**基于SSA和随机森林的股指分析和预测的新方法**

摘要：股指，尽管受到各种各样因素的影响，但在不同的时间维度上会有特定的特征，对股指特征的分析可以帮助我们更好地了解股指的本质规律并且提高预测的准确性。这篇文章，利用奇异谱分析（SSA）的方法，分解股指为趋势项，市场波动项和噪音项来在不同的时间维度上研究股指的特征，最后引入这些特征到随机森林中训练后进行股指预测。测试结果表明随机森林算法回归预测具有良好的表现，最后的预测效果表明通过SSA分解出原时间序列的特征序列再导入随机森林后可以对股指涨跌趋势作出较为准确的指导。

关键词：股指序列 奇异谱分析 随机森林 联合预测方法

1. **引言**

精确的股指预测有利于投资者有效地防范风险，管理者制定合理的经营战略和政府更好

地了解经济趋势。股指预测因此总是金融研究的热点话题。然而，这也是在金融研究中普遍被认为是最有挑战性的研究话题，因为它受大量经济的和非经济的因素影响。

在近年来，人工神经网络（ANN）已经成功地被用于金融时间序列的建模（Cheng et al. (1996)1， Sharda 和 Patil (1994)2，David和Suraphan (2005)3）。然而，因为研究股市数据是一个多因素分析问题，其中每一个因素可以看作一个维度的指标，这样股市数据就由多个维度的指标体系决定，其中还有大量的噪音，某些研究表明ANN有一些模式学习的局限性。最近，一种叫支持向量机（SVM）的新算法已经被用作金融时间序列的预测（Mukherjee etal.(1997)4. Tay 和 Cao (2001)5. Kim(2003)6.）另一种由Breiman（2001）7提出的叫随机森林的方法理论则已经被发现在许多研究中表现得更好，例如Creamer 和 Freund（2004）8，Lariviere和Poel (2004)9。这些机器学习方法例如支持向量机和随机森林成功地实现了在给定其它变量的情况下对某些特定变量进行精确的预测，同时还避免了对一些潜在分布建模时的问题，如神经网络模型可能会倾向于落入局部最优解，而SVM和随机森林的解往往可能是全局最优的。因此，随机森林在回归时往往还可以避免过拟合问题。尽管这些模型已经被科研工作实验过，但目前还没有人尝试过用随机森林的方法来对股市的时间序列预测进行实验。股市同时受经济，政治，金融和社会的因素，噪音和投资者行为10影响，股票价格可能在不同的时间范围内有不同的特征（例如分形和混沌的特征11）。但是很少研究引入这些价格特征到随机森林中来对股指进行预测。

SSA是一种最初由Broomhead和King(1986)12提出的分析非线性，非稳态的时间序列的方法。把时间序列的经典分析，包括多元几何学，多元统计学，动态系统和许多其它的理论组合在一起。它根据所观察到的时间序列构造出轨迹矩阵，并对自己矩阵进行分解、重构，从而提取出代表原时间序列不同特征的时间序列，如长期趋势项、周期项、噪声项等，从而对时间序列的结构进行分析。奇异谱分析目前较多地应用在气象学和地质学的研究中，并有少量学者将其引入到社会问题研究中。Hassani（2007）13使用SSA方法来从在美国意外死亡的时间序列提取信息并且获得趋势项，谐波项和噪音项；Lian JiJIan，et al (2008) 14使用基于奇异熵的阶数确定法来达到去除原始序列中的噪音的目的。由于这种方法不需要对模型进行假设，是一种非参数估计法并且不需要各种经济指标数据，不存在模型设计和过度拟合等问题，所以同样适合于金融时间序列的分析。Beneki，et al (2012) 15成功分解出趋势项和经济波动项，并且利用它们进行英国的旅游收入序列分析；Zhang Yi，et al(2012)16使用SSA来获得噪音项和趋势项并且它们进行汇率分析；Beneki et al（2013）17利用SSA对非参数时间序列分析的优势，分析和预测了欧盟的能源数据。Huang et al（2014）18通过HAR-CJ-M模型来预测中国股市波动。在这篇论文中，我们将使用SSA来分解深圳综指，提取它的特征来进行分析。

随着SVM和随机森林回归的引入，这些技术已经成功地被用于解决估计问题，并且这些算法均具有良好的表现。这表明随机森林能够在考虑股市各种复杂的特征的情况下，提高深证综指的涨跌方向预测的准确性。为了使预测更加准确，我们用SSA来分解股指序列为趋势项，股市波动项和噪音项，并且在不同的时间维度上研究股市价格特征，然后引入这些特征到随机森林来进行趋势预测。这篇文章分为4个部分：引言，预测方法的介绍，实证研究和结论。

1. **预测方法的介绍**
   1. **SSA**

SSA的核心思想是通过奇异值分解（SVD）来获得一系列包含原始序列信息的奇异值，然

后选择不同的奇异值来构造不同成分的序列。具体的步骤如下：

* + 1. **时间序列的相位空间重构**

给定一个时间序列， 是有效长度。重构序列的相位空间，然后我们可以推导出轨迹矩阵：



其中  是窗口长度， 是嵌入维数， 并且 ，，。

* + 1. **奇异值分解**

SSA是基于矩阵代数中一种叫奇异值分解的特定的变换。对 进行奇异值分解，，，其中。的集体被称之为SVD的第个三重特征向量。向量是矩阵的左边的奇异向量，数是奇异值，并且为矩阵提供奇异谱,由此把这种方法叫做奇异谱分析。向量被称之为主成分的向量。记，如果时间序列只包含有效信息，则矩阵的秩是; 如果时间序列同时包含有效信息和噪音，则。

* + 1. **基于奇异熵的去噪的阶数的确定**

奇异熵概念19的引入是为了研究时间序列中信息数量随着奇异谱阶数改变的规律：



其中，  是奇异熵的阶数，  代表奇异熵在阶数的增量， 通过下面的公式我们可以获得：



当奇异熵的增量接近渐近值时，时间序列的有效信息饱和，意味着信息已经完全包含，之后的奇异熵增量都是由噪音造成的，并且此时第阶的数字可以被认为是噪音之一而去除。

* + 1. **去噪后的时间序列重构**

去噪后，时间序列的重构可以被分成下面两个过程：

1. 我们可以由上述分析得到数字  从而确定去噪的阶数，令所有的阶数大于 的奇异值为0 来获得一个新的奇异值矩阵 ， 并且然后用公式 来推导出一个无噪音的轨迹矩阵。
2. 令重构的时间序列是 ， 并且  是 的一个元素， 我们有下面的公式 (这个公式代表矩阵副对角线上元素均值)：



根据上面的公式，我们可以获得一个去噪后的时间序列或者有效的信息序列。

* 1. **随机森林**

Bagging（Breiman (1996a)）20是并行使集成学习方法最著名的代表，它直接基于自主采样法（Efron and Tibshirani, 1993）21。给定包含个样本的数据集，我们先随机取出一个样本放入采样集中，再把该样本放回初始数据集，使得下次采样时该样本仍有可能被选中，这样，经过次随机采样操作，我们得到含个样本的采样集，初始数据集中有的样本在采样集里多次出现，有的则从未出现。照这样，我们可采样出个含个训练样本的采样集，然后基于每个采样集训练出一个基学习器，再将这些基学习器进行结合，这就是Bagging的基本流程。

Breiman（2001）7提出的随机森林是Bagging的一个扩展变体。随机森林在以决策树为基学习器构建Bagging集成的基础上，进一步在决策树的训练过程中引入了随机属性选择。具体来说，在随机森林中，一个随机的向量 被生成，独立同分布于原来的随机向量 ；此时有一棵树被用训练集和生成，形成一个分类器 ，其中 是一个输入向量。 随机选择组成大量的独立的从到的随机整数。的性质和维度取决于它在树的构成中的作用。在大量树被生成以后，它们投票选择出最适合的分类。这个过程被称作随机森林。

一个随机森林是一个由多个树型结构基分类器构成的分类器的集成。其中，  是独立同分布的随机向量，并且每棵树为输入的最适合的分类的投出一票。

给定一个分类器的集成 并且随着训练集从随机向量的分布被随机生成，定义边际函数为



其中  是指示函数。这个边际衡量了在 中，正确分类的平均选择数量超出其它分类的程度。这个边际越大，这个分类的置信度越高。总误差由下面的公式给出



其中表明在 空间上的可能性。

在随机森林中， 。 对于大量的树，它服从强大数定律，并且随着树的数量的增加，几乎可以确定出所有的序列 ，…

收敛于



由上面等式的收敛，我们可以知道随机森林不会随着树的增加而过拟合，但会产生一个总误差的极限值。

传统决策树在划分属性选择时是在当前节点的属性集合（假定有个属性）中选择一个最优属性；而在随机森林中，对基决策树的每个节点，先从该节点的属性集合中随机选择一个包含个属性的子集，然后再从这个子集中选择一个最优属性用于划分。这里的参数控制了随机性的引入程度：若令，则基决策树的构建与传统决策树相同；若令，则是随机选择一个属性用于划分；一般情况下，推荐值（Breiman（2001）7）。

这个策略被采用来达到下面的结果：

1. 在生成的树达到最大的深度的情况下，保证个体误差值足够低。
2. 通过下面两点保证足够低的随机残差相关性：
3. 通过来自训练的数据的样本生成每棵树。
4. 规定  (属性的数量)。在每棵树选择的节点上，选择  个属性并且选择出基于这些属性生成的节点的最佳分割。

随机森林中基学习器的多样性不仅来自样本扰动，还来自属性扰动，这就使得最终集成的泛化性能可通过个体学习器之间差异度的增加而进一步提升。随机森林的起始性能往往相对较差，但随着个体学习器数目的增加，随机森林通常会收敛到更低的泛化误差，而且训练效率常优于Bagging。

总之，对于二进制的输出， 随机森林构造了一个分类树的集成。每棵树是由在每一个分割的导入的样本数据所建立的；用一个随机样本的预测来检验。最终分类的确定是由分类树集成基于每个情况下的最多投票数来确定的。

**3. 实证研究**

**3.1 调查步骤**

我们参考Zhang, et al22对石油价格分析的研究来在三个不同的时间维度（长期，中期，短期）上研究股指的不同影响因素。在长期的时间维度上，我们集中考虑股指的趋势；在中期的时间维度上，我们关注各种与股指相关的经济的或非经济的事件的影响；在短期的时间维度上，我们考虑股市上的噪音。SSA的方法有它在股指分解上的优势:

1. SSA 可以被用于分析非线性的或非静态的时间序列；
2. 用SSA分解出来的股指长期趋势的序列并不会被单调性所限制；
3. SSA可以分解出噪音序列，这对于普遍存在于股市的因素——噪音的分析是有益的。

因此，这篇文章利用SSA来导出包含原始序列不同信息的奇异值。第一个奇异值包含原始序列最基本的信息，而且能够被用于构造用来研究股指长期趋势的趋势项。奇异熵去噪法被用于导出包含噪音的原始序列的奇异值，这被用来构造出研究股市噪音影响的噪音项。然后剩下的奇异值就被用来构造研究不同事件影响的股市波动项。有了对趋势项，市场波动项和噪音项的分析后，原始序列的本质规律就能够被懂得。最终的预测可以在随机森林的帮助下完成。具体的步骤如下：

1. SSA分解：首先选择一个合适的数字作为嵌入维来为原始序列构造一个相位空间来推导出轨迹矩阵，然后用SVD来分解轨迹矩阵来获得代表原始时间序列不同特征的奇异值；根据奇异熵增量的接近渐近值的程度来确定去噪阶数；最后提取出奇异值来重构趋势项，市场波动项和噪音项的时间序列。
2. 随机森林训练：我们通过对SSA分解出来的趋势项，市场波动项和噪音项重构时间序列作为对应时间段内的深圳综指收盘价的特征向量引入随机森林中进行训练，即每个训练样本是一个四维的向量。我们通过随机森林算法随机选择初始数据集的作为训练样本集。
3. 随机森林测试：经过（2）的训练后，我们利用初始数据集剩下的作为测试集来测试随机森林算法拟合的效果，通过比对通过随机森林算法得到的预测值和真实值的图像，可以评估随机森林算法的优良性。
4. 随机森林预测：最后，我们将选择另外时间段内的同种股指作为新样本，利用SSA得到新样本的趋势项，市场波动项和噪音项后，它们是代表新股指特征的时间序列。把它们引入到相同的经过训练和测试后的随机森林算法中进行预测，由此可以评估在金融环境变化的情况下，随机森林算法经过SSA分解后利用特征序列对原始序列随时间变化趋势的预测效果。

上述步骤呈现如图1的流程图：

原始序列

奇异谱分析

趋势信号

市场波动信号

噪声信号

随机森林训练

随机森林测试

随机森林预测

最终预测值

图表 1实证研究过程的流程图

**3.2 实证研究过程**

我们选择从2010年1月4日到2014年12月31日的深圳综指的收盘价作为研究对象，一共是1212个数据（数据来源是Wind资讯金融终端）。令原始序列为 ，然后用 来进行相位空间构建。在相位空间构建中，嵌入维的选择是非常重要的，因为它需要足够大，但不能超出原始序列长度的一半由此我们可以从原始序列中获得几乎所有的信息。特别地，如果我们知道原始序列的周期维数，嵌入维数的数量应该是周期维数的一个公倍数；如果存在不同的周期维数，则嵌入维数应该是所有周期维数的公倍数。实际操作中，三分之一或者四分之一长度的原始序列会被选择来从原始序列中获得所有的信息。这篇文章选择原始序列约55%的长度（668）作为嵌入维数，选择约44%的长度（545）作为窗口长度。通过分解轨迹矩阵来获得奇异值矩阵，然后构造出这些奇异值的奇异熵增量的图像，如图2所示：

图表 2 奇异熵增量

由图2我们可以看到，当阶数达到50时，奇异熵的增量的改变变得很少，有效信息饱和。重构阶数超过50的奇异值作为噪音项；由图还可以推得第一个奇异值包含序列最基本的信息，用它来重构出趋势项 ；剩下的阶数为2-50的奇异值就作为市场波动项 。然后可以得到一个和原始序列对比的图象如图3：

图表 3 原始序列和分解项

**3.2.1 分解项的经济含义分析**

利用统计学软件，可以得到趋势项和原始序列之间的皮尔逊相关系数为0.473，二者具有显著的中等正相关性，这表明趋势项对股指的长期趋势有决定性作用。长期趋势与一国的经济环境是密不可分的。经济增长在某种程度上决定了趋势项的改变，相应地，趋势项的改变也可以反映经济发展中的某些问题。例如，自从2008年的金融危机以后，全球股市大挫，包括国内股市和深圳综指。在这次危机以后，政府推行了大量刺激经济的计划。然后，在2009年，国家的经济开始反弹，股市指数开始上升。然而，在图像中，趋势项表明深圳综指在2010年再次开始下降。再例如中国GDP从2011年的9.2%直接大幅下降到2012年的7.8%。同时在图像中，趋势项表明深圳综指在2011年开始下降，即使期间略有小幅波动。这些都表明许多问题仍然存在于中国的经济增长中——太热的经济，投资上过多的依赖性，并由此导致的经济中过多的泡沫，股民对于投资的不自信，回归真实增长和股指上升到一定程度刺破泡沫的必要性。

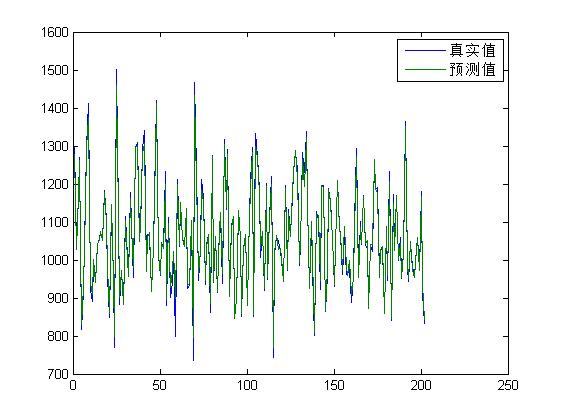
利用统计学软件，可以得到市场波动项和原始序列之间的皮尔逊相关系数为0.859，二者具有显著的强正相关性，这恰恰反映了中国股市体系的不完善。股市波动项主要是用来反映中期内对股指产生大型冲击的经济的和非经济的事件对于股指的影响。股市波动项的曲线表明从2010-2013年股市波动较为平稳，2013-2014年则波动相对剧烈，这表明一些事件对股指产生的冲击往往会持续几个星期到几个月不等，而重大事件的影响会持续更长时间。另外，在某些时间段内，数据会波动剧烈，表明这些事件对股指产生了重大的的影响。在长期趋势项改变缓慢和噪音项波动平稳的情况下，中期股指剧烈波动通常是由于各种各样的事件导致的。这就是深圳综指收盘价原始序列和它分解出来的市场波动项在图像上变化相似的原因。从原始序列中分解出这些事件对于股指预测是有益的。因为，当类似的事件发生的时候，例如频繁的经济事件或者某些特定的非经济事件，我们可以利用过去的相关研究来预测股指的增长或者下跌。

利用统计学软件，可以得到噪音项和原始序列之间的皮尔逊相关系数为0.098，二者具有显著的弱正相关性。噪音项主要反映了股市在股指上全体噪音的影响。中国的股市只有非常短的历史，股市规则和规章制度都没有完善。许多市场参与者缺乏金融经验，甚至存在羊群效应。更糟糕的是，中国的股市缺乏透明度，投资者通常没有足够的信息用于投资决策。由噪音项呈现在图像中的曲线可以看到，噪音项在长期和中期内是平稳的，但是在短期内会频繁地波动，不过，这些波动对深圳综指长期趋势的影响相对于趋势项，和股市波动项来说是微不足道的。这些噪音项的波动常常是由于短期内某些特殊事件发生引起的。因此，尽管这对于长期和中期的股指预测来说是可以忽略的，但对于短期的股指预测来说则是非常重要的。

**3.2.1 随机森林预测**

通过奇异谱分析（SSA）我们已经对深圳综指的原始序列进行了分解，得到了代表原始序列特征的趋势项，市场波动项和噪音项。接下来，我们将把这些特征序列导入到随机森林中，通过随机森林算法对这些数据进行训练和测试。最后，在考虑金融环境变化的情况下，我们会利用另一个时间段内的深圳综指同样通过SSA得到三个相关的特征时间序列后导入随机森林中检测随机森林对股指趋势预测的准确性。

首先，我们选取趋势项，市场波动项和噪音项作为深圳综指收盘价真实值的三个维度的特征，把总共1212组样本导入到随机森林进行训练，使用初始训练集的（约1010个样本）作为训练对象，则剩下的（约202个样本）可用作训练集来对泛化性能进行“包外估计”，得到的测试对象的预测值与真实值对比如下图：



图表 4随机森林测试对象预测值与真实值比较

从图中我们可以看出，预测值与真实值的图像拟合度极高，说明随机森林的算法表现十分良好。由此说明，只要股市所处的大环境没有发生重大变化，如果我们可以通过一系列经济的或者非经济的事件，通过与历史事件比较，准确地判定出长期特征趋势项，中期特征市场波动项和短期特征噪音项的取值，就可以对股指的涨跌趋势作出较为精准的预测。

下面我们选取2015年1月5号到2016年6月30号深圳综指的收盘价作为新的原始时间序列，同样通过奇异谱分析（SSA）提取出它的特征序列，包括趋势项，市场波动项和噪音项，导入到随机森林中作为预测的输入对象，由此可以得到对应时间段内的预测数值。

具体得到的图像如下图：

图表 5深圳综指股票价格指数预测值和真实值对比

把通过相同随机森林算法得到的深圳综指的预测值与真实值对比，通过图像可以看出，二者在波动上具有相当强程度的一致性。其中，测试结果的精确度与预测结果的精确度差别在于，测试的样本集来自于与训练集相同的SSA过程，而预测的样本集是独立于测试和预测的样本集而通过单独通过SSA得到的。所以，在SVD分解时，训练集和测试集具有相同的矩阵U和V，而预测集具有不同的矩阵。长期来看，我们可以视测试集与训练集来自于相同的金融环境，而预测集与训练集来自于不同的金融环境。

在2015年，中国金融市场可以说险象环生，不仅经历了A股市场的千股震荡、指数的巨大振幅、股市走势大涨后大跌、央行数次的降息降准，同时还迎来了互联网金融的蓬勃发展。考虑到金融市场本身的复杂性，排除这些史无前例的特殊事件的影响后，我们可以看到预测值的涨跌与真实值呈现相似的趋势，由此可以得到，尽管通过SSA分析提取出的特征序列在预测时会受到环境变化的限制而影响数值的精确预测效果，但通过SSA分解出特征序列后再引入到随机森林中进行预测的确对于股指涨跌趋势预测具有良好的指导作用。

**4．结论**

本文选择了2010年1月4号到2014年12月31号的深圳综指收盘价数据作为原始序列，通过SVD分解为趋势项，股价波动项和噪音项，然后在不同的时间维度上来研究其特征。研究结果表明，趋势项在长期中对于股指有决定性作用，因为趋势项与中国的经济状况密切相关，它也能在一定程度上反映国家的经济状况；明显的波动几乎是由于各种各样的事件的冲击所导致的，并且这些事件对股指的影响通常会持续一段时间，从几周到几个月不等。而且，对过去事件的分析有利于提高分析的准确性。股指的局部变化是由于噪音交易者导致的，尽管他们不能对股指的长期趋势造成影响，他们在短期预测中起到大作用。之后，我们利用代表不同特征的奇异值重构出代表原始序列三个特征的时间序列，并且把它们和原始序列一起引入到随机森林算法中，随机选择样本的作为训练对象，剩下的作为测试对象，然后同样对2015年1月5号到2016年6月30号的深圳综指收盘价数据进行SSA分解出趋势项，市场波动项和噪音项，导入相同的随机森林算法中进行预测。测试结果表明，在股市所处大环境不变的情况下，通过股市的长期特征，中期特征，和短期特征，随机森林算法即对股指的涨跌具有精确的预测效果。而预测结果表明，即使股市所处的环境发生变化，考虑到金融市场本身的复杂性，尽管通过SSA分析提取出的特征序列在预测时会受到环境变化的限制而影响数值的精确预测效果，但通过SSA分解出特征序列后再引入到随机森林中的股指分析和预测的方法对于股指的涨跌趋势仍然具有卓越的指导意义。

**参考文献：**

**Reference**

1. Cheng W, Wanger L, Lin CH., Forecasting the 30-year US treasury bond with a system of neural networks , *Journal of Computational Intelligence in Finance*, 4, (1996), pp.10-16
2. Sharda R, Patil RB. A connectionist approach to time series prediction: an empirical test. In: Trippi, RR, Turban, E, (Eds.), Neural Networks in Finance and Investing, Chicago: Probus Publishing Co., (1994), pp. 451-464.
3. David Enke, Suraphan Thawornwong. The use of data mining and neural networks for forecasting stock market returns, Expert Systems with Applications, 29, (2005), pp.927–940.
4. Mukherjee, S., Osuna, E., Girosi, F., Nonlinear prediction of chaotic time series using support vector machines , in: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, Amelia Island, FL, (1997), pp.511-520.
5. Tay, F. E. H., Cao, L., Application of support vector machines in financial time series forecasting , Omega, 29, (2001), pp.309-317.
6. Kim, K.J., Financial time series forecasting using support vector machines , *Neurocomputing* , 55,(2003),pp. 307-319
7. Breiman, L., Random Forests , *Machine Learning*, 45, (2001),pp. 5-32
8. Creamer, G.G. , Freund, Y. : Predicting performance and quantifying corporate governance risk for Latin American adrs and banks. In: Financial Engineering and Applications. MIT, Cambridge (2004)
9. Lariviere, B., Poel, D. V. D., Predicting Customer Retention and Profitability by Using Random Forests and Regression Forests Techniques , Working Paper, Department of Marketing, Hoveniersberg 24, 9000 Gent, Belgium, (2004).
10. Wen F, He Z, Dai Z, Yang X. Characteristics of investor’s risk preference for stock markets, *J. Economic Computation and Economic Cybernetics Studies and Research*, 48(3) (2014) pp.235-254.
11. Wen F, Li Z, Xie C, Shaw D. Study on the fractal and chaotic features of the Shanghai composite index, J. Fractals-complex geometry patterns and scaling in nature and society, 20(2) (2012) pp.133-140.
12. Broomhesd, D. S. King, G. P. Extracting Qualitative Dynamics from Experimental Data, J. *Physica D:Nonlinear Phenomena*, 20(2-3) (1986) pp.217-236.
13. Hossein Hassani, Singular Spectrum Analysis: Methodology and Comparison, *J. Data Sci*, 5(1) (2007)pp. 239-257.
14. Lian Jijian, Li Huokun, Zhang Jianwei. Hydranlic structure vibration modal ERA identification method based on the singular entropy order noise reduction, *J. SCI CHINA SER*, 09 (2008) pp.1398-1413.
15. Beneki,C.,Eeckels,B.,and Leon,L.,Signal Extraction and Forecasting of the UK Tourism Income Time Series: A Singular Spectrum Analysis Approach, *J. Journal of Forecasting*, 31(5) (2012) pp.351-400.
16. 张一，惠晓峰. 基于奇异谱分析的汇率预测研究, *统计与决策*, 06 (2012) pp.29-31.
17. Beneki C, Silva E S. Analysing and forecasting European Union energy data, *J. International journal of Energy and Statistics*, 02 (2013) pp.127-141.
18. Huang C, Gong X, Chen X, Wen F. Measuring and Forecasting Volatility in Chinese Stock Market Using HAR-CJ-M Model, J/OL. *Abstract and Applied Analysis*, 2013 (2013) pp.1-13.
19. Wu Zhao-hua, Huang N E. Ensemble empirical mode decomposition: a noise assisted data analysis method, *J. Advances in Adaptive Data Analysis*, 1(1) (2009) pp.1-41.
20. Breiman, L., Bagging predictors , *Machine Learning*, 24(2), 1996a, pp.123-140.
21. Efron, B. and R. Tibshirani. An Introduction to the Bootstrap. Chapman & Hall, New York, NY. (1993)
22. Zhang X, Lai k k, Wang S Y.A New Approach for Crude Oil Price Analysis Based on Empirical Mode Decomposition, *J. Energy Economics*, 30(3) (2008) pp.905-918.