

我国股票市场指数及指数证券投资组合

范龙振, 王海涛, 何 华

(复旦大学管理学院, 上海 200433)

摘要: 运用主成分分析法分析我国上海和深圳两个交易所几个市场指数对市场变化的反应情况, 结果表明两个股票市场的综合指数和A股指数可以反应市场的变化, 而其他指数不能反映各自代表的股票市场变化。由于这4个指数都不是一个好的投资组合, 要得到与指数一致的投资收益需要构造指数投资组合。本文利用多因子定价模型, 结合统计分析和优化方法, 从每个股票市场上选取20余支股票, 经过适当的组合就可以得到与指数一致的收益。

关键词: 市场指数; 指数证券组合; 聚类分析; APT模型

中图分类号: F830.91 **文献标识码:** A **文章编号:** 1007-9807(2002)05-0011-07

0 引言

在我国, 随着投资者足够理智和成熟, 证券市场的法规逐步健全和发展, 股票市场显示出了明显的效率。股票价格变动越来越具有个性, 大盘对个股的影响较1995年以前明显减弱。多元化投资分散个股特有风险的优势更为突出, 多元化、指数化投资的市场条件逐渐成熟。正是在这种市场条件和国家政策的鼓励下, 投资基金在我国问世, 并取得了快速的发展。投资基金分主动型和被动型。投资基金被动型投资基金的投资对象往往是某一支股票指数, 这种投资节省了大量的分析和交易管理费用, 其取得的收益直接与指数的变化一致。主动型的投资基金通过对市场的分析增大了某些股票的持有比重, 期望可以战胜市场, 但一般来讲它们的经营和收益也以市场指数为参照指标, 因此指数的质量对于基金运作具有重要的作用。

在我国, 推出指数期货也是一个热门话题, 构造一个能够作为指数期货标的的指数是推出指数期货首先要解决的问题。

市场指数一般来说有两个重要作用: (1) 反映

它所代表的股票的总体变化; (2) 能够告诉其所代表的投资组合策略的收益。作为代表在一个交易所交易股票的市场指数无论在上海还是在深圳都有几个, 哪一个能够准确反映市场的变动是首先研究的问题。在现有的指数中, 能够反应市场变化的指数可能不是一个好的投资对象, 如后面可以知道上证综合指数能够反应市场变化, 但现在在上海股票市场上市的公司有5百多家, 如果按每支股票的比重投资于每一支股票, 非有巨量资金不可。怎么能在有限的资金下投资一个股票资产组合, 其收益能与代表市场整体变动的股票指数一致是研究的另一个问题。

本文首先研究两个交易所现有指数能否反应市场变化, 然后从每个市场选取一个最能反映市场变动的股票指数作为投资模拟对象, 希望能从市场上选出小部分股票, 在一定的组合权重下组成一个资产组合, 其变化能与这个指数变化足够相同, 也就是具有一致的收益率。虽然不能期望这个资产组合在相当长的时间里与指数变动完全相同, 但希望它能够在这段时间里, 至少是3个月里应该足够相同。随着时间的增加, 这个组合可能有一定偏离, 可以对它进行动态调整。也就是可以

构造一个动态投资组合来与市场一样的收益
考虑到交易成本的存在, 希望调整的次数足够少,
每次调整的资金量也比较少.

1 市场变化与市场指数

为了研究指数对市场变化的反映, 首先要能够定量衡量市场变化. 本文选取在两个交易所交易的, 具有一定交易时间长度的所有股票作为的观察对象. 从数据分析上看, 观察时间要比较长, 观测的股票支数越多越好. 但由于证券市场在我国历史较短, 上市公司数量在前几年较少, 两者很难同时满足. 本文选取的观察时间长度是3年. 上海股票交易所的样本观测时间从1998年9月1日至2000年9月1日. 在这段时间里在上海股票交易所交易的股票中有406支股票有完整的周收盘价数据记录, 剩下的一些股票则存在不同程度的数据不足现象. 因此就以这406支股票来代表上海证券交易所在这段时间交易的所有股票. 深圳股票交易所的样本时间区间从1997年9月30日至2000年9月30日. 深圳股票市场这段时间共有488支股票交易, 舍弃上市日期在1997年10月1日以后的股票, 以及在样本区间内停牌日超过4周以上的股票. 停牌日小于4周的股票, 由于停牌日之前和之后一个交易日的股价差别不大, 因此采取在停牌日收益率补0的方法. 如此得到331支股票的时间序列数据. 本文利用周度数据进行分析. 数据来源于路透金融系统, 其价格数据已经对除现金红利以外的配股、送股、股票转赠进行了调整, 在这个价格的基础上, 再把现金红利的发放考虑进去, 计算出各支股票的周收益率. 周收益率用连续复利收益率表示, 即

$$x_{it} = \log(S_{it} + d_{it}) - \log S_{it-1} \quad (1)$$

式中: S_{it} 代表各支股票在时间 t 的价格; X_{it} 代表收益率; d_{it} 代表红利

为了从各支股票变化中, 分离出收益率的共同变动因素, 首先对这些收益率数据进行主成分分析. 从主成分分析结果看, 深圳股票市场第一个主成分能解释价格变动达到31.3%, 上海股票市场第1个主成分能解释价格变动达到32.2%. 如果说第1主成分反映了影响各只股票价格变化的共同因素导致价格发生的变动, 这些共同因素就

是市场的系统风险因素, 因此第1主成分反映了市场的变化. 从第1主成分的解释率30%左右可以看出, 从平均意义上来说, 个股变动30%左右是由市场变动引起的, 70%左右的变动是由非市场变动引起的, 而非市场变动主要归结于企业特有因素造成的变化, 这可以从第二主成分以后的各个主成解释力很差这一点上看出, 上海股票市场55个主成分的累计贡献率才刚超过90%. 从分析结果也可以看出, 个股变化已经显示出个性, 通过投资分散化可以分散掉个股风险的70%左右, 多元化投资在中国已经显示出投资优势. 但与美国市场相比, 美国市场的个股变化平均27%是由市场变化引起的, 分散化投资优势在美国市场变相更为明显.

第一主成分反映了市场的变化, 而反映市场变化也是指数的首要功能, 因此指数与主成分的区别可以用来判断指数是否反映了市场变化. 现在市场上存在的用来反映上海股票市场整体变化的指数主要有4个: 上证A股指数, 上证B股指数, 上证综指和上证30指数. 而深圳股票市场反应大盘的指数有深圳A股指数, 深圳A股成分指数, 深圳综合指数, 深圳综合成分指数, 深圳B股指数, 深圳B股成分指数. 它们与各自市场的第一主成分相关系数见表1.

表1 第一主成分和指数之间的相关系数

1A 深圳股票市场						
指数种类	SZSA	SZSA1	SZSC	SZSC1	SZSB	SZSB1
SZPR N1	0.976	0.873	0.976	0.8725	0.469	0.429

1B 上海股票市场						
指数种类	SSEA	SSEB	SSEC	SSE30		
SSPR N1	0.983	0.506	0.982	0.877		

注: SZSA 是深证A股指数, SZSA1 是深证A股成分指数, SZSC 和 SZSC1 分别是深证综合指数, 综合成分指数. SZSB 和 SZSB1 分别是深证B股指数和成分指数. SZPR N1 代表深证的第一主成分. SSEA, SSEB, SSEC, SSE30 分别是上海股票市场A股指数, B股指数, 综合指数和30成分指数. SSPR N1 是上证的第一主成分.

从结果来看, 在深圳市场, 第一主成分与深圳A股指数和深圳综合指数的相关性最大. 在上海股票市场, 第一主成分与上证A股指数和上证综指的相关性最大. 也就是两个市场的综合指数和A股指数最能反映市场的变化. 这个结果一方面

与直观的判断相吻合, 因为这两个指数分别是在两个股票市场上交易的所有股票的加权平均, 它们应当反映市场的整体变化; 97.6% 和 98% 以上的相关系数也说明了两个市场的综合指数虽然利用总市值作为权重, 但这种做法并没有影响它们对流通股股票市场变化的反映, 这种做法造成的误差极小

两个市场 B 股指数与第一主成分差异很大, 两者的相关系数只达到 0.5 左右. 由于第一主成分反映的是 A 股市场的价格变化, 这说明 A 股市场与 B 股市场是两个完全不同的市场, 两者的差异很大

深圳两个成分指数和上海 30 指数分别是由在两个市场上上市的部分股票组成的成分指数, 是容易实现的投资组合, 但它们与第一主成分有较大差异, 它们并不能充分反应市场的变化, 不是一个风险充分分散化的投资组合

2 指数证券组合的构造

上面已经看到两个市场 A 股指数最能反映 A 股市场的整体变化, A 股指数就可以看成 A 股市场充分分散化的证券投资组合, 这个组合只反映市场系统风险, 是一个极好的投资对象. 但是由于组合含有在整个市场交易的所有股票, 还要对不断上市的股票的进行调整, 投资起来操作性很差, 不是一个好的投资组合. 下面设法从整个股票市场选取较少的股票, 根据合适的权重进行投资, 其投资组合收益率能与 A 股指数有一样的收益. 也就是说, 希望能够构造一个证券投资组合, 其收益率的变动与选定的市场指数的变动具有一致性. 这个证券投资组合也叫指数证券组合.

要构造出一个指数证券组合, 关键有两点:

1) 从市场的所有股票中选取一部分股票, 作为构造指数证券组合的投资对象, 称之为证券组

合投资群体 (portfolio population). 本文选取 50 支股票, 这 50 只股票与选定指数的变动要尽可能的一致. 本文采用聚类分析方法来做这一点

2) 对证券投资群体中的每一支股票, 选择合适的投资比例, 使得到的证券投资组合即指数证券组合与选定的市场指数充分接近. 本文以 APT 模型作指导, 通过时间序列主成分分析和最优化求解来做这一点

2.1 证券组合投资群体

利用聚类方法选取投资群体, 其中的每一支股票要与市场指数变化尽量一致, 指数证券组合模拟指数的效果才会更好. 证券或指数变化的一致性可以用 $E(x_{it} - x_{jt})^2$ 的大小来衡量, 其中: x_{it} 和 x_{jt} 是证券或指数的收益率. $E(x_{it} - x_{jt})^2$ 越小, 证券或指数的变动越具有一致性. 根据样本数据, $E(x_{it} - x_{jt})^2$ 的大小可以由式 (2) 来估计.

$$\hat{E}(x_{it} - x_{jt})^2 = \frac{1}{T} \sum_{j=1}^T (x_{it} - x_{jt})^2 \quad (2)$$

等式右边的 $x_{it}, x_{jt}, t = 1, 2, \dots, T$ 是样本观测期间证券或指数的实际观测值, 其中 T 是样本观测长度. 因此定义聚类用的距离为

$$D_{ij} = \sum_{j=1}^T (x_{it} - x_{jt})^2 \quad (3)$$

分别对两支股票市场进行分析. 在上海股票市场, 把上证 A 股指数收益率时间序列与 406 支股票的收益率时间序列放在一个数据集中, 进行转置后, 使用 SAS 中的快速聚类 (FASTCLUS) 过程, 该过程根据指定类中心的数据集 (此处为上证 A 股指数周收益率时间序列), 求出各股票周收益率时间序列与它的距离. 然后可以按照输出数据集找到收益率与上证 A 股指数最接近的股票, 选择 50 支股票作为投资对象. 在深圳股票市场, 把深圳 A 股指数和所有 331 支股票收益率序列放在一起使用同样的方法做聚类分析, 选择与指数距离最短的 50 支股票作为投资对象. 表 2 给出的是上海股票市场选出的 50 支股票

表 2 被选中的作为投资对象的 50 支股票

code	distance								
R 6002	0.34	R 6080	0.35	R 6168	0.29	R 6678	0.35	R 6764	0.34
R 6054	0.33	R 6084	0.34	R 6170	0.28	R 6684	0.28	R 6773	0.33
R 6055	0.35	R 6087	0.35	R 6609	0.35	R 6698	0.35	R 6810	0.36

续表2 被选中的作为投资对象的50支股票

code	distance								
R 6059	0 35	R 6093	0 31	R 6619	0 35	R 6702	0 34	R 6820	0 34
R 6060	0 30	R 6104	0 36	R 6626	0 33	R 6713	0 32	R 6823	0 36
R 6062	0 34	R 6113	0 33	R 6638	0 36	R 6717	0 30	R 6854	0 35
R 6063	0 35	R 6116	0 35	R 6639	0 35	R 6726	0 35	R 6860	0 33
R 6066	0 33	R 6129	0 33	R 6649	0 35	R 6727	0 34	R 6866	0 33
R 6068	0 32	R 6162	0 33	R 6655	0 34	R 6736	0 30	R 6872	0 35
R 6074	0 34	R 6163	0 36	R 6663	0 32	R 6740	0 31	R 6890	0 36

注: Code 指股票代码, Distance 指股票与上证A 指的样本距离, R 6002 代表 600002, 其他股票代码同样

2.2 指数证券组合的构造

已经确定了50支可供投资的股票, 下面要确定这50支股票的投资权重 d_1, d_2, \dots, d_p (此处 $p = 50$), 使这个投资组合与指数变化一致

根据APT模型, 假定市场上各个股票的回报率受 q 个风险因子的影响, 它们的回报率可以由下式表达:

$$x_{it} = \gamma_{i0} + \gamma_{i1}g_{1t} + \dots + \gamma_{iq}g_{qt} + \epsilon_{it} \quad (4)$$

$i = 0, 1, 2, \dots, p$

其中:

$$E(\epsilon_{it}) = 0$$

$$\text{cov}(\epsilon_{it}, \epsilon_{jt}) = 0, \quad i \neq j$$

$$\text{cov}(\epsilon_{it}, \epsilon_{t-k}) = 0, \quad k \neq 0$$

假定 $\{x_{pt}\}$ 是所要构造的指数证券组合的收益率, 其在股票 i 的投资权重为 d_i , 则有

$$x_{pt} = d_1x_{1t} + \dots + d_px_{pt} \quad (5)$$

$t = 1, \dots, T$

x_{pt} 可由多因子模型表示为

$$x_{pt} = \gamma_{p0} + \gamma_{p1}g_{1t} + \dots + \gamma_{pq}g_{qt} + \epsilon_{pt} \quad (6)$$

式中

$$x_{pt} = \sum_{i=1}^p d_i x_{it}, \quad \gamma_{pj} = \sum_{i=1}^p d_i \gamma_{ij}, \quad \epsilon_{pt} = \sum_{i=1}^p d_i \epsilon_{it}$$

$j = 0, 1, \dots, q$

指数证券组合的收益率 $\{x_{pt}\}$ 应充分接近于选定的市场指数的收益率 $\{x_{0t}\}$, 其中: $x_{0t} = \log S_{0t} - \log S_{0t-1}$, S_{0t} 是指数的取值 两者充分近相当于要求 $E(x_{pt} - x_{0t})^2$ 尽可能的小 根据样本观测值, 可以估计出 $E(x_{pt} - x_{0t})^2$ 的取值为

$$\hat{E}(x_{pt} - x_{0t})^2 = \frac{1}{T} \sum_{t=1}^T (x_{pt} - x_{0t})^2 \quad (7)$$

x_{0t} 也服从多因子模型, 即

$$x_{0t} = \gamma_{00} + \gamma_{01}g_{1t} + \dots + \gamma_{0q}g_{qt} + \epsilon_{0t} \quad (8)$$

要使指数证券组合和指数充分近, 首先要保证投

资组合的风险因子系数和指数的风险因子系数一样 在这个前提下, 希望随机误差尽量接近 因此, 使两者充分接近的投资组合比例相当于如下二次规划问题:

$$\min_{d_i} \sum_{t=1}^T (x_{pt} - x_{0t})^2 \quad (9)$$

s t $\gamma_{pj} = \gamma_{0j} \quad j = 0, 1, \dots, q$

虽然APT从理论上给出了一个理想的求解投资组合比例的方法, 但在实用中, 最大的难题是怎样找出这些风险因子 常用的做法有3种: 分析外部经济因素对股票市场的影响, 然后找出影响股票市场的这些外部风险因子; 使用统计分析方法对价格回报率数据进行分析, 从而从统计意义上找出影响回报率变化的风险因子; 分析影响价格回报率的财务因素和股票市场因素, 用股票市场内部因素和公司财务因素分析股票市场各个回报率的相对变化 这3种做法各有优缺点, 单纯从对市场解释的意义上看, 第1种最差, 第2种较好, 第3种最好 由于第3种方法过于复杂, 在此采取第2种方法

把每个市场选择得到50支股票和A股指数收益率时间序列放在一起, 进行主成分分析 结果表明, 上海股票市场需要前30个主成分才能解释这50支股票或指数变动的95%, 而深圳股票市场前27支主成分的解释程度达到90%, 第一主成分可以解释的程度为49% 在上海股票市场, 用前30支主成分作为多因子模型的因子, 求解上面的二次规划问题, 求得各个股票的权重, 结果见表3 实际上, 50支股票中只有19支股票的权重不为零, 也就是只用到19支股票 在深圳股票市场, 使用前27支主成分作为多因子模型的因子, 使用同样的方法, 有26支股票的权重不为零

需要指出的是, 权重为0是由于在excel上求

解线性约束的二次优化造成的 由于 excel 求解优化的功能并不是很强, 可能在角点附近就把某些权重的取值设为 0 如果换一种计算语言(如 Matlab), 这些为 0 的权重可能不会为 0, 但会很小 由于这儿不追求计算精度, 就没有在其他计算语言上计算 这些权重为 0 的股票是无法从投资

群体中直接剔除的 因为可以看出, 投资组合的权重与距离大小无直接关系 实际上, 从金融理论上讲, 决定股票投资权重的往往是它们的相关性 如果一个股票与其他资产或其他资产的组合的相关性很强, 这个资产在组合中的权重可能很弱

表 3 各个股票在投资组合中所占的比例

code	weight								
R 6002	0.043 3	R 6080	0	R 6168	0	R 6678	0	R 6764	0
R 6054	0.008 06	R 6084	0	R 6170	0	R 6684	0	R 6773	0
R 6055	0.081 17	R 6087	0.084 89	R 6609	0	R 6698	0	R 6810	0
R 6059	0	R 6093	0	R 6619	0.026 02	R 6702	0	R 6820	0
R 6060	0	R 6104	0	R 6626	0	R 6713	0	R 6823	0.107 54
R 6062	0.047 65	R 6113	0	R 6638	0	R 6717	0	R 6854	0
R 6063	0	R 6116	0	R 6639	0	R 6726	0.113 8	R 6860	0.049 63
R 6066	0.055 32	R 6129	0.124 5	R 6649	0.017 59	R 6727	0	R 6866	0.005 75
R 6068	0	R 6162	0.003 48	R 6655	0.032 44	R 6736	0	R 6872	0.037 78
R 6074	0	R 6163	0	R 6663	0.056 29	R 6740	0.011 04	R 6890	0.095 67

注: code 指股票代码, weight 指股票权重

3 结果的分析讨论

首先看这些股票构成的指数投资组合 A 股指数变动是否一致 把由这些股票构成的投资组合的收益率序列与上证 A 股指数的变化率(收益率)序列在样本期内进行相关系数分析, 两者

的相关系数是 0.977 8, 而相应的深圳股票市场的投资组合和深证 A 指的相关系数为 0.975 027, 可见在两个市场指数组合对指数拟合效果很好 其实从操作性上看, 更关心两者在样本期过后的一段时间里两者能否足够相同, 把样本期和样本后 3 个月收益率直接进行比较, 上海市场的比较结果如图 1.

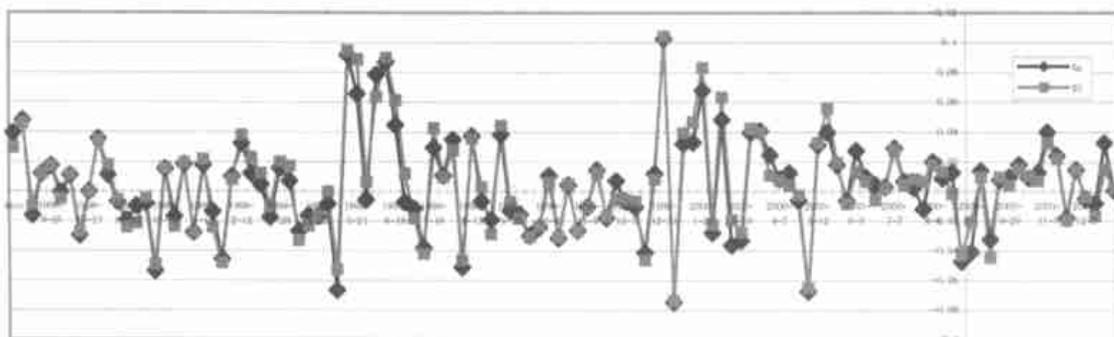


图 1 指数和指数证券组合变动的比较

再以上海股票市场为例分析一下组成指数组合的股票的行业特征和股本结构特征 可以发现, 这些股票的分布十分广泛, 不仅在几个主要行业, 即工业、商业、房地产业、公用事业和综合业存在, 而且在工业和综合业中也是多属于几个不同

的子行业 从它们的股本分布, 从各个股票的流通股大小来看, 属于大盘股(10 000 万股以上)的股票共有 8 支, 属于中盘股(5 000 至 10 000 万股)的股票有 7 支, 小盘股(5 000 万以下)则有 4 支, 应该说分布也是比较理想的 这说明这 19 支股票对

市场有很好的代表性

4 结 束 语

(1) 分散化投资在中国股票市场已经显示出很大优势,可以分散掉个股变化70%左右的风险;

(2) 两个市场综合指数和A股指数能够反映A股市场变化,其他指数相对较差

(3) 根据APT模型,使用聚类分析方法和主成分分析方法,选择大约20支股票和适当的权

重构造的投资组合就可以在短期内和A股指数变化一致

(4) 本文的分析方法也有一定的局限性 如在选择投资对象时,依据的是历史回报率的统计特性,因此从历史数据看,投资组合可以对指数有很好的逼近效果,但选择投资组合主要是为了满足将来的投资需要,投资组合在将来的表现好坏要看证券回报率的这些统计特性能否在很近的将来维持一段时间,实证虽然表明在不远的将来资产组合仍有很好的表现,但在更长的时间内指数证券组合的表现还有待于进一步检验

参 考 文 献

[1] Sharpe W. Capital asset prices: A theory of market equilibrium [J]. Journal of Finance, 1964, (4): 109- 113

[2] John L. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets [J]. Review of Economics and Statistics, 1965, (1): 56- 67

[3] Jan M. Equilibrium in a capital asset market [J]. Econometrica, 1966, (1): 118- 121

[4] Bodie, Kane, Marcus Investments[M]. third edition, Boston: Irwin McGraw-Hill, 1996

[5] Kariya T. Quantitative Method for Portfolio Analysis: A Method Approach [M]. Boston: Academic Press, 1993: 127- 136

[6] Connor G. The tree types of factor models: A comparison of their explanatory power [J]. Financial Analysts Journal, 1995, (3): 42- 46

[7] Connor G, Korajczyk R. Risk and return in an equilibrium APT: Application of a new test methodology [J]. Journal of Financial Economics, 1988, (2): 255- 289

[8] Fama E, Macbeth J. Risk, return and equilibrium: Empirical test [J]. Journal of Political Economy, 1973, (2): 607- 636

[9] Rosenberg B. Extra market components of covariance in security markets [J]. Journal of Financial and Quantitative Analysis, 1974, (1): 263- 274

[10] Fama E, French K. The cross-section of expected stock returns [J]. Journal of Finance, 1992, (2): 427- 465

[11] Hakansson GR. Higher return, lower risk: Historical returns on long-run actively managed portfolios of stocks, bonds, and bills [J]. Financial Analysts Journal, 1982, (2): 2- 16

[12] Grinold R, Kahn R. Active Portfolio Management [M]. New York: McGraw-Hill, 2000

[13] 朱宏泉, 卢祖帝, 汪寿阳. 中国股市的 Granger 因果关系分析 [J]. 管理科学学报, 2001, (5): 7- 12

[14] 陈怡玲, 宋逢明. 中国股市价格变动与交易量关系的实证研究 [J]. 管理科学学报, 2000, (2): 62- 68

Stock indexes and indexing portfolios in China stock markets

FAN Long-zhen, WANG Hai-tao, HE Hua

School of Management, Fudan University, Shanghai 200433, China

Abstract With the approach of principal component analysis, the main indexes are studied to determine if they can reflect the stock market change of the Shanghai Stock Exchange or the Shenzhen

Stock Exchange. The results indicate that the two composite indexes and two A-share indexes can accurately reflect the market changes, and other indexes are not so good in this function. Because all the four indexes are not good investment portfolios, we need to make indexing portfolios to track them. With the theory of APT and the approaches of statistics and optimization, we choose principal components as factors, make the factor loadings of tracking portfolios the same as the indexes, and then let the sum of squared differences of residuals minimized in the sample period to obtain the weight of each stock in the portfolio. About 20 stocks are selected from every market to obtain the tracking portfolio with specified weight. The indexing portfolios have almost the same returns as the indexes both in the sample period and after sample period (three months).

Key words: market indexes; indexing portfolio; cluster analysis; APT model

(上接第 5 页)

Study on common factors of vector stochastic volatility model

DU Zi-ping, ZHANG Shi-ying

School of Management, Tianjin University, Tianjin 300072, China

Abstract In this paper, from the idea of common trend, we put forward the definition of co-persistence in volatility of vector SV model, and establish the equivalent relationship between common persistence in volatility and cointegration. Meanwhile, we give two representations of common persistent factors under the Stock-Watson and Gonzalo-Granger factor decompositions, and verify the existence of cointegration between these two representations. Moreover, under the unobserved variables to be VAR(1), the representation of co-persistence factors is presented, and necessary conditions of existing co-persistence factors are also studied.

Key words: vector SV model; common persistence in volatility; common persistent factor; cointegration; long-run common trend