发布时间: 2018-07-18

证券研究报告 / 金融工程研究报告

港股研究系列: Barra 模型及应用

报告摘要:

随着陆股通的推进,内地投资者对港股的关注度也越来越高。在本报告中,我们对港股 Barra 模型进行了介绍,并对数据结果及应用进行展示。

数据是分析的基础。我们在本地搭建了港股基础信息、行情序列、 交易数据、恒生行业分类、股票估值数据、资产负债、利润和现金 流量等数据表。港股数据质量相对于 A 股存在一定不足,还需进一 步完善。

对于行业分类,我们使用的是恒生一级行业分类:公用事业、原材料业、地产建筑业、工业、消费品制造业、消费者服务业、电讯业、综合企业、能源业、资讯科技业和金融业;风格包括市值、贝塔、动量、波动率、非线性市值、BP、流动性、盈利、成长和杠杆等十个因子。数据清洗包括奇异数据的处理(去极值)、数据规范化和部分风格因子的因子载荷正交化以及缺失风险载荷的补足;同时参考Barra报告,在进行截面回归得到风格因子收益时,使用了加权最小二乘的回归方法。

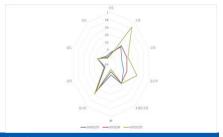
对于模型结果,首先对其可靠性进行分析。整体来看,样本内 R²均值为 40.63%,相对处于较高的水平,模型可信度较高。对于风格因子收益率,贝塔、动量和流动性因子累计收益为正值;波动率、市值和非线性市值这三个因子均为负值;盈利、成长和 BP 累计收益均为正,且 BP 最大;杠杆因子累计收益为负。这十个风格因子在港股市场的表现与 A 股存在一定差异,这也是后续研究的重要内容。

作为应用,我们对公募基金港股持仓风格和陆港通港股持仓风格进行了分析。公募基金持仓在市值、非线性市值、流动性、盈利、成长和杠杆上的暴露整体变动较小;在贝塔和波动率上的暴露增加较多,在动量上的暴露有所提高。对于陆港通持股,在市值上的暴露有所降低,而在贝塔和波动率上的暴露先升后降,在非线性市值上的暴露有所提高,其他风格暴露变动不大。

这一研究可为后续港股策略分析提供必要的工具基础,我们在后续报告中将会逐步完善港股相关研究内容。

模型 R² 走势 1000 1000 1000 1000 1000 1000

公募基金港股持仓风格



相关报告

《基于因子筛选的特质波动率研究》

2017-10-25

《因子优化: 动量因子再研究》

2017-12-04

《基本面分析因子研究-基于估值视角》

2018-03-04

《因子收益拆分与组合构建》

2018-06-08

证券分析师: 高建



目 录

1.	研究思路	3
2.	数据和方法	3
	1. 基础数据	
	3. 数据清洗	
	结果展示	
3.	1. 模型解释度	6
	2. 风格因子收益	
3.3	3. 组合分析	7
4.	总结	9
5. 1	附录	10



随着陆股通的推进,内地投资者对港股的关注度也越来越高,我们也开始对这一块 内容进行关注:港股数据库的搭建、港股量化策略的研究等。在本报告中,我们将 对港股 Barra 模型进行介绍,并对数据结果及应用进行展示。这一内容可为后续研 究提供工具性基础,带来较多便利。

1. 研究思路

Barra 模型的理论基础是多因子模型,即:

其中 r_i 为股票收益率; r_f 为无风险收益率; f_c 、 f_i 和 f_s 分别表示国家因子、行业因子和风格因子,I是由虚拟变量构成的行业风险载荷矩阵,S为风格因子的风险载荷。

利用多因子模型而非收益率的时间序列估计方差协方差矩阵的优势已为人所熟知。 其一,此种方法降低了估计协方差矩阵的所需要的时间跨度;其二,减小了矩阵的 不稳定性;其三,增强了方差矩阵的可预测性。而多因子模型通过选择解释能力强 且数量有限的各类因子对股票收益进行截面回归,并通过估计因子的方差-协方差矩 阵求得估计样本的协方差矩阵,并以此确定投资组合的绩效分解和边际风险归因, 即如以下公式所示。

$$R_{p} = \sum_{n} w_{n} r_{n} = \sum_{k} \psi_{k} f_{k} + \sum_{n} w_{n} u_{n}$$

$$Var(R_{p}) = \sum_{kl} \psi_{k} \psi_{l} F_{kl} + \sum_{n} w_{n}^{2} Var(u_{n})$$

其中, w, 为第 k 个因子对应的风险载荷与权重的线性组合。

在此基础上,通过计算预测方差协方差矩阵,即可对组合风险进行预测和管理。

在本报告中,我们会对港股 Barra 涉及的数据及因子计算方法进行说明,并给出模型数据结果;同时,我们也会依据港股组合持仓数据,对其风格进行分析。

2. 数据和方法

在这一部分,我们首先对涉及的数据进行说明,然后对因子计算方法进行分析。

2.1. 基础数据



我们在本地搭建了港股基础信息表、行情序列、交易数据、恒生行业分类、股票估值、资产负债、利润和现金流量等数据表,数据来源为 WIND; 因子构建及回归中涉及的数据均来源于这些数据表。相比于 A 股数据,港股数据质量相对差一些,后续还需要进一步核对完善。

2.2. 因子和数据处理方法

参考 Barra GEM3,涉及的十个风格因子定义、行业基准和其他数据说明如下(具体计算方法见参考报告):

表 1: 风格因子定义1

风格因子	计算方法及构成	因子含义	
Size	总市值自然对数	规模(大盘)	
Beta	CAPM 回归斜率 系统		
Momentum	超额收益(relative strength)		
Momentum	CAPM 回归截距项	一	
	日超额收益波动率(DASTD)	波动性	
Residual Volatility	累计超额收益(CMRA)		
_	CAPM 残差的标准差(HSIGMA)	_	
Non-Linear Size	Size 的立方剔除线性部分	规模(中盘)	
Book-to-Price	市净率倒数	价值	
	1月度平均换手率	流动性	
Liquidity	3个月平均换手率		
_	12 个月平均换手率		
C4h	过去5年每股盈利增长率	+ 12 14	
Growth -	过去5年每股营业收入增长率	— 成长性	
T	市现率(TTM)倒数	历山山	
Earning Yield -	市盈率(TTM)倒数	_ 盈利性	
	市场杠杆		
Leverage	负债/资产	_ 杠杆率	
_	账面杠杆	_	
行业	恒生一级行业分类	行业	
无风险收益率	Hibor-3M	无风险收益	

数据来源:东北证券,《The Barra Global Equity Model (GEM3)》

基于基础数据表,我们首先构建得到基础因子数据表(原始数据,未进行去极值和标准化处理)、风格因子数据表(标准化数据)和因子收益表(回归 R^2 、行业和风格因子收益及对应 T 值)。

2.3. 数据清洗

数据清洗包括奇异数据的处理(去极值)、数据规范化和部分风格因子的因子载荷正交化以及缺失风险载荷的补足。下面分别对其进行说明。

-

¹ 因子具体定义见附录。



2.3.1. 极值与规范化处理

首先,对数据进行如下标准化:

$$X_{nk}^{(std)} = \frac{X_{nk}^{(raw)} - \mu_k}{\sigma_k}$$

其中 μ_{L} 为市值加权平均, σ_{L} 为简单平均标准差。

与传统的极值限制在3个标准差的做法不同,我们采取以下极值处理方式,将风险暴露限制在[-3.5,3.5]。

$$\tilde{X}_{nk}^{(std)} = \begin{cases} 3 \cdot (1 - s_{(+)}) + X_{nk}^{(std)} \cdot s_{(+)} ; & X_{nk}^{(std)} > 3 \\ X_{nk}^{(std)} & ; -3 \le X_{nk}^{(std)} \le 3 \\ -3 \cdot (1 - s_{(-)}) + X_{nk}^{(std)} \cdot s_{(-)} ; & X_{nk}^{(std)} < -3 \end{cases}$$

其中,

$$s_{(+)} = Max \left[0, Min(1, \frac{0.5}{Max(X_{nk}^{(std)} - 3)}) \right]$$

可以检验, 此种极值处理方式不仅限制了数据的取值, 同时保留了原始数据的排序。

极值处理后,需要重新对数据进行标准化;此外,若是描述变量,还需对组合为风险暴露后的数据进行标准化处理。

2.3.2. 缺失值处理

当构成风格因子的描述变量数据只有部分缺失时,我们利用其余未缺失的因子指标 来构成风格因子(权重进行归一化处理);当某只股票因子指标全部缺失时,我们 运用数据替换算法来补足缺失值。具体做法为:对其他未缺失数据的部分因子暴露 值进行回归,并得到回归系数后,从而近似计算出风格因子的暴露值。

2.3.3. 截面回归

在得到风格因子数据后,通过截面回归的方法,即可得到风格因子收益值和其他统计量结果。

通常,大盘股收益相较于小盘股收益具有更多的共性,此外在进行分析时理应对市值占比较高的股票给予较多的关注。参考 Barra 报告,我们在回归模型中,考虑市值的影响,以降低异方差的影响。这里我们以总市值平方根为权重进行加权最小二乘回归,得到风格因子收益数据。

3. 结果展示²

在这一部分我们将分别对模型可靠性结果 (回归 R^2)、风格因子收益进行说明;在

²我们计算了 2008 至今的日度基础因子值、风格因子值和因子收益值。

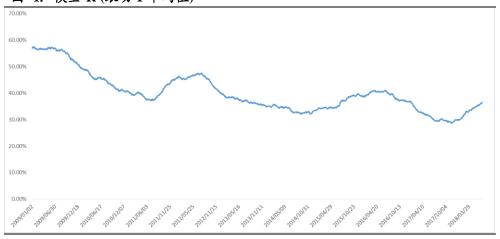


此基础上,我们以基金港股持仓为样本,对其持股风格进行分析;同时,对陆港通 持股中港股作为样本,分析其持仓风格变动。

3.1. 模型解释度

首先给出可决系数 R²的滚动均值结果(过去1年),见图 1。整体来看,样本内 R2 均值为 40.63%,相对处于较高的水平,模型可信度较高;随时间推移初步呈下降趋势,这也在一定程度上反映出因子有效性有所降低。

图 1: 模型 R²(滚动 1 年均值)



数据来源: 东北证券, Wind

3.2. 风格因子收益

这一部分, 我们给出 10 个风格因子的累计收益走势(2008-2018年), 见图 2 和图 3。

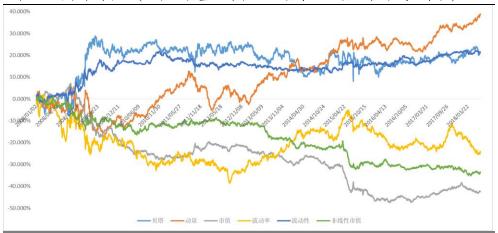
图 2 中给出了贝塔、动量、市值、波动率、流动性和非线性市值等 6 个因子的累计收益。其中,贝塔、动量和流动性因子累计收益为正值;同期,A 股这三个因子的累计收益分别为正值、负值和负值,且贝塔累计收益绝对值大小较港股大很多。而港股波动率、市值和非线性市值这三个因子均为负值,从方向上来看与 A 股一致。

图 3 中给出了盈利、成长、BP 和杠杆等四个基本面指标的累计收益走势。其中,盈利、成长和 BP 累计收益均为正,且 BP 最大; A 股盈利和 BP 因子累计收益为正值,而港股成长因子相对表现更好。杠杆因子累计收益为负,这一结果与 A 股市场一致。

从上述结果看出,这十个风格因子在港股市场的表现与 A 股存在一定差异,这也是后续研究的重要内容。

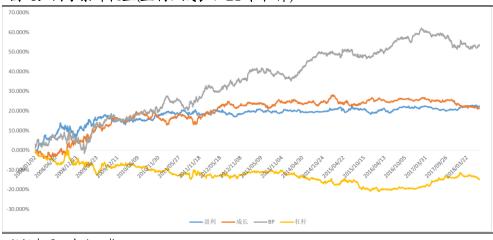


图 2: 因子累计收益(贝塔、动量、市值、波动率、流动性和非线性市值)



数据来源:东北证券, Wind

图 3: 因子累计收益(盈利、成长、BP和杠杆)



数据来源:东北证券, Wind

3.3. 组合分析

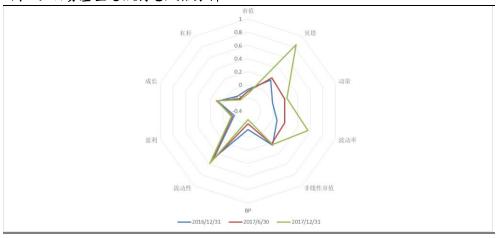
在这一部分基于港股 Barra 模型,对公募基金港股持仓、陆港通港股持仓的风格及行业分布进行分析(2016 四季度、2017 二季度和 2017 四季度),结果见图 4、图 5、图 6 和图 7。

对于持仓风格: 公募基金持仓在市值、非线性市值、流动性、盈利、成长和杠杆上的暴露整体变动较小; 在贝塔和波动率上的暴露增加较多, 在动量上的暴露有所提高。对于陆港通持股, 在市值上的暴露有所降低, 而在贝塔和波动率上的暴露先升后降, 在非线性市值上的暴露有所提高, 其他风格暴露变动不大。

对于行业:公募基金持仓行业占比最高的为金融业、消费品制造业、资讯科技业和公用事业;从时间上来看,消费品制造业、消费者服务业、资讯科技业和金融业占比均逐渐提升。陆港通持仓行业占比最高的为金融业、消费品制造业、地产建筑业和资讯科技业;从时间上来看,各行业占比变动相对较小,消费品制造业占比有所提高,而金融业则有所降低。

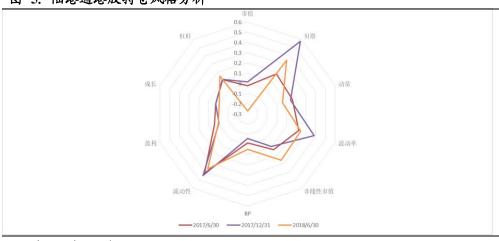


图 4: 公募基金港股持仓风格分析



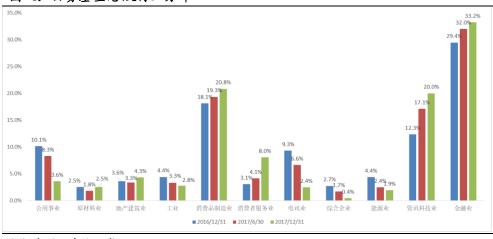
数据来源:东北证券, Wind

图 5: 陆港通港股持仓风格分析



数据来源:东北证券, Wind

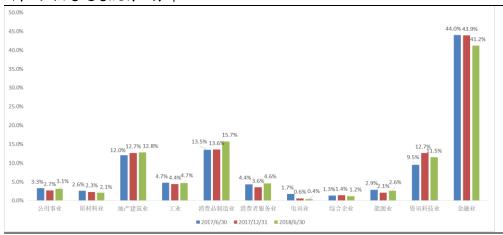
图 6: 公募基金港股行业分布



数据来源:东北证券, Wind



图 7: 陆港通港股行业分布



数据来源:东北证券, Wind

4. 总结

随着陆股通的推进,内地投资者对港股的关注度也越来越高。在本报告中,我们对港股 Barra 模型进行了介绍,并对数据结果及应用进行展示。

我们首先对 Barra 模型的含义和形式进行了说明: 其是截面上的多因子模型,将国家、行业和风格因子纳入模型中,以对组合收益和风险进行归因和预测。

数据是分析的基础。我们在本地搭建了港股基础信息、行情序列、交易数据、恒生 行业分类、股票估值数据、资产负债、利润和现金流量等数据表。港股数据质量相 对于 A 股存在一定不足,还需进一步完善。

对于行业分类,我们使用的是恒生一级行业分类:公用事业、原材料业、地产建筑业、工业、消费品制造业、消费者服务业、电讯业、综合企业、能源业、资讯科技业和金融业;风格包括市值、贝塔、动量、波动率、非线性市值、BP、流动性、盈利、成长和杠杆等十个因子。数据清洗包括奇异数据的处理(去极值)、数据规范化和部分风格因子的因子载荷正交化以及缺失风险载荷的补足;同时参考 Barra 报告,在进行截面回归得到风格因子收益时,使用了加权最小二乘的回归方法。

对于模型结果,首先对其可靠性进行分析。整体来看,样本内 R² 均值为 40.63%,相对处于较高的水平,模型可信度较高。对于风格因子收益率,贝塔、动量和流动性因子累计收益为正值;波动率、市值和非线性市值这三个因子均为负值;盈利、成长和 BP 累计收益均为正,且 BP 最大;杠杆因子累计收益为负。这十个风格因子在港股市场的表现与 A 股存在一定差异,这也是后续研究的重要内容。

作为应用,我们对公募基金港股持仓风格和陆港通港股持仓风格进行了分析。公募基金持仓在市值、非线性市值、流动性、盈利、成长和杠杆上的暴露整体变动较小;在贝塔和波动率上的暴露增加较多,在动量上的暴露有所提高。对于陆港通持股,在市值上的暴露有所降低,而在贝塔和波动率上的暴露先升后降,在非线性市值上的暴露有所提高,其他风格暴露变动不大。



这一研究可为后续港股策略分析提供必要的工具基础,我们在后续报告中将会逐步完善港股相关研究内容。

5. 附录

在这一部分,我们对十个风格因子的计算方法进行说明。

1. SIZE

取个股总市值自然对数,作为因子值。

2. Beta

$$r_{t} - r_{ft} = \alpha + \beta R_{t} + e_{t}$$

 R_t 为全市场个股超额收益总市值加权结果,作为市场指数;使用过去 252 个交易日进行回归计算;回归之前对 r_t-r_{ft} 和 R_t 进行指数加权,半衰期为 63 个交易日,即 $x_t=R_t(0.5^{1/63})^{252-t}\,,\;\;y_t=(r_t-r_{ft})(0.5^{1/63})^{252-t}\,.$

3. Momentum

这一指标由 RSTR 和 HALPHA 进行加权得到。

使用过去 504 个交易日收益率数据, 计算得到 RS 指标:

$$RS = \sum_{t=1}^{504} w_{t} (\ln(1+r_{t}) - \ln(1+r_{ft}))$$

其中, $w_{t} = (0.5^{1/126})^{504-t}$; r_{t} 为个股收益率; r_{ft} 为无风险收益率。

对 T-20 至 T-10, 共 11 个交易日的 RS 值进行简单平均, 得到 RSTR 指标:

$$RSTR = \frac{1}{11} \sum_{\tau=11}^{21} RS(\tau)$$

对 T-20 至 T-10, 共 11 个交易日的截距项值(回归得到 Beta 时的截距项)进行简单平均,得到 HALPHA 指标。

4. Residual Volatility

这一指标由 DASTD、CMRA 和 HSIGMA 三个指标加权得到; 风格因子需要对 Beta 和 Size 进行回归正交处理,以剔除共线性的影响。

DASTD:使用过去 252 个交易日半衰期加权的个股超额收益率, 计算得到标准差; 半衰期为 42 个交易日;

CMRA:首先计算每月个股和无风险复利收益率 r_{τ} 和 $r_{f\tau}$;然后依次计算过去 1 个月,



2 个月, ..., 12 个月的累计超额对数收益率
$$Z(T) = \sum_{\tau=1}^{T} \left[\ln(1+r_{\tau}) - \ln(1+r_{f\tau}) \right]$$
 ;

最后计算
$$CMRA = \max\{Z(T)\}_{T=1}^{12} - \min\{Z(T)\}_{T=1}^{12}$$
 ;

HSIGMA:利用计算 Beta 回归时的残差收益率序列,计算其标准差。

5. Non-Linear Size

$$Size^3 = bSize + e$$

将标准化 Size 的立方对 Size 做加权回归,得到参数估计值 b; 然后取 $Size^3 - bSize$ 作为非线性市值因子。

6. Book-to-Price

取市净率倒数 (最近报告期净资产值)。

7. Liquidity

该指标由 STOM、STOQ 和 STOA 三个指标加权得到; 风格因子需对 Size 进行回归 正交处理。

根据个股换手率进行计算: V, 成交量; S, 总股本。

STOM:利用过去 21 个交易日数据进行计算:
$$STOM = \ln(\sum_{t=1}^{21} \frac{V_t}{S_t})$$

STOQ:利用过去 63 个交易日数据进行计算:
$$STOQ = \ln(\frac{1}{3}\sum_{r=1}^{3} \exp(STOM_r))$$

STOA:利用过去 252 个交易日数据进行计算:
$$STOA = \ln\left(\frac{1}{12}\sum_{r=1}^{12} \exp(STOM_r)\right)$$

8. Earnings Yield

该指标由 CETOP 和 ETOP 加权得到。

CETOP:经营性现金净流入 TTM 除以当前市值;

ETOP:净利润 TTM 除以当前市值。

9. Growth

该指标由 EGRO 和 SGRO 加权得到。

EGRO:使用过去 5 年每股盈利数据,对时间序号 (1-5)进行回归;取回归斜率值除以每股盈利均值作为因子值;

SGRO: 使用过去 5 年每股营业收入数据,对时间序号 (1-5)进行回归;取回归斜率值除以每股营业收入均值作为因子值。

10. Leverage

该指标由 MELV、DTOA 和 BLEV 指标加权得到。



MLEV: $_{MLEV} = \frac{ME + PE + LD}{ME}$, 其中 ME 为股票总市值, PE 为优先股, LD 为公

司长期负债;

DTOA: $DTOA = \frac{TD}{TA}$, 其中 TD 为总负债, TA 为总资产;

BLEV: $_{BLEV} = \frac{BE + PE + LD}{BE}$, BE 为所有者权益,PE 为优先股,LD 为公司长

期负债。



分析师简介:

徐忠亚:金融工程研究助理,南京大学金融学硕士,2017年加入东北证券研究所。

重要声明

本报告由东北证券股份有限公司(以下称"本公司")制作并仅向本公司客户发布,本公司不会因任何机构或个人接收到本报告而视其为本公司的当然客户。

本公司具有中国证监会核准的证券投资咨询业务资格。

本报告中的信息均来源于公开资料,本公司对这些信息的准确性和完整性不作任何保证。报告中的内容和意见仅反映本公司于发布本报告当日的判断,不保证所包含的内容和意见不发生变化。

本报告仅供参考,并不构成对所述证券买卖的出价或征价。在任何情况下,本报告中的信息或所表述的意见均不构成对任何人的证券买卖建议。本公司及其雇员不承诺投资者一定获利,不与投资者分享投资收益,在任何情况下,我公司及其雇员对任何人使用本报告及其内容所引发的任何直接或间接损失概不负责。

本公司或其关联机构可能会持有本报告中涉及到的公司所发行的证券头寸并进行交易,并在法律许可的情况下不进行披露; 可能为这些公司提供或争取提供投资银行业务、财务顾问等相关服务。

本报告版权归本公司所有。未经本公司书面许可,任何机构和个人不得以任何形式翻版、复制、发表或引用。如征得本公司同意进行引用、刊发的,须在本公司允许的范围内使用,并注明本报告的发布人和发布日期,提示使用本报告的风险。

本报告及相关服务属于中风险(R3)等级金融产品及服务,包括但不限于A股股票、B股股票、股票型或混合型公募基金、AA级别信用债或ABS、创新层挂牌公司股票、股票期权备兑开仓业务、股票期权保护性认沽开仓业务、银行非保本型理财产品及相关服务。

若本公司客户(以下称"该客户")向第三方发送本报告,则由该客户独自为此发送行为负责。提醒通过此途径获得本报告的投资者注意,本公司不对通过此种途径获得本报告所引起的任何损失承担任何责任。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格,并在中国证券业协会注册登记为证券分析师。本报告遵循合规、客观、专业、审慎的制作原则,所采用数据、资料的来源合法合规,文字阐述反映了作者的真实观点,报告结论未受任何第三方的授意或影响,特此声明。

投资评级说明

#EX 1	**	
	买入	未来6个月内,股价涨幅超越市场基准15%以上。
股票	增持	未来6个月内,股价涨幅超越市场基准5%至15%之间。
投资 评级	中性	未来6个月内,股价涨幅介于市场基准-5%至5%之间。
说明	减持	在未来6个月内,股价涨幅落后市场基准5%至15%之间。
	卖出	未来6个月内,股价涨幅落后市场基准15%以上。
行业	优于大势	未来6个月内,行业指数的收益超越市场平均收益。
投资 评级	同步大势	未来6个月内,行业指数的收益与市场平均收益持平。
说明	落后大势	未来6个月内,行业指数的收益落后于市场平均收益。



东北证券股份有限公司

网址: http://www.nesc.cn 电话: 400-600-0686

地址	邮编
中国吉林省长春市生态大街 6666 号	130119
中国北京市西城区锦什坊街 28 号恒奥中心 D 座	100033
中国上海市浦东新区杨高南路 729 号	200127
中国深圳市南山区大冲商务中心 1 栋 2 号楼 24D	518000

机构销售联系方式

	· ·						
姓名	办公电话	手机	邮箱				
华东地区机构销售							
袁颖 (总监)	021-20361100	13621693507	y uany ing@nesc.cn				
王博	021-20361111	13761500624	wangbo@nesc.cn				
李寅	021-20361229	15221688595	liyin@nesc.cn				
杨涛	021-20361106	18601722659	yangtao@nesc.cn				
阮敏	021-20361121	13564972909	ruanmin@nesc.cn				
李喆莹	021-20361101	13641900351	lizy@nesc.cn				
齐健	021-20361258	18221628116	qijian@nesc.cn				
陈希豪	021-20361267	13956071185	chen_xh@nesc.cn				
华北地区机构销售							
李航(总监)	010-58034553	18515018255	lihan g@nesc. cn				
殷璐璐	010-58034557	18501954588	yinlulu@nesc.cn				
温中朝	010-58034555	13701194494	wenzc@nesc.cn				
曾彦戈	010-58034563	18501944669	zengy g@nesc.cn				
颜玮	010-58034565	18601018177	yanwei@nesc.cn				
华南地区机构销售							
邱晓星 (总监)	0755-33975865	18664579712	qiuxx@nesc.cn				
刘璇	0755-33975865	18938029743	liu_xuan@nesc.cn				
刘曼	0755-33975865	15989508876	liuman@nesc.cn				
林钰乔	0755-33975865	13662669201	linyq@nesc.cn				
周逸群	0755-33975865	18682251183	zhouyq@nesc.cn				