曲线拟合总结

**曲线拟合的目的及概念**

通常，债券购买者得到的期限结构数据是二维空间中一系列离散的点，但是为了更加直观的观察债券收益走势，更加合理的利用期限结构获取更多有用的信息，通常会利用这些离散的点构造一条平滑的曲线，即为收益率曲线，利用离散的点构造这条收益率曲线即为收益率曲线拟合。

比如，我们从中债网上可以获取到2018年4月13号当日的债券收益日信息，填写在表格中，即：

|  |  |
| --- | --- |
| Maturity(Year) | Yield(%) |
| 0.0 | 2.2316 |
| 0.08 | 2.5617 |
| 0.17 | 2.5693 |
| 0.25 | 2.6842 |
| 0.5 | 3.0242 |
| 0.75 | 3.1843 |
| 1.0 | 3.2013 |
| 3.0 | 3.3682 |
| 5.0 | 3.5205 |
| 7.0 | 3.6916 |
| 10.0 | 3.7062 |
| 15.0 | 3.9689 |
| 20.0 | 4.0079 |
| 30.0 | 4.1207 |
| 40.0 | 4.162 |
| 50.0 | 4.1808 |

使用Excel描绘散点图：



由此可以得到一个期限结构数据描绘的散点图，但是这张散点图不够直观，如果债券购买者想要获得到期期限为35个月的债券的收益率数据时，将无从下手，所以需要根据这些二维空间中的散点，构造一条光滑曲线，这条曲线并不只是简单的将散点用光滑曲线连接，还需要对未给出的节点做出价格推测。这就是曲线拟合以及曲线拟合的目的和意义。

下图是中债网所给出的收益率曲线：



**曲线拟合方法及模型的选择**

为了得到一条这样的拟合曲线，首先要选择合适的拟合方法和模型。常见的收益率曲线拟合方法或模型有多项式插值法、Hermite模型插值法、Nelson-Siegel模型、Hull-White等等。接下来会对其中几种模型或方法的拟合精度进行评估，使用到的评估标准是均方根误差（RMSE）。

1. 多项式插值法

构造多项式拟合已有收益率数据，然后通过插值构造平滑收益率曲线的方法即为多项式插值法。

多项式中拟合收益率曲线效果最好的是三次多项式，即为立方多项式。可以通过Python程序进行拟合，并查看拟合效果。

首先在Python中通过numpy工具进行三次多项式拟合，拟合的三次多项式系数：

|  |  |  |  |
| --- | --- | --- | --- |
| 三次项系数 | 二次项系数 | 一次项系数 | 常数 |
| 5.775974408183625e-05 | -0.005396803801863335 | 0.15538879446115653 | 2.7450925385495 |

然后计算出所有的估计值，然后计算出均方根误差进行输出，下面是程序运行的结果：



说明对于TestData.csv这一组数据而言，利用三次多项式进行拟合，计算出的均方根误差是0.045。

下面是拟合数据与原始数据的对比：



从图中可以看出，三次多项式的拟合效果较差，误差较大，甚至出现了折现的情况，不符合我们需要平滑曲线的要求。

1. Hermite插值法

Hermite插值法在国外使用的不是很多，但是这里之所以要提Hermite插值法，是因为它比较适合于拟合中国债券市场的收益率曲线，比较有代表的使用者就是中债网，中债网使用Hermite多段插值法进行收益率曲线拟合，具有较好的拟合效果。

中债网使用的是多段插值法，具体做法是在每两个节点之间，使用三次Hermite进行插值，所以根本上而言，这还是使用的三次多项式模型，但是这种方法不仅要求节点处的值要相同，而且在节点处的倒数也要相同，所以拟合这与一般的三次多项式拟合存在不同。

接下来我们仍然使用Python程序进行拟合，并计算RMSE进行效果比较，这里Python程序没有实现Hermite算法，使用的是已有的工具包——numpy。

通过Python程序拟合，我们得到了拟合的三次Hermite多项式，其各项系数如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 三次项系数 | 二次项系数 | 一次项系数 | 常数 |
| 7.21996801e-06 | -1.34920095e-03 | 7.77377170e-02 | 2.74239414 |

拟合后计算出的均方根误差为：



这说明相对于TestData.csv这组数据而言，Hermite插值法拟合曲线得到的数据均方根误差为0.094，这比三次多项式拟合的效果还要差。

其拟合数据与原数据对比如下：



1. Nelson-Siegel插值法

Nelson-Siegel模型是金融届（尤其是债券行业）大名鼎鼎的一个模型，它是1987年Nelson C.和Siegel A.提出来的一个参数拟合模型，它需要确定的参数少（只有四个），但是拟合的效果却很好，而且每一个参数都具有很强的金融含义。

其具体公式如下：

其中，、、是三个具有金融含义的参数；是与时间有关的常数，可称为时间常数；是到期期限，y是到期收益率。

同样的，通过Python程序进行Nelson-Siege模型拟合，由于有四个参数需要确定，所以会存在多组解，我们通过优化方法，使得拟合出的效果最好（以各项数据方差之和为参考标准），最终效果最好的一组参数即为我们拟合的Nelson-Siegel模型的四个参数。

最终拟合的Nelson-Siegel模型的四个参数如下：

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 4.34118883 | -1.28633087 | 4.45571531e-03 | 0.194545496 |

最终计算出的RMSE为：



拟合效果：



结合RMSE来看，三次多项式的拟合效果要好于Nelson-Siegel模型拟合的效果，但是通过曲线观察，可以发现Nelson-Siegel模型的拟合效果要好于三次多项式的拟合效果，这是因为在使用Nelson-Siegel模型拟合曲线的时候，由于其他因素限制，只使用了整数到期期限的数据进行拟合，所以导致其RMSE比三次多项式更低，由于这个原因，可以认为Nelson-Siegel模型的拟合效果好于三次多项式，所以该毕设还是采用Nelson-Siegel模型结合插值法进行收益率曲线的拟合。

此外，采用Nelson-Siegel模型进行拟合的原因还有以下几点：

1. Nelson-Siegel在业界久负盛名，非常适合与拟合收益率曲线。根据统计[参考中国动态Nelson-Siegel利率期限结构模型研究]，比利时、芬兰、法国、意大利、西班牙、等多个国家的中央银行采用Nelson-Siegel模型作为利率期限结构的估计方法。还有许多其他国家采用的是基于Nelson-Siegel模型的Nelson-Siegel-Svensson模型，所以Nelson-Siegel模型的名气可见一斑。
2. Nelson-Siegel模型构造简单，只需要计算四个参数即可完成构建。
3. Nelson-Siegel模型的四个参数具有强烈的金融含义，所以容易该模型更理解。之前说过参数是只与时间有关的常数，不同日期的收益率数据可能不相同。如果我们确定了，以、、和为参数，我们可以做出、、三个参数的系数随发生变化的图形：



上图实际上就是影响债券收益率走势的三个主成分，