限公司（以下简称“本公司”）制作及发布。

本研究报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本研究报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的报告之外，绝大多数研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买投资标的的邀请或向人作出邀请。

在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者需自行承担风险。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

如需引用、刊发或转载本报告，需注明出处为东证衍生品研究院，且不得对本报告进行任何有悖原意的引用、删节和修改。

## 东证衍生品研究院

地址：上海市中山南路 318 号东方国际金融广场 2 号楼 21 楼

联系人：梁爽

电话：8621-63325888-1592

传真：8621-33315862

网址：[www.orientfutures.com](http://www.orientfutures.com)

Email：[research@orientfutures.com](mailto:research@orientfutures.com)