

GO WITH THE CUSTOMERS

股指期货跨期套利量化模型专题

【2010 年 12 月 14 日】

Wind 资讯金融情报所

上海万得信息技术股份有限公司
Shanghai Wind Information Co., Ltd.

地 址: 上海市浦东新区福山路 33 号建工大厦 9 楼
邮编 Zip: 200120
电话 Tel: (8621) 6888 2280
传真 Fax: (8621) 6888 2281
Email: sales@wind.com.cn
<http://www.wind.com.cn>

引言：

本文主要以期指的跨期套利为研究对象，期指投机和期现套利暂不做展开，主要介绍运用统计套利手段来揭示跨期套利机会。跨期套利实质上是一种价差交易，即合约间的价差偏离均衡价差而出现过高或过低的状态时，相应地卖出或买入价差。价差的不确定性决定了跨期套利不是一种无风险套利。跨期套利的困难之处在于何时采取行动构建头寸和何时获利了结。对于跨期套利的时机和价差波动范围以及趋势的把握是跨期套利的关键。

传统的股指期货跨期套利方法是基于持有成本理论，而持有成本理论中的股息收益率不易确定，限制其方法的应用，且基于持有成本理论的套利通常需要经历较长时间才能完成，因为它的价差会一直处于偏高或偏低的状态，只有在合约临近到期时才有回归的压力，因此这里主要介绍基于统计套利的跨期套利。

统计套利实质上是一种利用不同到期月份合约间价差偏离长期均衡状态构建对冲套利头寸而获利的交易方式。该方法有别于纯粹对赌近期和远期合约走势的投机套利交易行为。统计套利的最大特点是不需要对市场未来的方向进行判断，该方法使得价差序列分析更具有客观性。

基于统计套利的跨期套利模型，本文主要介绍了基于协整的跨期套利和基于 GARCH 模型的跨期套利。股指期货的两个不同期货合约因为对应着同一个股票指数，所以可能存在长期的均衡关系。基于协整的跨期套利的核心在于准确发现价差偏离均衡的时机和概率，利用协整方法构建不同到期月份合约之间的长期均衡关系，估计价差的分布，制定适当的价差交易策略。基于 GARCH 模型的统计套利是考虑到了很多金融时间序列数据具有异方差性，GARCH 模型相应地在时变方差的基础上制定交易策略，使得交易策略更加准确和稳健。本文分析了 IF1103 和 IF1012 合约之间基于统计套利的跨期套利。

1. 股指期货跨期套利的概念

股指期货跨期套利是在同一交易所进行同一指数但不同交割月份的套利活动。股指期货远月合约价格与近月合约的价格往往有一定的价差关系，如果价差偏离一定水平，就存在跨期获取低风险收益的机会。实际上，跨期套利买卖的是不同交割期的股指期货合约间的价差。跨期套利按操作方向的不同分为三种类别，分别为牛市跨期套利（多头套利）、熊市跨期套利（空头套利）和蝶式跨期套利。

牛市跨期套利是指如果我们判断远月合约涨幅将大于近月合约涨幅，或远月合约跌幅将小于近月合约跌幅，就可以买入远月合约，同时卖出近月合约。从价差的角度看，做牛市套利的投资者看多股市，认为远月合约涨幅将大于近月合约的涨幅，或者说远月合约跌幅将小于近期合约的跌幅。

熊市跨期套利是指如果我们判断近月合约涨幅将大于远月合约涨幅，或近月合约跌幅将小于远月合约跌幅，我们就可以买入近月合约，同时卖出远月合约。熊市套利与牛市套利相反，即看空股市，认为远月合约的跌幅将大于近月合约，或者说远月合约的涨幅将小于近月合约涨幅。

蝶式套利是指两个方向相反、共享中间交割月份的跨期套利的组合，即同时进行三个交割月份的合约买卖，通过中间交割月份合约与前后两交割月份合约的价差的变化来获利。当投资者认为中间交割月份的股指期货合约与两边交割月份合约价格之间的价差将发生变化时，会选择采用蝶式套利。

跨期套利的难点在于何时采取行动构建头寸和何时获利了结，合理估计并预测价差波动的范围和趋势是跨期套利交易的关键问题。

2. 统计套利

统计套利方法，是指通过研究一段时间两个不同交割月份的实际价差的波动情况，确定一个合理的波动范围，当之后的实际价差偏离该范围时，就可以进行

套利操作。它是一种基于历史价差水平的统计方式来挖掘价差稳定性和变量间的长期均衡关系，从而制定相对客观的跨期套利，其无需对行情进行预计和估计，并能够挖掘最大化的套利机会。

2.1 基于协整的跨期套利

2.1.1 协整的概念

假定一些经济指标被某经济系统联系在一起，那么从长远看来这些变量应该具有均衡关系，这是建立和检验模型的基本出发点。在短期内，因为随机干扰，这些变量有可能偏离均值。如果这种偏离是暂时的，那么随着时间推移或回到均衡状态；如果这种偏离是持久的，就不能说这些变量之间存在均衡关系。协整（co-integration）可被看作这种均衡关系性质的统计表示。

因为金融时间序列数据往往具有非平稳性，早期对金融时间序列进行普通线性回归就容易产生“伪回归”问题，而对非平稳性序列进行差分虽然解决了“伪回归”的问题，但又丢失了数据之间的长期信息。

20 世纪 80 年代 Engle 和 Granger 提出了协整的概念。协整定义：对于随机向量 $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{Nt})^T$ ，如果已知：（1） $\mathbf{x}_t \sim I(d)$ （即 \mathbf{x}_t 中每一个分量都是 d 阶非平稳的）；（2）存在一个 $N \times 1$ 阶列向量 $\boldsymbol{\beta} (\boldsymbol{\beta} \neq 0)$ ，使得 $\boldsymbol{\beta}^T \mathbf{x}_t \sim I(d-b)$ ，则称变量 $x_{1t}, x_{2t}, \dots, x_{Nt}$ 存在阶数为 (d, b) 的协整关系，用 $\mathbf{x}_t \sim CI(d, b)$ 表示。 $\boldsymbol{\beta}$ 称为协整向量。 $\boldsymbol{\beta}$ 的元素称为协整参数。

对于协整的检验有两种方法，即 EG 两步法和 Johansen 协整检验。其中 EG 两步法适用于检验两个变量之间的协整关系，而 Johansen 协整检验适用于检验多个变量之间的协整关系。下面主要是检验两个序列之间的协整关系，所以我们采用 EG 两步法。

2.1.2 基于协整的跨期套利思路

基于协整的跨期套利思路如下：

第一，对各序列进行处理，利用单位根检验检验序列的稳定性，一般大多数经济金融时间序列都是不平稳的，但若两个序列是同阶单整的，那么两个序列之间可能存在协整关系。

第二，检验协整关系的存在，通过 OLS 回归后对得出的残差进行平稳性检验，如果残差是平稳的，则两个序列之间存在长期的均衡关系，基于协整的统计套利方法就可以实施。

第三，根据对价差的统计分析来制定跨期套利的策略。

2.1.3 协整检验

选取 IF1103 和 IF1012 两个合约 2010 年 11 月 24 日至 2010 年 12 月 7 日之间的 1 分钟高频价格数据进行协整分析，所用软件是 EViews5.0。首先对 IF1103、IF1012 序列进行单位根检验，检验结果如表 1 所示：

表 1 IF1012 和 IF1103 的单位根检验结果

变量	检验类型 (c, t, p)	ADF值	临界值			是否平稳
			1%	5%	10%	
IF1012	(0, 0, 0)	0.5958	-2.5658	-1.9409	-1.6166	否
IF1103	(0, 0, 0)	0.5958	-2.5658	-1.9409	-1.6166	否
Δ IF1012	(0, 0, 0)	-48.8201	-2.5658	-1.9409	-1.6166	是
Δ IF1103	(0, 0, 0)	-57.4037	-2.5658	-1.9409	-1.6166	是

数据来源：wind 资讯

图 1 IF1012 和 IF1103 原始和一阶差分数据平稳性检验 EViews 截图

Null Hypothesis: IF1012 has a unit root
Exogenous: None
Lag Length: 1 (Automatic based on SIC, MAXLAG=27)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.532438	0.8309
Test critical values: 1% level	-2.565814	
5% level	-1.940940	
10% level	-1.616621	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(IF1012) has a unit root
Exogenous: None
Lag Length: 0 (Automatic based on SIC, MAXLAG=27)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-48.82013	0.0001
Test critical values: 1% level	-2.565814	
5% level	-1.940940	
10% level	-1.616621	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: IF1103 has a unit root
Exogenous: None
Lag Length: 1 (Automatic based on SIC, MAXLAG=27)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.595772	0.8450
Test critical values: 1% level	-2.565814	
5% level	-1.940940	
10% level	-1.616621	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(IF1103) has a unit root
Exogenous: None
Lag Length: 0 (Automatic based on SIC, MAXLAG=27)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-57.40370	0.0001
Test critical values: 1% level	-2.565814	
5% level	-1.940940	
10% level	-1.616621	

*MacKinnon (1996) one-sided p-values.

由表 1 可知，IF1012 和 IF1103 序列都含有单位根，但两个序列的一阶差分是平稳的，即都是一阶单整，既然是同阶单整，就可进行协整分析。

利用一分钟高频价格数据对 IF1103、IF1012 进行回归可得：

$$\text{IF1103} = 1.023539\text{IF1012} + \text{resid} \\ (t=21931.5)$$

模型的残差单位根检验结果如下：

表 2 残差的单位根检验结果

变量	检验类型 (c, t, p)	ADF值	临界值			是否平稳
			1%	5%	10%	
Resid	(0, 0, 0)	-1.8806	-2.5658	-1.9409	-1.6166	是

数据来源：wind 资讯

图 2 残差 Resid 平稳性检验 EViews 截图

Null Hypothesis: RD has a unit root
Exogenous: None
Lag Length: 8 (Automatic based on SIC, MAXLAG=27)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.880610	0.0573
Test critical values: 1% level	-2.565816	
5% level	-1.940941	
10% level	-1.616621	

*MacKinnon (1996) one-sided p-values.

由表 2 和图 2 可知，残差的单位根检验表明其基本上是平稳的，所以 IF1103 和 IF1012 存在协整关系。

2.1.4 交易策略和结果

因为 IF1103 和 IF1012 之间的价差与残差 Resid 基本表现出同步变动，所以用残差 Resid 表示 IF1103 和 IF1012 之间的价差 Spread，根据协整分析结果，可以得到价差序列 Spread。其基本统计量如下：

表 3 Spread 基本统计量

Mean	Median	Maximum	Minimum	Std.Dev.	Skewness	Kurtosis	Jarque-Bera	Probability
0.001116	-1.8969	30.39187	-30.49464	7.74135	0.30663	2.20286	114.5113	0.00000

数据来源：wind 资讯

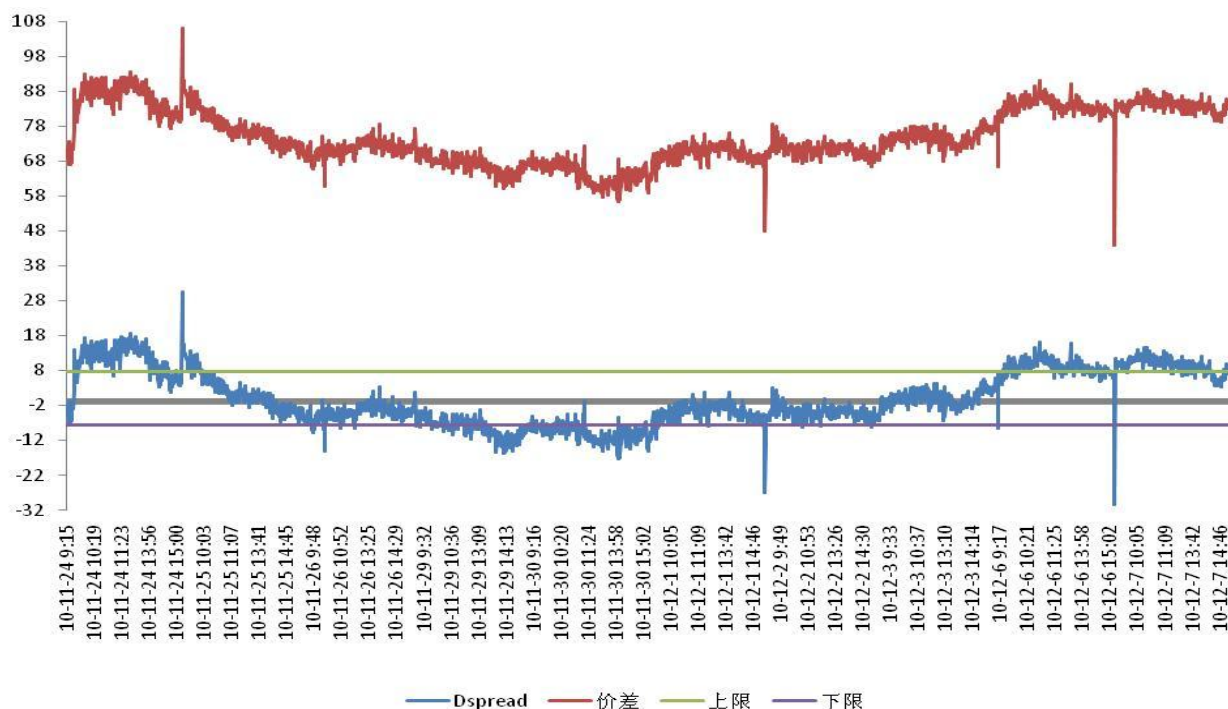
根据价差序列的均值 mean 将 Spread 中心化，令 $Dspread = Spread - mean$ ，利用得到的 Dspread 序列建立如下套利策略：

1. 当 $Dspread_t < -\sigma$ ($Dspread_t$) 时，进行牛市跨期套利；
2. 当 $Dspread_t > \sigma$ ($Dspread_t$) 时，进行熊市跨期套利；
3. 建立套利头寸后，当 $Dspread_t$ 回落至正负一个标准差区间时，进行反向操作，了解套利头寸；

4. 当近期合约到期后，无论价差是否回落，反向对冲了结套利头寸，不进行展期操作。

得到的套利区间和走势如下图：

图 3 基于协整的跨期套利



数据来源：wind 资讯

表 4 中列出了基于协整的跨期套利结果，其中毛盈利是指相应阈值得出的最大的理论上不包括成本的盈利，操作上的毛盈利可能低于这个值，仅供参考。实际上我们还可以 $\pm 1.1\sigma$, $\pm 1.2\sigma$, $\pm 1.3\sigma$, $\pm 1.4\sigma$, $\pm 1.5\sigma$... 为跨期套利阈值，通过试验来看，阈值设定的越大（即上下限之间的距离越大），相应的套利机会越少，每次套利收益可能越大。另外，价差一超过上下限，理论上可以立即套利，但有时价差可能继续朝着有利的方向运动，也许等一下可能会收获更大的收益，所以除了阈值的设定外，跨期套利和投资者的风险偏好、投资经验等有很大关系，因此理论上的收益只是作为参考，这里只提供方法，具体的最大投资收益还需要投资者根据自己的情况找出自己的阈值和投资模式。表 6 中列出的情况也只是一种阈值假定下的收益，仅供参考。

表 4 基于协整的跨期套利结果统计

统计指标	统计值
样本个数	2717
牛市套利次数	5
熊市套利次数	13
跨期套利总次数	18
毛盈利	190.6（点）
平均毛盈利	10.6（点）
一次套利所需最长时间	1.06 天
一次套利所需最短时间	1 分钟
一次套利平均所需时间	0.56 天

数据来源：wind 资讯

2.2 基于 GARCH 模型的跨期套利

2.2.1 GARCH 模型的概念

恩格尔和克拉克（Kraft, D., 1983）在分析宏观数据时，发现一些现象：时间序列模型中的扰动方差稳定性通常比假设的要差。恩格尔的发现说明在分析通货膨胀模型时，大的及小的预测误差常常会成群出现，表明存在一种异方差，其中预测误差的大小取决于后续扰动项的大小。

从事于股票价格、通货膨胀率和外汇汇率等金融时间序列预测的研究工作者发现，对这些变量的预测能力随时期的不同而有相当大的变化。预测的误差在某一时期中相对地小，而在某一时期中则相对地大，然后，在另一时期又是较小的。这种变化很可能由于金融市场的波动性易受谣言、政局变动、政府货币政策与财政政策变化等的影响，从而有理由相信误差项的条件方差不是某个自变量的函数，

而是随时间变化且依赖于过去误差的大小。

在 GARCH 模型中，要考虑两个不同的设定：一个是条件均值，另一个是条件方差。标准的 GARCH(1,1)模型为：

$$y_t = x_t \gamma + u_t, t=1, 2, \dots, T$$

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2$$

其中： x_t 是 $1 \times (k+1)$ 维外生变量向量， γ 是 $(k+1) \times 1$ 维系数向量。

2.2.2 GARCH 模型的建立

上面是利用协整检验方法，采取了较简单的策略制定方法，即以恒定不变的标准差作为套利交易的阈值。实际上， $Spread_t$ 序列自身的方差并非恒定不变，和其他大多数金融时间序列一样， $Spread_t$ 也具有时变方差特性，需要对价差的方差进行预测，下面建立 GARCH(1,1)模型对样本外价差的时变方差进行估计。

选取 2010 年 11 月 24 日至 12 月 7 日的 IF1103 与 IF1012 之间的价差 1 分钟数据作为样本内数据，以 12 月 8 日-12 月 9 日的价差 1 分钟数据作为样本外数据。通过样本内数据建立 GARCH 模型，之后用该模型对样本外价差的方差作出估计，在这基础上制定交易策略。所用软件是 EViews5.0。

$$Spread_t = 1.64750 + 0.31385 Spread_{t-1} + 0.26212 Spread_{t-2} + 0.24793 Spread_{t-3} + 0.15310 Spread_{t-4} + \varepsilon_t$$

$$\sigma_t^2 = 2.33201 + 0.29602 \varepsilon_{t-1}^2 + 0.18890 \sigma_{t-1}^2$$

对均值方程的残差项 ε_t 进行 ARCH LM 检验，表明不存在 ARCH 效应，见表 5。模型的 $R^2=0.92969$ ， $DW=1.99807$ ，表明 AR(4)-GARCH(1,1)模型较好的拟合了序列 $Spread_t$ ，可用此模型预测样本外方差分布。

表 5 ARCH TEST

F-statistic	0.012114	Probability	0.999711
Obs*R-squared	0.048543	Probability	0.999710

数据来源: wind 资讯

将 GARCH(1,1)模型的结果带入样本外数据:

$$resid02 = Spread_t - 0.31385 Spread_{t-1} - 0.26212 Spread_{t-2} - 0.24793 Spread_{t-3} - 0.15310 Spread_{t-4} - 1.64750$$

$$resid03 = resid02 / Std.Dev(resid02)$$

图 4 AR(4)-GARCH(1,1)模型 EViews 截图

Dependent Variable: SPREAD
Method: ML - ARCH
Date: 12/10/10 Time: 10:40
Sample (adjusted): 5 2717
Included observations: 2713 after adjustments
Convergence achieved after 144 iterations
Variance backcast: ON
GARCH = C(6) + C(7)*RESID(-1)^2 + C(8)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
C	1.647498	0.372511	4.422685	0.0000
SPREAD(-1)	0.313849	0.024691	12.71124	0.0000
SPREAD(-2)	0.262122	0.023043	11.37547	0.0000
SPREAD(-3)	0.247933	0.023238	10.66950	0.0000
SPREAD(-4)	0.153102	0.021731	7.045292	0.0000

Variance Equation				
C	2.332010	0.109816	21.23571	0.0000
RESID(-1)^2	0.296021	0.024101	12.28243	0.0000
GARCH(-1)	0.188901	0.038945	4.850412	0.0000

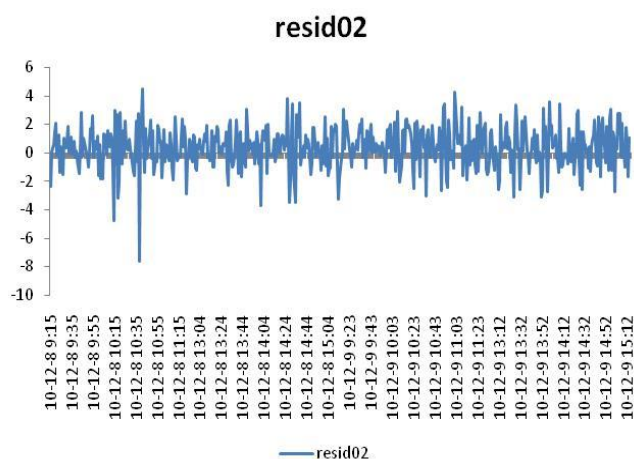
R-squared	0.929694	Mean dependent var	74.90992
Adjusted R-squared	0.929512	S.D. dependent var	7.768211
S.E. of regression	2.062424	Akaike info criterion	4.187590
Sum squared resid	11505.97	Schwarz criterion	4.205005
Log likelihood	-5672.466	F-statistic	5109.964
Durbin-Watson stat	1.998069	Prob(F-statistic)	0.000000

2.2.3 交易策略和结果

所采取的交易策略是:

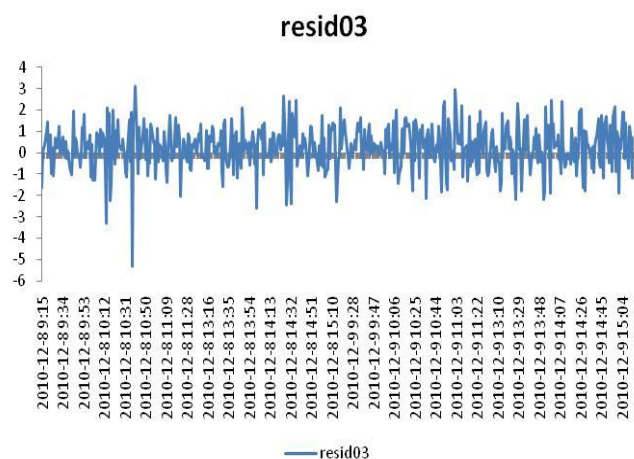
1. 当 $resid03 < -1$ 时, 进行牛市跨期套利;
2. 当 $resid03 > 1$ 时, 进行熊市跨期套利;
3. 建立套利头寸后, 当 $resid03$ 回落至区间 $[-1,1]$ 时, 进行反向操作, 了解套利头寸;
4. 当近期合约到期后, 无论价差是否回落, 反向对冲了结套利头寸, 不进行展期操作。

图 5 resid02 序列



数据来源: wind 资讯

图 6 resid03 序列



数据来源: wind 资讯

表 6 基于 GARCH 模型的跨期套利结果统计 (12 月 8 日-9 日)

统计指标	统计值
样本个数	544
牛市套利次数	3
熊市套利次数	5
跨期套利总次数	8
毛盈利	49 (点)
平均毛盈利	6.1 (点)
一次套利所需最长时间	201 分钟
一次套利所需最短时间	1 分钟
一次套利平均所需时间	0.25 天

数据来源: wind 资讯

比较基于协整的统计套利和基于 GARCH 模型的统计套利, 我们看出基于 GARCH 模型的统计套利每次所需的套利时间明显缩短了, 虽然基于 GARCH 模型的统计套利平均毛盈利比基于协整的统计套利有所下降, 但基于 GARCH 模型的

统计套利相对稳健和精确，风险相对较小。

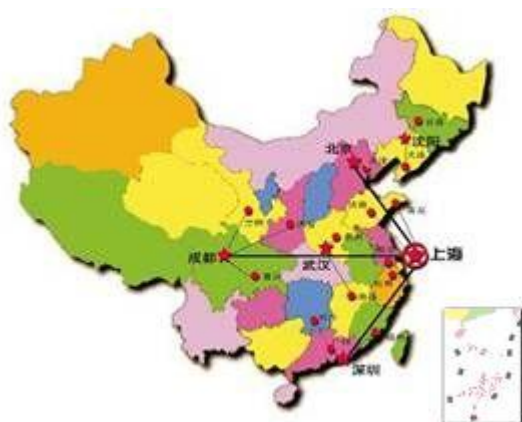
3 . 小结

统计套利实质上是一种利用不同到期月份合约间价差偏离长期均衡状态构建对冲套利头寸而获利的交易方式。该方法有别于纯粹对赌近期和远期合约走势的投机套利交易行为，为股指期货跨期套利提供了一种新的思路和方法。

然而，基于统计套利的跨期套利在执行过程中还有一些地方值得思考：一是，统计套利选取历史数据来建立模型，数据变化时无法反映变量之间的新的关系，因此，模型应不断地根据新包括进来的数据及时更新；二是，阈值的设定和投资者的风险偏好问题，不管基于协整还是基于 GAECH 模型，不同的阈值设定会导致不同的交易行为，从而带来不同的收益，另外不同的风险偏好和投资模式的投资者也会在相同的阈值设定下有不同的收益。

需要强调的是，跨期套利并不是无风险套利。除了价差风险，保证金水平也是跨期套利的风险来源。市场波动剧烈时，要求投资者能有一定的活动资金应对保证金风险。另外，设定好止盈止损的阈值对于投资者来说也比较重要。

The World' s Leading Provider of Chinese Financial Data and Solutions



上海万得信息技术股份有限公司
Shanghai Wind Information Co., Ltd.
 地 址: 上海市浦东新区福山路 33 号
 建工大厦 9 楼
 邮编 Zip: 200120
 电话 Tel: (8621) 6888 2280
 传真 Fax: (8621) 6888 2281
 Email: sales@wind.com.cn
<http://www.wind.com.cn>

由于本文档含有具体的与技术、业务相关的专有信息，文中涉及的所有资料必须严格保密。贵公司作为我方授权客户，需承诺绝不向第三方透露本文档所有内容。