答辩演讲稿

尊敬的各位老师、亲爱的各位同学：

大家好，我是测控zy1401班的钟德鸣，接下来由我给大家展示我的毕设成果。我的毕设题目是《基于PCA算法与N-S模型的债券数据分析系统的设计》，我的指导老师是牟新刚副教授。

接下来的展示，主要包括以下５个方面，分别是①……②……③……④……⑤……。

首先是介绍研究背景及意义，可以概括为四点。第一，债券行业在中国的发展势头十分迅猛，根据中债登的数据，中国在2009年全国债券发行总额为86474.77亿元，而2017年全国发行债券总额达到了135795.45亿元。债券发行总额在８年时间接近翻了一倍，足以说明债券行业在中国的迅猛发展。随着越来越多的人开始购买债券，消费者的专业素质也在不断提升，他们对于债券购买的态度不再随便，需要更加可靠的工具帮助他们实现债券定价、风险评估和趋势分析，而本设计正是基于这些需求和背景，帮助用户实现所需功能。

那么具体实现了哪些功能呢？主要有以下六个：①……②……③……④……⑤……⑥……。

研究成果是一个网站，具有以下六个功能：①……②……③……④……⑤……⑥……。如果有时间的话，稍后将会播放一个视频展示我的网站。

算法部分包括PCA算法、N-S模型和M-C模拟法。读PPT……

我先介绍PCA算法，这里有一个例子，我们有一份如上图的数据用于描述各省的经济发展状况，针对这么多的维度，比较难得出各省经济发展状况的结论。这时可以对这些数据进行降维，通过PCA算法，可以将10个维度的元数据降为两个维度甚至一个维度。这样对比起来将简单很多，从降维后的数据可以看出来经济发展最好的是广东省。这就是PCA降维算法的作用。我们之所以可以得出结论，是因为PCA算法在信息损失最小的情况下，将多维数据降维，降维后的数据仍然保留了元数据中大部分的信息。PCA算法在图像识别等领域应用广泛。

手动计算或编写代码实现PCA算法，需要进行以下步骤：①……②……③……④……⑤……。我先进行实证分析，通过SPSS软件进行PCA处理。

下面这组数据是中国国债在2002年～2017年全年度每日债券收益率数据，共有3996条数据，14个维度，分别是0,0.08,0.17,0.25,0.5,0.75,1,3,5,7,10,15,20,30年到期期限。

将这3996条数据导入到SPSS软件，效果如图。

经过SPSS分析之后，输出结果如图所示。右边图片是总方差解释图，代表了１４个维度对于债券收益率曲线走势的影响的贡献度，总的贡献度一定为１。从左边的图可以看出，前三个主成分可以解释债券收益率曲线走势的绝大部分因素。

这是成分矩阵表，可以理解为特征值排行前三的三个维度对应的特征向量所组成的矩阵。

成分矩阵的每一项，除以它对应的特征值的平方根，就可以得到各指标对应系数。所有指标对应系数可以组成因子载荷矩阵。原始数据乘以因子载荷矩阵可以得到降维后的数据。所以通过研究因子载荷矩阵，可以知道每个主成分对债券走势的影响。

通过研究中美国债历史数据以及SHIBOR的历史数据，可以得到发现第一主成分很平稳，基本一直为正数；第二主成分先负后正，对长期债券有正面影响，短期债券有负面影响；第三主成分最为复杂，对短期和长期债券往往有相同的影响，而对中期债券的影响往往相反。根据Diebold和Li的研究，这三个主成分根据它们对债券收益率走势的影响可以依次称为水平因子、斜率因子和曲率因子，它们分别影响债券收益率曲线的水平位移、斜率和曲度。

接下来介绍N-S模型，由于通常得到的数据只包含标准期限结构，所以在图上是一系列离散的点，我们现在要根据这些点构造一条光滑的曲线，以实现非标准期限结构的收益率预测，这个过程就称为收益率曲线拟合。

N-S模型的公式如下，其中…………。其实还有其他的拟合模型可以选择，比如多项式拟合、Hermite拟合，为什么选择N-S模型呢？

针对下面这组数据，我对三个拟合方案进行了测试。拟合效果评判标准是RMSE（均方根误差，标准差），拟合效果如下，RMSE如下，…………。综上，选取N-S模型拟合收益率曲线。

N-S模型拟合还有以下优点，比如名气大，构造简单以及参数的金融含义强。

软件开发完成后，实际的曲线拟合效果如下，与中债网的拟合结果进行对比，可以发现误差非常小。

接下来介绍M-C模拟法，这是诞生于二战中的一种随机数模拟算法，科学家为了解决核武器研发中无法解决的计算，通过随机数模拟结果解决问题，后来用著名的“赌城”蒙特卡洛命名这个算法。上图就是一个最基本的M-C算法的使用实例——通过随机数模拟计算pi。

M-C模拟法适用于计算VaR在险价值的，在险价值可以用如下数学表达式表示，它的含义是在指定置信区间下可能的最大损失，VaR方便了用户对于投资风险的评估。

计算VaR还使用了德尔塔正态法，其公式如下……。

结合M-C模拟法计算在险价值的步骤如下……。

实证分析，在中债网上找到了当时债券代码为1282337的债券的当日VaR如图。

在万德网上，我购买到了这支债券４月２０号之前一个月左右的日收益率数据。

通过Python模拟，得到一万条符合正态分布的数据表。

这是这一万条数据的分布图。

通过我参照之前步骤所实现的Python程序的计算，得到在险价值如图，跟中债网的在险价值的对比可以发现相差不多。

下面介绍软件开发部分的工作，也是主要分为三个部分，分别是……。

架构设计图如图，其中用户视图层负责与用户进行交互，即为GUI；后端业务逻辑层由Node和Python程序共同实现（包括数据查询、分析、收集等），数据存储在服务器的数据库中；中间代理层由Node实现，我基于Express搭建了一个HTTP服务器，中间层的Node也可以调用Python脚本。

这是我的开发流程，基本上是前端后台一起开发，这样方便联动调试API。

如图是前端开发所用的主要模块或框架，React是最主要的视图层开发框架，Highcharts实现收益率曲线的绘制，Antd实现UI开发，Axios是Ajax操作的封装库，可以实现调用API，React-Router实现前端路由开发，当网址变化时，对应的组件也会发生变化。

这是我的路由设计，根据路由就可以开始设计组件。

开发完成后，通过Webpack打包代码，得到static文件夹中的文件，拷贝到服务器目录即可实现部署。

下面是Python开发，利用Python脚本实现的功能有这三个。

通过Python爬虫获取数据，分为两种情况，基本原理如下，此外如果网站进行了防爬虫处理（和讯），就需要伪造请求头，否则无法爬取到数据。

本设计Python爬虫开发主要依赖以下两个模块。

Python爬虫获取到的原始数据有许多缺陷，比如维度不统一，数据缺失等等，所以需要对原始数据进行处理。利用Python进行数据处理，主要遵循以下原理，……依赖以下两个模块。

Python实现数据分析主要依赖之前介绍的算法实现步骤，实现起来也比较简单，这里不在介绍了。

Node框架Express实现的服务器主要通过一系列中间件实现请求处理，这些中间件基本结构如下，其中……。

数据库我选择MongoDB，它是由C++编写的NoSQL非关系型数据库，存取快速，由于使用BSON格式保存数据，所以和JSON的关系非常亲密，适合于JavaScript和Python的程序开发。

通过MongoDB可视化程序查看数据如图，共有八个集合。

Python程序通过pymongo驱动进行数据库操作，Node通过mongodb驱动进行数据库操作。

这是我在本次毕设中的几点收获，①……②……③……④……。

这是我在工作中的几点不足，还可以再进行优化。

我的介绍基本完毕，最后，感谢各位老师和各位同学的耐心倾听！