**摘 要**

21世纪是计算机行业与金融行业蓬勃发展的黄金时代，我国的债券交易产业也在进入21世纪之后发展得如火如荼，然而随着金融产业的发展，债券交易市场越来越具有复杂性，许多不可预知的风险威胁着债券交易市场，传统的金融知识以及金融工具已经很难再保障债券交易的安全性，这就需要使用新的工具、新的方法来为债券交易进行保驾护航。众所周知金融知识中运用了许多的数理统计、线性代数与概率论知识，计算的过程中还需要采集大量的历史数据进行分析和推断，使用这些方法进行人工计算往往需要消耗大量的人力物力，但是计算机与大数据概念的出现，使得金融问题的解决有了更高效的方案。该毕业设计主要解决了以下一些问题：

1. 使用Python爬虫获取债券交易数据，实现了用户自定义收益率数据的收集。
2. 通过PCA主成分分析方法研究债券标准期限结构，提取出了3个主成分，计算出了三个主因子对固定收益产品价值波动的贡献度，并分析了这些主因子对债券期限结构的影响。
3. 通过蒙特卡洛模拟法计算了给定债券的在险价值（风险价值），使得用户对于债券的投资风险有了直观的概念。
4. 通过插值法实现了基于Nelson-Siegel模型的收益率曲线拟合功能，使得用户能够拟合自定义的收益率曲线。
5. 通过Web前端技术结合Node Express代理和MongoDB数据库实现软件开发，为用户提供数据展示、数据查询、曲线拟合、PCA分析、VaR计算等功能。

关键词：收益率曲线；Nelson-Siegel；PCA；数据采集；VaR； Node Express；MongoDB

**Abstract**

The 21st century is the golden age for the booming development of the computer industry and financial industry. China's bond trading industry is also developing rapidly in the 21st century. However, with the development of the financial industry, the bond trading market is increasingly complex and unpredictable. The risk poses a threat to the bond trading market. Traditional financial knowledge and financial instruments can no longer guarantee the security of bond trading. This requires to use new tools and new methods to protect bond transactions. It is well known that many mathematical statistics, linear algebra and probability theory are used in financial knowledge, and a large amount of historical data needs to be collected in the calculation process for analysis and inference. Manual calculation using these methods often requires a lot of manpower and material resources, But with the emergence of computer and the concept of big data, a more efficient solution has been made to the solution of financial problems. This graduation design mainly solves the following problems:

1. Using Python spiders to obtain bond transaction data, realized the collection of user-defined yield data.
2. Using the principal component analysis(PCA) method to study the bond term structure, extracting three principal components, calculating the contribution of the three main factors to the value fluctuation of fixed income products, and analyzing the term structure of bonds with these principal factors’ impact.
3. Calculate the value at risk (VaR) of a given bond through the Monte-Carlo simulation method, so that the user has an intuitive concept of the investment risk of the bond.
4. The Nelson-Siegel model-based yield curve fitting function is implemented by the interpolation method, so that the user can fit a customized yield curve.
5. Software development through Web front-end technology combined with Node Express agent and MongoDB database to provide users with data display, data query, curve fitting, PCA analysis, VaR calculation and other functions.

**Key words**: yield curve; Nelson-Siegel; PCA; data acquisition; VaR; Node Express; MongoDB

**目录**

第1章 绪论

贵阳高登世德金融科技公司是一家中国资产证券化行业内顶尖的金融科技公司，主要负责为金融机构设计资产证券化产品以及优化证券化产品的评级优化，而在这家公司的实习促进了这个毕业设计的诞生，公司的客户需要一款能够实现债券数据采集、分析、期限结构影响因子挖掘以及在险价值计算的软件，所以这个毕业设计的主要目的就是设计一款符合客户需求的债券分析软件。

1.1 研究背景

在停止发行国债22年之后，1981年我国恢复发行国债，从1981年以来，中国政府债券市场已经经历了近30多年的发展。中国国债在筹集财政资金、支持国家重大项目建设、改善经济结构等多方面都发挥了很大的作用。在这近三十年的发展过程中，政府债券规模不断扩大，政府债券品种不断完善，政府债券市场也从无到有，建立了多层次的政府债券市场体系。但与此同时，中国国债市场还具有许多的不足，比如立法基础薄弱、市场分割、投资者结构不合理、市场流动性不足等等。

我国最早在1984年发行了企业债券，到现在为止，经历了萌芽阶段（1984-1986）、快速发展阶段（1987-1992）、整顿阶段（1993-1995）和规范发展阶段（1996至今）。目前而言，我国上市公司债券规模还不够大，远远落后于股票的发行，但是从1984年发行企业债券以来，我国企业债券发行总量有了大幅度增长，发行主体也越来越呈现出多元化的趋势，企业债券品种越来越丰富，其作为投资工具的地位也越来越重要。随着对企业债券的限制越来越少，企业债券发行节奏明显加快。发行债券的品种和主体都越来越丰富。

伴随着我国国债和企业债市场的飞速发展，诞生了一些优秀的债券数据分析管理企业，比较著名的比如中债登、万得、和讯等企业，这些企业积累了大量的中国债券市场交易数据，比如收益率数据、估值数据、VaR数据、成交指数等等，这些企业还拥有了比较丰富的国内债券数据处理经验和数据，比如中债登采用独特的Hermite插值法拟合中国国债收益率曲线，采用历史模拟法实现VaR的计算。通过这些优秀的企业，可以获取到许多宝贵的数据以及数据处理方法，进而可以对债券进行评估定价和风险预测，达到毕业设计的目的。

1.2 研究目的和意义

在购买债券时，应该如何判断一支债券是否值得购买呢？债券购买者往往通过收益率曲线来判断，研究债券收益率曲线具有重要的意义,对于投资者而言,可以用来作为预测债券的发行投标利率、在二级市场上选择债券投资券种和预测债券价格的分析工具；对于发行人而言,可为其发行债券、进行资产负债管理提供参考，所以收益率曲线以及收益率数据的收集与研究就显得十分有价值。

本设计通过从各大证券交易所收集收益率数据并且对这些数据进行曲线拟合、主因子提炼、分析主要因子的影响以及判断未来曲线的走势预测债券购买价值并给出投资策略。

本设计的意义：通过对已有数据的分析挖掘，提炼出曲线以及方程，根据PCA方法分析主要影响因子，根据这些因子可以给出未来的债券投资策略，达到预测债券走势的目的。

虽然该毕业设计主要虽然研究的是固定收益债券定价和投资策略分析，但是主要研究对象还是收益率曲线。要了解设计的目的，首先要了解收益率曲线。

收益率曲线是根据不同到期期限债券的收益率所绘制的图形，它反映市场当前的收益率水平。收益率曲线不是历史数据图形，不能反映一段时间内的收益率水平，反映一段时间内收益率水平的图形称为历史价格图或者历史收益率图形。

1.3 国内外研究现状

在中国，由于包括国债市场在内的债券市场起步较晚，流动性和市场化程度尚需进一步加强，有关利率期限结构和国债收益率曲线的研究也相对较薄弱，这增加了对国债收益率曲线进行研究的必要性和迫切性。国内最早的关于收益率曲线的公开研究文献是杨大楷和杨勇姚长辉和梁跃军。

由于中国债券市场的发展起步较晚，对中国债券市场的理论研究落后，收益率曲线研究始于90年代后期。近年来值得一提的工作是应用国外广泛使用的模型对我国债券收益率曲线进行实证研究和比较分析。研究表明，针对中国的现状短期债券发行量小，长期性差，中期债券比例大，不健康的期限结构，Nelson-Siegel模型适用于构建中国国债收益率曲线。应当利用中国债券市场数据，利用国际上公认的几种收益率曲线拟合模型进行实证研究，寻找适合中国债券市场的拟合收益率曲线的方法。

Diebold和Li（2006）、Diebold、Rudebusch、Aruoba（2006）、Diebold Piazzesi和Rudebusch（2005）表明，在美国的封闭经济环境中，广义的Nelson-Siegel模型精确地预估了收益率曲线动态，提供了良好的预测。

Diebold-Li分解单个国家的Nelson-Siegel收益率曲线是：

其中表示月债券的连续复利的零息票名义收益率，、、、、均为待定参数，并且具有标准偏差的干扰。继Diebold和Li之后，通过允许参数随时间变化来动态化模型：

式中解释为为潜在因素，特别是，如Diebold和Li所示，它们分别是水平参数，斜率参数和曲率参数，因为这些因子参数是常数，所以这是一个关于 的递减函数和凹函数。由于收益率因子随时间而变化，这种广义的Nelson-Siegel模型可以生成各种随时间变化的收益率曲线形状。

1.4 研究内容

该论文主要描述一款基于Web前端页面、Node Express后台接口、Python数据采集分析脚本、PCA算法、Monte Carlo算法以及Mongo DB数据库的债券数据采集分析软件的开发过程。这款软件实现了用户自定义债券收益率数据的查询、自定义债券收益率曲线的拟合、自定义债券期限结构数据的PCA分析以及自定义债券的在险价值计算。

这款软件的核心在于后台Python脚本程序，通过Python脚本程序，实现了债券数据采集和过滤、债券数据的PCA分析、债券收益率曲线的拟合以及债券在险价值的计算，这一部分算法以及数据都是整个软件的核心。通过Python程序采集到数据后并进行过滤整理后，将所有数据分类保存在MongoDB数据库中，以便之后Node脚本程序和Python程序对数据进行查询调用。在前端部分，采用了目前最为流行的MVC框架——React进行开发，负责呈现软件的用户交互界面，包括数据展示、收益率曲线展示、用户输入界面等等。后台部分，使用当前比较流行的Node.js以及Node.js构建的Express框架作为后台代理，实现了RESTful风格的后台接口开发以及服务器搭建。另外通过npm模块实现了Node脚本程序与Python脚本程序之间的交互，使得Node服务器可以调用Python程序进行数据处理。