

# RISK SIMULATOR

## 用户手册

**Johnathan Mun, Ph.D., MBA, MS, BS, CRM, FRM, CFC, MIFC**  
**Real Options Valuation, Inc.**

R I S K  
R I S K



本手册及其中涉及的软件由 Real Options Valuation, Inc 公司授权提供, 用户须在遵守《最终用户许可协议》条款的情况下使用和复制。本文件中所提供的信息仅供参考, 随时可能更改, 不另行通知, 不代表 Real Options Valuation, Inc 公司的商业义务。不管出于何种目的, 在没有 Real Options Valuation, Inc 公司书面授权情况下, 任何人不得以任何形式、任何方式, 包括影印和录音等方式复制或传播本手册的任何章节。本文参考了 Real Options Valuation, Inc 公司创始人及首席执行官 Dr. Johnathan Mun 的出版物。由 Dr. Johnathan Mun 编写。在美国境内编写设计出版。如想购买本手册的补充读物, 请按以下的 e-mail 地址联系 Real Options Valuation, Inc 公司: [admin@real-consulting.com](mailto:admin@real-consulting.com) or visit [www.realoptionsvaluation.com.cn](http://www.realoptionsvaluation.com.cn)。

© 2005 -2012 by Dr. Johnathan Mun. 公司保留一切解释权利。Microsoft®是微软公司在美国境内及其它国家的注册商标。下文中涉及的其他产品可能是属于相关拥有者的商标和/或注册商标。

© Copyright 2005-2012 Dr. Johnathan Mun. All rights reserved.

Real Options Valuation, Inc.

4101F Dublin Blvd., Ste. 425

Dublin, California 94568 U.S.A.

Phone 925.271.4438 • Fax 925.369.0450

[admin@realoptionsvaluation.com](mailto:admin@realoptionsvaluation.com)

[www.risksimulator.com](http://www.risksimulator.com)

[www.realoptionsvaluation.com](http://www.realoptionsvaluation.com)



## 目录

<b>1. 系统简介</b> .....	<b>5</b>
1.1 欢迎使用 Risk Simulator 软件 .....	5
1.2 安装和授权过程 .....	6
1.3 授权过程 .....	6
1.4 2011/2012 版有哪些新增功能? .....	9
1.4.1 一般功能 .....	9
1.4.2 仿真模块 .....	10
1.4.3 预测模块 .....	10
1.4.4 优化模块 .....	11
1.4.5 分析模块 .....	11
1.4.6 统计学和 BizStats 模块 .....	12
<b>2. Monte Carlo 仿真</b> .....	<b>14</b>
2.1 什么是 Monte Carlo 仿真? .....	14
2.2 开始学习 Risk Simulator .....	15
2.2.1 软件的总体介绍 .....	15
2.2.2 运行 Monte Carlo 仿真 .....	15
2.3 相关性和精度控制 .....	26
2.3.1 相关性的基础知识 .....	26
2.3.2 在 Risk Simulator 中应用相关性 .....	27
2.3.3 相关性对 Monte Carlo 仿真的影响 .....	27
2.3.4 精度和误差控制 .....	29
2.3.5 对预测统计量的理解 .....	31
2.3.6 了解 Monte Carlo 仿真的概率分布 .....	34
2.4 离散分布 .....	36
2.5 连续分布 .....	42
<b>3. 预测</b> .....	<b>59</b>
3.1 不同类型的预测方法 .....	59
3.2 运行 Risk Simulator 中的预测工具 .....	60
3.3 时间序列分析 .....	61
3.4 多元回归 .....	64
3.5 随机预测 .....	67
3.6 非线性外推法 .....	69
3.7 Box-Jenkins ARIMA 高级时间序列 .....	71
3.8 自回归求和移动平均模型(Box-Jenkins ARIMA 高级时间序列 ) .....	76
3.9 基本计量经济学模型 .....	76
3.10 J-S 曲线预测 .....	77

3.11 GARCH 波动率预测.....	79
3.12 马尔可夫链.....	81
3.13 最大似然估计模型 (MLE) : Logit, Probit, Tobit .....	81
3.14 样条模型 (三次样条内插和外插模型).....	83
<b>4. 优化 .....</b>	<b>86</b>
4.1 优化方法.....	86
4.2 连续型决策变量的优化.....	88
4.3 离散整数型优化.....	92
4.4 有效边际和高级优化选项.....	95
<b>5. 风险仿真分析工具.....</b>	<b>97</b>
5.1 仿真中的飓风和敏感性分析工具.....	97
5.2 敏感性分析.....	104
5.3 分布拟合: 单变量和多元变量.....	107
5.4 Bootstrap 仿真 (拔靴法) .....	111
5.5 假设检验.....	113
5.6 数据输出和保存仿真结果.....	114
5.7 创建报告.....	115
5.8 回归和预测诊断工具.....	117
5.9 统计分析工具.....	123
5.10 分布分析工具.....	127
5.13 Risk Simulator 2011/2012 新工具 .....	132
5.14 随机数生成, 蒙特卡罗与拉丁超立方, 以及相关 Copula 方法.....	132
5.15 数据季节性因素和趋势性因素消除.....	133
5.16 主成分分析.....	134
5.17 结构性中断分析.....	135
5.18 趋势线预测.....	136
5.19 模型检验工具.....	137
5.20 百分位分布拟合工具.....	138
5.21 分布图表: 概率分布工具.....	139
5.22 ROV BizStats.....	143
5.23 神经网络和组合模糊逻辑预测方法.....	146
5.24 优化器目标搜寻.....	149
5.25 单变量优化.....	150
5.26 遗传算法优化.....	151
5.27 ROV 决策树模组 .....	153
5.27.1 决策树.....	153
5.27.2 仿真模型.....	154
5.27.3 贝叶斯分析.....	155
5.27.4 精确信息的预计数值, 最大最小分析, 风险文件, 不精确信息的数值 .....	155
5.27.5 敏感性分析.....	155

5.27.6 情景表 .....	156
5.27.7 生成效用函数 .....	156
<b>6. 帮助贴士和技巧 .....</b>	<b>164</b>
贴士：假设（设定输入假设用户界面） .....	164
贴士：复制和粘贴 .....	164
贴士：相关性 .....	165
贴士：数据诊断和统计分析 .....	165
贴士：分布分析，图表 .....	165
贴士：有效边际 .....	165
贴士：预测单元格 .....	165
贴士：预测图表 .....	165
贴士：预测 .....	166
贴士：预测：ARIMA .....	166
贴士：预测：基本计量经济学 .....	166
贴士：预测：Logit, Probit 以及 Tobit .....	166
贴士：预测：随机过程 .....	166
贴士：预测：趋势线 .....	166
贴士：函数调用 .....	166
贴士：入门练习和视频 .....	167
贴士：硬件 ID .....	167
贴士：在线资源 .....	167
贴士：优化 .....	167
贴士：文档 .....	167
贴士：右键快捷键或者其他快捷键 .....	168
贴士：保存 .....	168
贴士：抽样和仿真技术 .....	168
贴士：软件卡发工具（SDK）和动态链接库 .....	168
贴士：在 Excel 中使用 Risk Simulator .....	168
贴士：极速仿真 .....	169
贴士：飓风图分析 .....	169
贴士：Troubleshooter .....	169

# 1. 系统简介

## 1.1 欢迎使用 Risk Simulator 软件

Risk Simulator 版本是一套集 Monte Carlo 仿真，预测和优化为一体的软件。本软件使用 microsoft.NET #C 作为开发工具，并附加具备 Excel 功能。此外本软件还与 Real Option Valuation 公司的 SLS 软件和 ESOV 软件兼容，可同时使用。注意，尽管我们努力让本手册包含所有的内容，但是它不能代替配套的 DVD，培训课程以及软件开发者的相关著作（例如：Dr.Johnathan Mun 的 *Real Option Analysis* 《实物期权分析》，第二版，Wiley Finance 2005；*Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《风险建模：应用 Monte Carlo 仿真，实物期权分析，预测及优化》，第二版，Wiley 2010；*Valuing Employee Stock Options (2004 FAS 123R)*，Wiley Finance 2004。）更多相关信息请访问我们的网站：[www.realoptionsvaluation.com.cn](http://www.realoptionsvaluation.com.cn)。

Risk Simulator 软件包含以下几个模块：

- Monte Carlo 仿真模块（利用不同的仿真文档、截距和相关性仿真、正态分布、控制置信度和误差的仿真以及其它一些运算法则对 42 种概率分布进行参数和非参数仿真）
- 预测模块（进行 Box-Jenkins ARIMA、多元回归、非线性推断、随机过程和时间序列分析）
- 不确定情况下的优化模块（利用离散整数和连续变量对有仿真和无仿真的有价证券和项目进行优化）
- 建模和分析工具模块（进行飓风图、蜘蛛图和敏感性分析，同时还有内部抽样仿真，假设检验和分布拟合等）

**Real Options SLS** 软件可以用于计算简单和复杂的期权，同时具备创造自定义期权模型的功能。**Real Options SLS** 软件包括以下几个模块：

- 单资产 SLS（用于解决放弃期权，选择期权，延迟期权，扩展期权以及自定义期权）
- 多重资产和多阶段 SLS（用于解决多阶段连续期权，包含多重资产和多个阶段，多阶段期权与放弃、选择、收缩、延迟、扩展及转换等期权的结合，也可以解决自定义的期权问题）
- 多叉 SLS（用于解决三叉均值回复期权，四叉跳跃-扩散期权以及五叉彩虹期权）

- Excel 加载宏功能（在 Excel 环境下解决上述所有期权模型及闭合模型和自定义期权模型）

## 1.2 安装和授权过程

跟随屏幕上的指示安装本软件。本软件的最低要求如下：

- Pentium IV 处理器及以上
- Windows XP/Vista/Windows 7
- Windows Excel XP, 2003, 2007, 2010 版及以上
- Microsoft .NET Framework 2.0/3.0
- 500M 空间
- 推荐 2GB 内存
- 安装软件的管理员权限

现在大多数新电脑都已经自带安装了 Microsoft .NET Framework 2.0/3.0 组件。但是，如果在安

## 1.3 授权过程

装 Risk Simulator 的过程中出现错误提示，要求安装 .NET Framework，那就要首先退出安装程序，然后从 CD 里（可以自己选择语言）安装相关的 .NET Framework 组件。完成之后重启电脑，再重新安装 Risk Simulator 软件。

我们的软件有一个 30 天的试用期。如果想到得到完整的公司授权，请联系 Real Option Valuation 公司邮箱：[admin@real-consulting.com](mailto:admin@real-consulting.com) or [admin@realoptionsvaluation.com](mailto:admin@realoptionsvaluation.com) 或者拨打电话（925）271-4438，还可以访问我们的网站：[www.realoptionsvaluation.com.cn](http://www.realoptionsvaluation.com.cn)。

如果您已安装本软件，并购买了使用该软件的授权，您需要将您的硬盘码用电子邮件的方式发送给我们，我们将会为您生成所需的授权文件。具体可参照如下指导：

Windows XP 操作系统，使用 Excel XP, Excel 2003, Excel 2007 和 Excel 2010：  
首先，在 Excel 中单击仿真|授权，[并将得到的包括字母和数字的 11-20 位硬盘码发送 Email 到 admin@real-consulting.com or admin@realoptionsvaluation.com](#)。一旦收到您的硬盘码，我们将会为您生成一个永久使用软件的授权文件，并将该授权文件发送 Email 给您。当您得到该授权文件后，下载存放到您的硬盘上，然后运行 Excel，点击仿真|授权，并单击**安装授权**，载入授权文件。然后重新启动 Excel 就完成了。整个过程只需短短的几分钟，您就可以成为授权用户来使用该软件了。

Windows Vista/Windows 7 使用 Excel XP, Excel 2003 或 Excel 2007 和 Excel 2010：  
首先，在 Windows Vista/Windows 7 中打开 Excel2007/2010，并选中仿真菜单栏，单击授权图标或仿真|授权，[将得到的包括字母和数字的 11-20 位硬盘码发送 Email 到 admin@real-consulting.com or admin@realoptionsvaluation.com](#)。一旦收到您的硬盘码，我们将会为您生成一个永久使用软件的授权文件，并将该授权文件发送 Email 给您。当您得到该授权文件后，下载存放到您的硬盘上。运行 Excel，点击仿真|授权或授权图标，并单击**安装授权**，载入授权文件。

然后重新启动 Excel 就完成了。整个过程只需短短的几分钟，您就可以成为授权用户来使用该软件了。

安装完成之后，开启 Microsoft Excel，如果安装成功的话，会看到 Excel 的菜单列中会出现一项额外的“仿真”项，还有一个新的图标，如下图 1.1。除此以外，还会出现下图 1.2 中的程序启动画面，表示软件被激活并加载到 Excel 中。图 1.3 是 Risk Simulator 的工具列。如果可以在 Excel 中找到以下项目，那您现在就可以开始使用本软件了。接下来的章节将会一步步地介绍如何使用本软件。

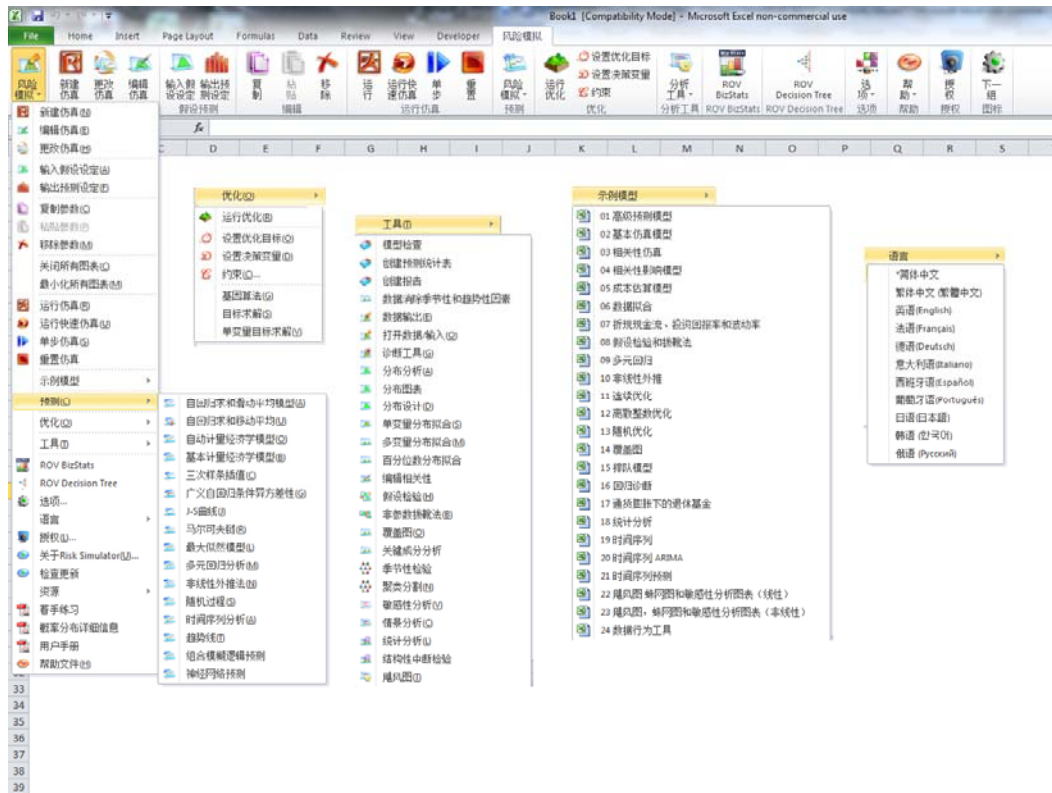


图 1.1 ——Risk Simulator 菜单和图标





图 1.2 ——Risk Simulator 启动画面

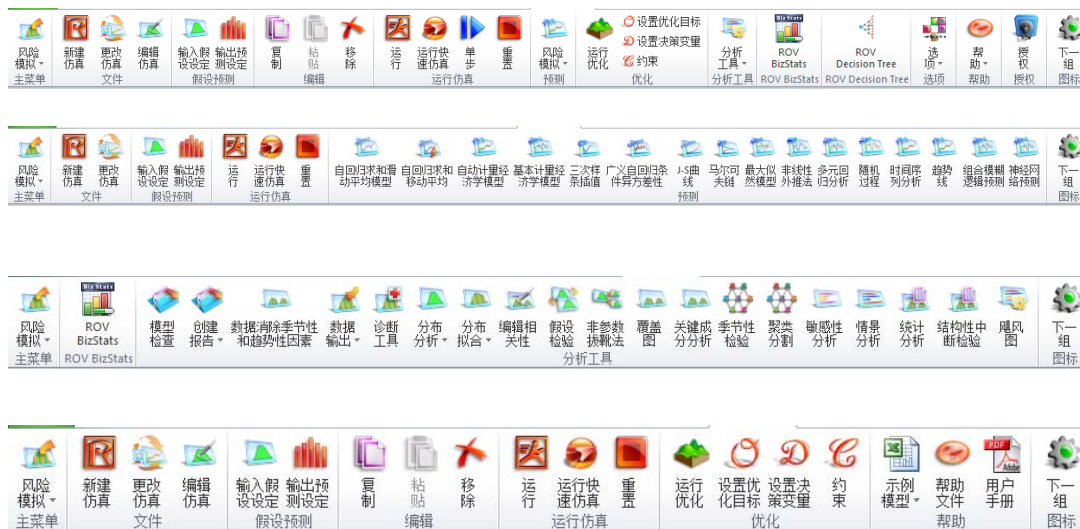


图 1.3 ——Risk Simulator 工具栏

## 1.4 2011/2012 版有哪些新增功能?

### Risk Simulator 功能完整清单

以下清单列出了 Risk Simulator 的主要功能，强调显示的部分是 Risk Simulator 2011/2012 版增的最新功能。

#### 1.4.1 一般功能

1. 支持 11 种语言 - 英语，法语，德语，意大利语，日语，**韩语**，葡萄牙语，西班牙语，简体中文，**俄语**和**繁体中文**。
2. ROV 决策树软件用于创建和评估决策树模型。高级方法和分析包括：
  - 决策树模型
  - 蒙特拉罗仿真
  - 敏感性分析
  - 情景分析
  - 贝叶斯定理(连接及后可能性更新)
  - 期望数值信息
  - 极大极小
  - 极大极小
  - 风险文件
3. 书籍 - 由 10 本书籍提供分析理论，应用和案例研究支持。
4. 批注单元格 - 打开或关闭单元格的批注和决定你是否想在输入假设，输出预测和决策变量的单元格显示批注。
5. 详细的示例模型 - Risk Simulator 里面的 24 种示例模型和 Modeling Toolkit 里面的超过 300 个模型。
6. 详细报告 - 所有的分析包含都详细报告。
7. 详细的用户手册 - 分步介绍的用户手册。
8. 灵活的授权方式 - 你可以打开或关闭的软件的某些功能，能够让你定制你的风险分析经验。例如，如果你只对 Risk Simulator 的预测工具感兴趣，您可以获取一个特定的授权来激活预测工具模块而不激活其它的模块从而节省一些软件费用。
9. 灵活的软件需求 - 适用于 Window 7, Vista 和 XP 系统；与 Excel 2010, 2007, 2003 和 Mac 操作系统上运行的虚拟机。
10. 完全可定制的颜色和图表 - 倾斜，立体，颜色，图表类型，以及更多！
11. 动手练习 - 详细 Risk Simulator 的分步使用指南，包括解释结果的指南。
12. 多个单元格复制和粘贴 - 允许对输入假设，决策变量，预测进行复制和粘贴。
13. 在一个模型中创建多个仿真文档 - 允许在一个模型中创建多个仿真文档（在一个模型中可以创建，复制，编辑和运行仿真文档）。
14. 在 Excel 2007/2010 中修订了的图标 -- 完全重新设计的图标工具栏，更加直观和用户友好。将有四套图标来匹配大部分的屏幕分辨率（1280 × 760 及以上）。
15. 右键快捷键 - 可以右击鼠标来进入 Risk Simulator 的工具和菜单。
16. ROV 软件集成 - 与其他 ROV 软件兼容的软件包括，包括 Real Options

SLS, Modeling Toolkit, Basel Toolkit, ROV Compiler, ROV Extractor 和 Evaluator, ROV Modeler, ROV Valuator, ROV Optimizer, ROV Dashboard, ESO Valuation Toolkit, 以及更多!

17. Excel 中的 RS 函数功能 – 在 Excel 中插入 RS 功能来进行输入假设和预测设置, 并支持用鼠标右键单击应用。
18. 疑难解答: 此工具允许您重新启用该软件, 检查您的系统要求, 获取硬件 ID, 等等。
19. 高速分析: 这种新的功能能够高速运行预测和其它分析 (在版本 5.2 基础上提升)。分析和结果与就版本仍是相同的, 但现在计算和报表生成的速度更快。
20. 网络资源, 案例研究和视频 – 免费下载模式, 开始使用视频, 案例研究, 白皮书, 和我们网站的其它材料。

### 1.4.2 仿真模块

21. 6 种随机生成器 – ROV 高级减法生成器, 减法随机打乱生成器, 长期打乱生成器, 可转移随机打乱生成器, 快速 IEEE Hex 打乱生成器, 基本最小可转移生成器。
22. 两种抽样方法 - 蒙特卡罗和拉丁超立方。
23. 3 种相关性 Copula 函数 - 应用正态度 Copula 函数, T Copula 函数, 和相关仿真的拟正态 Copula 函数。
24. 42 种概率分布 - 反正弦, 伯努利, Beta, Beta 3, Beta 4, 二项, 柯西, 卡方, 余弦, 自定义, 离散均匀, 双对数, Erlang, 指数, 指数 2, F 分布, 伽玛, 几何, Gumbel 分布最大, 最小 Gumbel 分布, 超几何, 拉普拉斯 (Laplace), 逻辑, 对数正态分布 (算术) 和对数正态分布 (对数), 对数正态分布 3 (算术) 和对数正态分布 3 (逻辑), 负二项, 正态, 抛物, 帕累托, 帕斯卡尔, 皮尔逊 5, 皮尔森 6, PERT 法, 泊松, 功率, 功率 3, 瑞利 (Rayleigh), T 和 T2, 三角, 均匀, 韦伯, 韦伯 3。
25. 替代参数 – 使用输入百分位作为输入参数的替代方法。
26. 自定义非参数分布 – 定制自己的分布, 运行历史仿真, 和运用德尔菲法。
27. 分布截断 – 设定数据的界限。
28. Excel 函数 – 使用 Excel 函数来设定假设和预测。
29. 多维仿真 - 对不确定性输入参数的仿真。
30. 精密控制 - 确定模拟试验次数是否足够。
31. 超高速模拟- 在几秒钟内运行 100,000 次仿真。

### 1.4.3 预测模块

32. ARIMA 模型 - 自回归整合移动平均模型 ARIMA (P, D, Q)。
33. 自动 ARIMA 模型 - 运行 ARIMA 模型最常见的组合, 找到最佳的拟合模型。
34. 自动计量经济学 - 运行数以千计的模型组合及排列获取对现存数据拟合最佳的模型 (线性, 非线性, 相互作用, 滞后, 超前, 按速率变化, 差异)。
35. 基本计量经济学 - 计量经济学和线性/非线性和相互作用回归模型。

36. 三次样条 - 非线性插值和外推。
37. GARCH 模型 - 适用广义自回归条件异方差模型来预测波动率：GARCH, GARCH-M, TGARCH, TGARCH - M, EGARCH, EGARCH-T, GJR - GARCH, 以及 GJR - TGARCH。
38. J 曲线 - 指数 J 曲线。
39. 有限相依变量 - Logit 模型, Probit, 与 Tobit。
40. 马尔可夫链 - 随着时间推移的两竞争因素和市场占有率预测。
41. 多元回归 - 常规线性和非线性回归, 与逐步方法 (向前, 向后, 相关性, 向前-向后)。
42. 非线性外推 - 非线性时间序列预测。
43. S 曲线 - 逻辑 S 曲线。
44. 时间序列分析 - 8 个时间序列分解模型, 预测基数, 趋势和季节性型。
45. 趋势线 - 使用线性, 非线性多项式, 指数, 对数, 指数和移动平均法进行预测和拟合, 并有拟合优度。
46. 神经网络预测 (线性, 逻辑, 双曲正切, 双曲正切的余弦)。
47. 组合模糊逻辑预测。

#### 1.4.4 优化模块

48. 线性优化 - 多相优化和一般线性优化。
49. 非线性优化 - 详细的结果包括 Hessian 矩阵, 拉格朗日函数和更多。
50. 静态优化 - 快速运行连续, 整数离散和二元制优化。
51. 动态优化 - 仿真与优化。
52. 随机优化 - 二次, 切, 中, 向前, 收敛标准。
53. 有效前沿 - 多元有效前沿的随机和动态优化组合。
54. 遗传算法 - 适用于各种优化问题。
55. 多相优化 - 局部与全局优化测试, 从而更好地控制优化如何运行, 提高了结果的准确性和相互关系。
56. 百分位和条件均值 - 随机优化的其它统计信息, 包括百分位以及条件均值, 这对计算在险价值很关键。
57. 搜索算法 - 对于单个决策变量和目标搜索应用是一种简单, 快速, 有效的算法。
58. 动态和随机优化中的超高速仿真 - 与优化结合的超高速仿真。

#### 1.4.5 分析模块

59. 模型检查 - 测试你的模型中最常见的错误。
60. 相关性编辑器 - 允许直接输入和编辑大量相关性矩阵。
61. 创建报告 - 自动生成模型里面的假设和预测报告。
62. 建立统计量报告 - 生成所有预测统计量比较报告。
63. 数据诊断 - 运行异方差, 微数缺测性, 离群, 非线性, 自相关, 正态, 球形, 非平稳性, 多重共线性和相关性测试。
64. 数据提取和导出 - 提取数据到 Excel 或文本文件和 RiskSim 文件, 运行统计报告和预测结果报告。

65. 数据打开和导入 – 重新检索以前的仿真运行结果。
66. **去季节化和和去趋势化** – 对您的数据去季节化和和去趋势化处理。
67. 分布分析 – 对 42 种分布计算精确的 PDF, CDF 和 ICDF, 并生成概率分布表。
68. 分布设计器 - 创建自己的定制分布。
69. 分布拟合 (多个变量) - 同时运行多个变量的分布拟合并考虑相关性和相关性的影响。
70. 分布拟合 (单个变量) - 对连续分布进行柯尔莫哥洛夫, 斯米尔诺夫和卡方检验, 生成完整的报告和分布假设。
71. 假设检验 – 测试两个统计量是否统计显著相同或者有差异。
72. 非参数拔靴法 - 对统计量进行仿真, 以获取结果的精度和准确性。
73. 覆盖图 – 对假设和预测生成完全定制的覆盖图 (CDF, PDF, 2D/3D 图表类型)。
74. **主成分分析** - 测试最佳预测变量和方法, 以减少数据的数组。
75. 情景分析 – 可以生成数以万计的静态二维情景。
76. 季节性测试 - 测试各种季节性滞后。
77. 分割聚类 – 分割你的数据, 将数据集纳入不同的统计类别。
78. 敏感性分析 - 动态敏感性测试 (同时进行分析)。
79. **结构突变测试** - 测试你的时间序列数据是否有统计结构突变。
80. 龙卷风分析 - 敏感性的静态扰动, 蜘蛛网分析和龙卷风分析, 情景分析表格。

#### 1.4.6 统计学和 BizStats 模块

81. **百分位分布拟合** – 使用百分位和优化, 以找到最佳拟合分布。
82. **概率分布'图表和表格** - 运行 45 种概率分布, 计算它们的四个矩, CDF, ICDF, PDF, 图表, 叠加多个分布图, 并生成概率分布表。
83. 统计分析 - 描述性统计, 分布拟合, 直方图, 图表, 非线性外推, 正态性测试, 随机参数估算, 时间序列预测, 趋势预测等。
84. **ROV BIZSTATS** - 超过 130 种商业统计和分析模型:

Absolute Values, ANOVA: Randomized Blocks Multiple Treatments, ANOVA: Single Factor Multiple Treatments, ANOVA: Two Way Analysis, ARIMA, Auto ARIMA, Autocorrelation and Partial Autocorrelation, Autoeconometrics (Detailed), Autoeconometrics (Quick), Average, Combinatorial Fuzzy Logic Forecasting, Control Chart: C, Control Chart: NP, Control Chart: P, Control Chart: R, Control Chart: U, Control Chart: X, Control Chart: XMR, Correlation, Correlation (Linear, Nonlinear), Count, Covariance, Cubic Spline, Custom Econometric Model, Data Descriptive Statistics, Deseasonalize, Difference, Distributional Fitting, Exponential J Curve, GARCH, Heteroskedasticity, Lag, Lead, Limited Dependent Variables (Logit), Limited Dependent Variables (Probit), Limited Dependent Variables (Tobit), Linear Interpolation, Linear Regression, LN, Log, Logistic S Curve, Markov Chain, Max, Median, Min, Mode, Neural Network, Nonlinear Regression, Nonparametric: Chi-Square Goodness of Fit, Nonparametric: Chi-Square Independence,

## Risk Simulator

Nonparametric: Chi-Square Population Variance, Nonparametric: Friedman's Test, Nonparametric: Kruskal-Wallis Test, Nonparametric: Lilliefors Test, Nonparametric: Runs Test, Nonparametric: Wilcoxon Signed-Rank (One Var), Nonparametric: Wilcoxon Signed-Rank (Two Var) , Parametric: One Variable (T) Mean , Parametric: One Variable (Z) Mean , Parametric: One Variable (Z) Proportion , Parametric: Two Variable (F) Variances , Parametric: Two Variable (T) Dependent Means , Parametric: Two Variable (T) Independent Equal Variance , Parametric: Two Variable (T) Independent Unequal Variance , Parametric: Two Variable (Z) Independent Means , Parametric: Two Variable (Z) Independent Proportions , Power, Principal Component Analysis, Rank Ascending, Rank Descending, Relative LN Returns, Relative Returns, Seasonality, Segmentation Clustering, Semi-Standard Deviation (Lower), Semi-Standard Deviation (Upper), Standard 2D Area, Standard 2D Bar, Standard 2D Line, Standard 2D Point, Standard 2D Scatter, Standard 3D Area, Standard 3D Bar, Standard 3D Line, Standard 3D Point, Standard 3D Scatter, Standard Deviation (Population), Standard Deviation (Sample), Stepwise Regression (Backward), Stepwise Regression (Correlation), Stepwise Regression (Forward), Stepwise Regression (Forward-Backward), Stochastic Processes (Exponential Brownian Motion), Stochastic Processes (Geometric Brownian Motion), Stochastic Processes (Jump Diffusion), Stochastic Processes (Mean Reversion with Jump Diffusion), Stochastic Processes (Mean Reversion), Structural Break, Sum, Time-Series Analysis (Auto), Time-Series Analysis (Double Exponential Smoothing), Time-Series Analysis (Double Moving Average), Time-Series Analysis (Holt-Winter's Additive), Time-Series Analysis (Holt-Winter's Multiplicative), Time-Series Analysis (Seasonal Additive), Time-Series Analysis (Seasonal Multiplicative), Time-Series Analysis (Single Exponential Smoothing), Time-Series Analysis (Single Moving Average), Trend Line (Difference Detrended), Trend Line (Exponential Detrended), Trend Line (Exponential), Trend Line (Linear Detrended), Trend Line (Linear), Trend Line (Logarithmic Detrended), Trend Line (Logarithmic), Trend Line (Moving Average Detrended), Trend Line (Moving Average), Trend Line (Polynomial Detrended), Trend Line (Polynomial), Trend Line (Power Detrended), Trend Line (Power), Trend Line (Rate Detrended), Trend Line (Static Mean Detrended), Trend Line (Static Median Detrended), Variance (Population), Variance (Sample), Volatility: EGARCH, Volatility: EGARCH-T, Volatility: GARCH, Volatility: GARCH-M, Volatility: GJR GARCH, Volatility: GJR TGARCH, Volatility: Log Returns Approach, Volatility: TGARCH, Volatility: TGARCH-M, Yield Curve (Bliss), and Yield Curve (Nelson-Siegel).

## 2

## 2. Monte Carlo 仿真

以著名的摩纳哥赌城命名的 Monte Carlo 仿真是一种非常有效的方法。仿真可以轻而易举地为从业者解决那些困难复杂的现实问题。Monte Carlo 通过生成成百上千种甚至上百万种试验路径的结果来模拟和分析未来可能发生的事件，并观察它们的普遍共性。作为公司的分析师，仅仅学习大学水平的数学课程是不合乎实际的。一名优秀的分析师会尽可能地利用其所有可利用的工具，用最简单和最实用的方法来得到相同的答案。在所有案例之中，只要模型正确，Monte Carlo 仿真与其它数学上优美的方法一样，也可以提供近似的答案。现在我们的问题是什么是 Monte Carlo 仿真以及它是如何运作的？

### 2.1 什么是 Monte Carlo 仿真？

简单来说 Monte Carlo 仿真就是一个用于预测、估计和风险分析的随机数发生器。仿真就是通过反复地从一个事先确定的变量的概率分布中挑选出一些数值，并将这些数值用于模型来计算一个模型的无数种情景。由于所有这些情景在模型中产生关联的结果，每个情景可以产生一种预测。预测通常是包含公式或者函数的事件，它被定义为模型的一个重要输出量。这些事件通常是诸如总量、净利润和总支出等。

可以简单地将 Monte Carlo 仿真方法想象成从一个大篮子里取出高尔夫球然后放回的重复过程。篮子的大小和形状取决于**输入假设**的分布(例如，一个均值为 100，标准差为 10 的正态分布相比于均匀分布或三角分布)，那些有深度并且比较对称的篮子会令到某些球被取出的频数高于其它球，重复取出球的次数取决于仿真试验的次数。对于一个很大，有着多重相关性的模型，将它想象成一个超级大的篮子，里面有很多小篮子。每一个篮子里面都有一堆高尔夫球跳来跳去。有时候这些小篮子会手挽着手(在这些变量之间存在相关性的情况下)，这时有些高尔夫球会一前一后跳动，其它一些会彼此独立跳动。在存在这些交互作用的情况下，将每次取出的球的编号输入表格，这就提供了一份仿真的预测输出结果。

## 2.2 开始学习 Risk Simulator

### 2.2.1 软件的总体介绍

Risk Simulator 软件包含几种不同的应用模块，具体有 Monte Carlo 仿真，预测以及优化。

- 仿真应用模块允许您在基于 Excel 的模型中进行仿真，生成仿真预测（结果的分布），进行分布拟合（自动寻找最优的统计分布），计算相关性（变量之间的关系），确定敏感度（制作飓风图和敏感度图表），同时还可以运行自定义和非参数仿真（利用历史数据，但不需指定分布或者参数的仿真）。
- 预测应用模块可以用于生成 Box-Jenkins ARIMA 计量经济预测，自动时间序列预测（包含季节性和趋势），多元回归（线性和非线性回归），非线性外推（曲线拟合），随机过程（随机游走，均值回归以及跳跃扩散过程）。
- 最优化应用模块用于多元离散整数变量、连续变量和混合决定变量，在某个约束条件下，最大化或最小化某个目标。并且可以和 Monte Carlo 仿真一起用于静态最优化或动态最优化，以及在不确定情况下的随机最优化，还可以用于解决线性和非线性最优化问题。
- Real Options SLS 是与 Risk Simulator 软件相独立又相互补充的一款软件，可用于解决多种简单或复杂的实物期权问题。

### 2.2.2 运行 Monte Carlo 仿真

一般情况下，为了在已有 Excel 模型下运行仿真，需要遵循以下步骤：

1. 新建一个仿真文档或打开一个现有文档
2. 在相关的单元格定义输入假设
3. 在相关的单元格定义输出假设
4. 运行仿真
5. 结果解析
6. 如果想要试验一下，可以打开名叫“基本仿真模型”的示例文件，然后跟随下面的例子

来进行仿真。这个示例文件可以在开始菜单的目录下找到：开始|Real Options Valuation|Risk Simulator|示例。

#### 新建一个仿真文档

为了开始一次新的仿真，您必须首先创建一个仿真文档。这个仿真文档包含一整套关于

如何进行一次仿真的指导说明，也就是所有的假设、预测、运行选项等等。这个文件简化了创建多个仿真的过程。也就是说利用一个相同的模型可以创建几个文档，每个都对应自己的仿真特性和要求。同一个人可以利用不同的分布假设和输入量来创建不同的测试情景，多个人也可以在同一个模型下根据他们自己的假设量和输入量进行测试。

- 打开 Excel 表格，创建一个新的或打开一个现有的模型（您可以使用基本



仿真模型的例子来继续)

- 点击**仿真**图标，选择**新建仿真**
- 为仿真设定一个标题，并填入其它相关信息（图 2.1）
- 

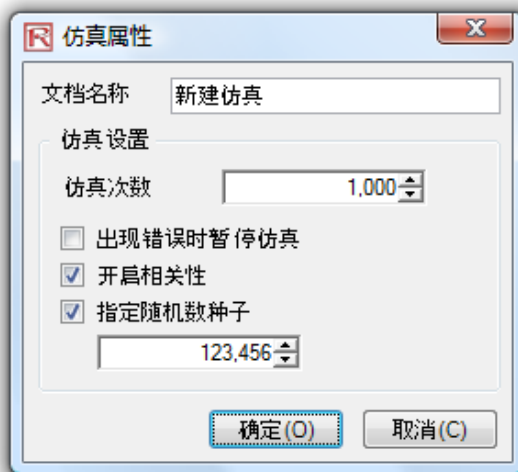


图 2.1——新建仿真文档

- 标题:** 对仿真文档进行命名，可以在同一个 Excel 模型中创建多个仿真文档。这意味着可以在不取消现存假设和不因新仿真情景的需要而改变假设的情况下，在同一模型中保存不同的仿真情景的文档。之后您也可以任意改变文档名（**仿真|编辑仿真**）。
- 试验次数:** 这里要求输入的是需要仿真试验的次数。进行 1000 次试验意味着会产生基于输入假设的 1000 个不同的结果。可以根据需要改变试验次数，但是键入的数值必须是正整数，系统默认的运行次数是 1000 次。同时可以利用置信度和误差控制来自动帮助决定需要进行仿真的次数（更多详情请参考**置信度和误差控制**章节）。
- 仿真错误时停止:** 如果选中此项，那么在 Excel 模型中每次出现错误时仿真就会停止。也就是说，如果您的模型遇到一个计算错误（例如，在某次仿真试验中可能会有某张电子数据出现输入值除以零的错误），仿真就会停止。这点对于审核您的模型以确保在 Excel 模型中没有任何计算错误很有帮助。当然，如果您对自己的模型有足够的信心，那么也就没有必要选这一项了。
- 开启相关性:** 如选中这一选项，就会考虑输入假设之间的相关性。否则，相关性就会被设置为零，仿真会在输入假设间不存在相关性的假设下进行。举例来说，如果存在正相关的话应用相关性会得到更精确的结果，但如果存在负相关性的话，得到的预测可信度就会下降。如果在此处开启相关性的话，稍后您可以在生成的每个假设处设置相应的相关系数（更多详情请参见**相关性**章节）。
- 设置随机数种子:** 每次仿真所得到的结果都会有细微的差别。这是由 Monte Carlo 仿真中随机数据生成方法的特点决定的，这也是所有随机数据生成器的理论事实。但是在作报告的时候，可能需要相同的结果（尤其是当报告要展示一系列结果时，在报告过程中想要展示相同的生成结果；或是当与他人采用的是相同的模型时，希望每次都得到相同的结果），那么就选中这一选项，然后设置一个种子值。这个种子值可以是任意正整数。使用相同的初始随机数种子，相同的试验次数，相同的输入假设，进行仿真后，通常会生成相同序列的随机数，以保证最终结果的相同。

注意一旦创建一个新的仿真文档，您可以在任何时候返回去修改这些设置。但是要保证当前文件是您想要修改的文档。否则，点击**仿真|更改仿真**，选择您想要修改的那个文档，点击**确定**（下图 2.2 有一个例子显示当前有几个文档，如何激活想要修改的文档）。然后点击**仿真|编辑仿真**，作相应的修改。同时还可以复制或重命名现存的文档。

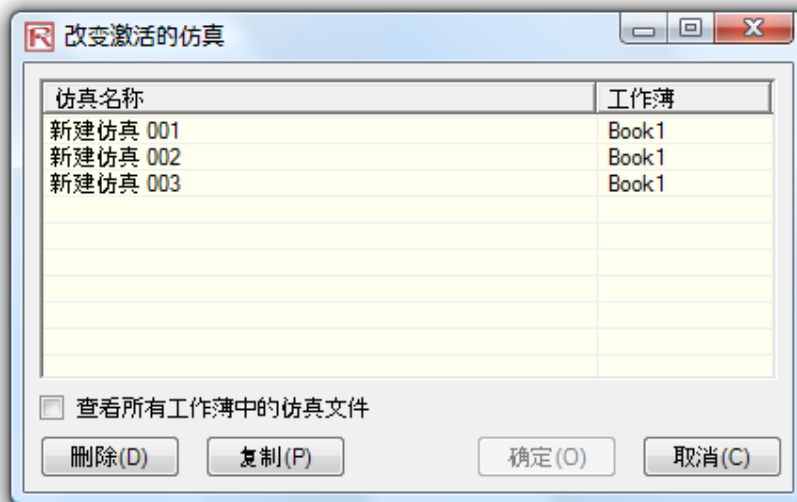


图 2.2 ——修改当前仿真

## 定义输入假设

接下来这一步是在您的模型中设置输入假设，注意假设不能以任何等式或函数的形式分配到单元格中。例如，在模型中只能对数值单元格设定输入假设，反之只能对含有等式或函数的单元格设定输出预测。记住只有在仿真文件已经存在的条件下才可以设置假设和预测。按照下列步骤来设置模型中新的输入假设：

- 确保仿真文档已经存在，您可以打开一个现有文档，或是新建一个文档（仿真|新建仿真）
- 选择您想要设置输入假设的单元格（例如，在基本仿真模型例子中的 G8 单元格）
- 点击仿真|输入假设设定或是点击 Risk Simulator 工具栏上的第三个图表
- 选择想要的相关分布，并键入相关的分布参数，点击确定将这些输入假设插入到模型中（图 2.3）

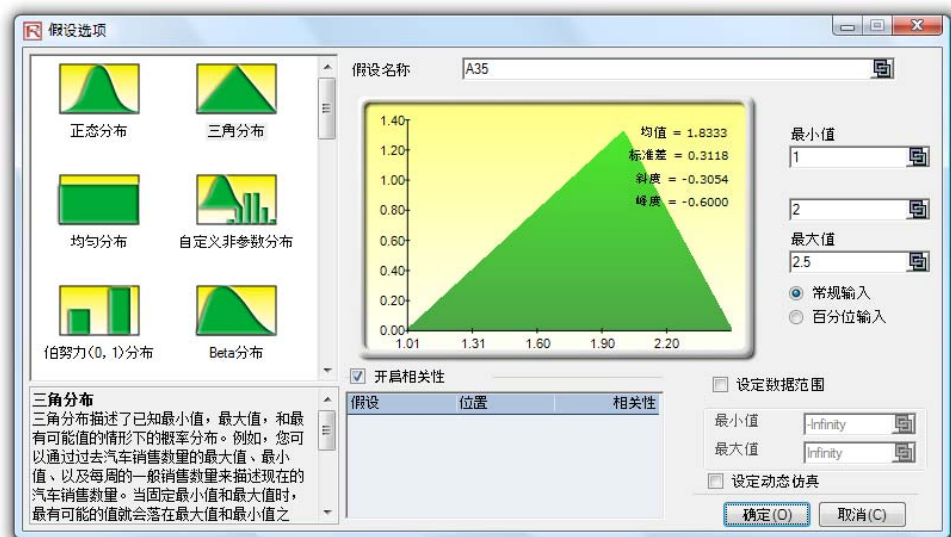


图 2.3 设定输入假设

注意，在假设属性中有几个需要留意的地方。下图 2.4 指出了这些地方：

假设名称	<ul style="list-style-type: none"> <li>这块可选择的功能允许您为假设输入一个独特的名称，以便区分不同的假设，好的模型通常使用精简的假设名称。</li> </ul>
分布图库	<p>左边这块区域列示了软件中所有可用的分布。可以通过右击图库的任何地方，选择大图表，小图表或列表来变换视图，有超过 24 种的分布可以使用。</p>
输入参数	<p>基于选定的分布来确定所需的参数。您既可以直接输入参数，也可以链接到具体的工作表中。在输入参数需要可见或是允许改变（点击链接图标将输入参数与工作表单元格链接）的情况下，链接到工作表单元格比较适用。</p>
分布边界	<p>一般情况下分析师是不会使用的，它们只是用于缩小分布假设的范围。例如，如果选择正态分布，那么它的理论边界就是正无穷和负无穷之间。然而在实际情况中，仿真变量只在一个小范围内存在，这个范围可以用来适当缩减分布范围。</p>
相关性	<p>可以将输入假设设置为一对相关量。如果假设需要的话，记住点击<b>仿真 编辑仿真</b>来开启相关性选择。更多详情请参考本章末的关于相关性设置和其对模型影响的讨论。</p>
简短描述	<p>图库里的每个分布都有一个简短描述。它说明了何时使用该分布，以及输入参数的要求。参见<a href="#">了解 Monte Carlo 仿真中的概率分布</a>章节关于软件中可用的分布类型的详细介绍。</p>

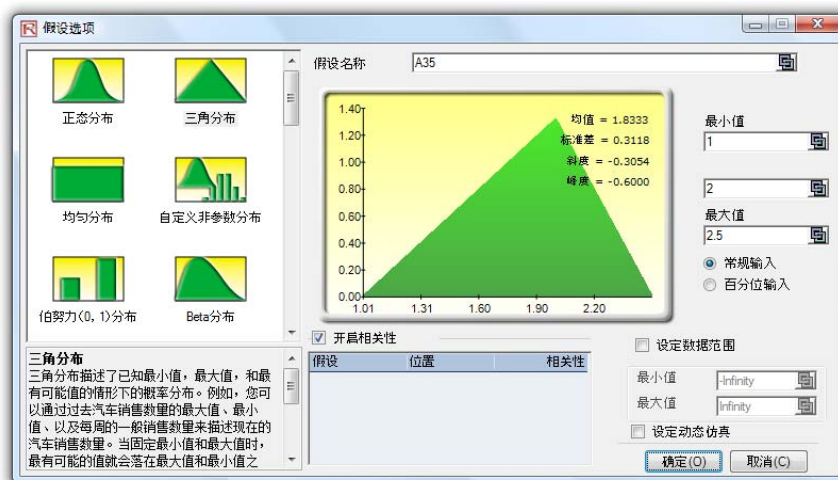


图 2.4 假设属性

注意：如果继续本例的话，在 G9 单元格中设定其它的假设。使用最小值为 0.9，最大值为 1.1 的均匀分布。然后，在下一个步骤中定义输出预测。

## 定义输出预测

下一步是在模型中定义输出预测，输出预测只能设置在含有等式或函数的单元格上。以下是设定预测的程序：

- 选择想要设置假设的单元格（例如，基本仿真模型例中的单元格 G10）
- 点击**仿真**，选择**输出预测设定**或是点击 Risk Simulator 工具栏上的第四个图标（图 1.3）
- 键入相关的信息按**确认**键。

下图 2.5 说明了预测属性的设置

- **预测名称**：定义预测单元格，这点非常重要，因为当您的模型很大并且包含很多预测单元格时，单独命名预测单元格有助于您迅速找到正确的结果。不要小瞧这一步骤的重要性，最好是使用精确简短的假设名称。
- **预测准确性**：您可以通过设置置信水平和误差控制而不是根据推测来估计所需的试验次数。当仿真达到所需的误差-置信水平结合的要求之后，仿真就会停止并告知您所达到的置信水平，通过这样的方式可以自动地获得试验次数，不需要我们去估计它的数值。更多详情请参见**误差和置信度控制**章节。
- **显示预测窗口**：允许使用者显示或不显示具体的预测窗口，系统的默认选项是显示预测窗口。

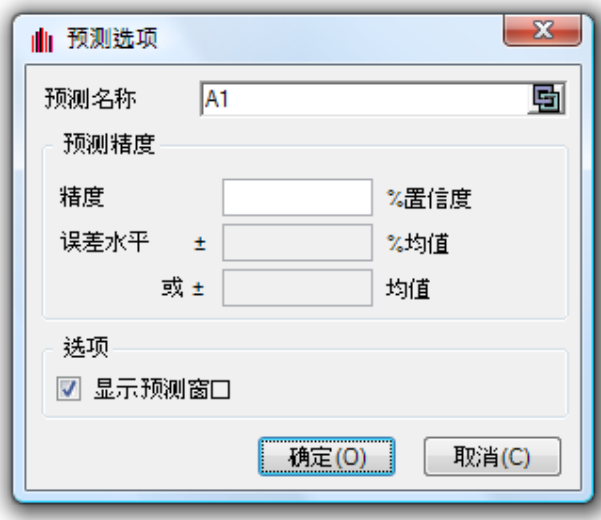


图 2.5 输出预测设定

### 运行仿真

如果一切准备就绪，就可以点击**仿真|运行仿真**或是点击**运行**图标（Risk Simulator 工具栏上的第八个图标）开始仿真。还可以在运行之后重新设置仿真再运行一次（**仿真|重置仿真**或是工具栏上的第十个图标），或是在运行过程中中止它。单步功能（**仿真|单步仿真**或是工具栏上的第九个图标）可以允许您进行一次仿真试验，一次一个，这对于初学者来说很有用（例如，可以在每次试验中，将所有假设单元格的值替换掉，并重新计算整个模型）

### 预测结果解析

Monte Carlo 仿真的最后一个步骤就是对预测结果图进行解释说明。图 2.6 到图 2.13 是运行仿真之后生成的预测图和统计表，以下是在解释仿真结果时需要说明的重要因素：

- **预测图：**图 2.6 中的预测图显示的是整个仿真试验中数值出现频数的概率直方图。竖条代表的是某个具体的 x 值在总试验次数中出现的次数，累积频数（一条光滑的曲线）代表的是不大于 x 值的所有数值在预测中出现的总概率。
- **预测统计量表：**图 2.7 中的预测统计量表根据分布的四矩对预测值的分布进行了总结。更多关于这些统计指标意义的解释请参见 [了解预测统计量](#) 章节。通过按空格键您可以在柱形图和统计量表之间进行切换。



图 2.6 预测图

Statistics	结果
仿真次数	1000
均值	0.9386
中位数	0.9320
标准差	0.1132
方差	0.0128
变异系数	0.1206
极大值	1.2358
极小值	0.6857
极差	0.5501
偏度	0.1212
峰度	-0.7517
25分位数	0.8486
75分位数	1.0262
95%置信度的百分误	0.7475%

图 2.7 预测统计表

- 用户设置：预测统计表里的用户设置标签可以让您改变图表的外观。例如，如果选择**总在最前端显示**，那么不管您的电脑里有什么别的软件在运行，预测图表都是可见的。直方图清晰度允许您改变直方图的竖条数量，从 5 到 100 之间的任意数均可。此外，**数据更新**板块允许您控制仿真的运行速度(相对于预测图表的更新频数)。也就是说，如果您想要每次试验时都更新预测图表，那就要牺牲仿真的速度，因为用于仿真的内存将被分配用于更新图表。这仅仅是一种用户选择，对仿真的结果不会有任何影响，只是影响了仿真的完成速度。为了进一步增加仿真的运行速度，您可以在仿真

过程中将 Excel 表格最小化，这样减少了更新 Excel 电子数据表所需的内存，并释放了运行仿真的内存。关闭所有窗口及最小化能控制所有打开的预测图表。

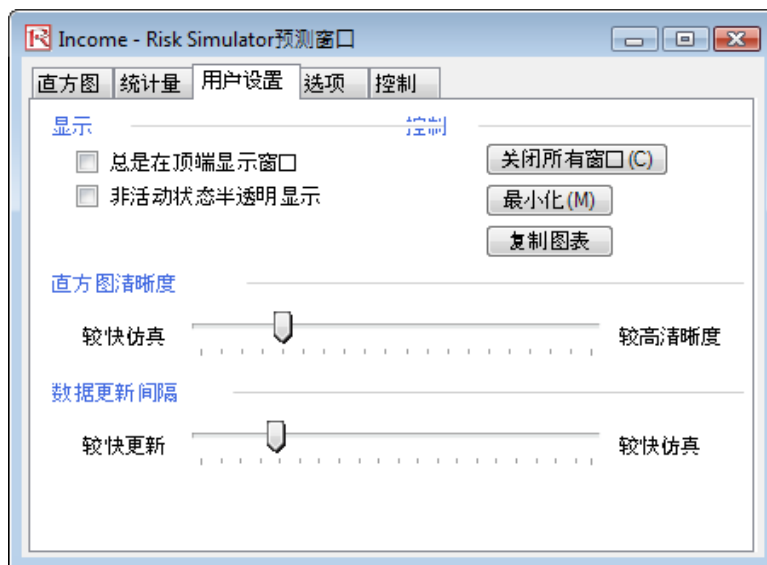


图 2.8 预测图参数选择

- 选项：预测图表里的选项功能允许您显示所有的预测数据或过滤在某一区间或某一标准差间隔内的输入/输出数据。此外，这里还可以为此预测设置置信度，以便从统计学的观点显示误差水平。更多详情参见 *误差和置信度控制* 章节，如果想要在预测图表中列出均值，中位数，第一分位线以及第四分位线（第 25 百分点和第 75 百分点），那么显示下面这些统计量是一个很好的选择。

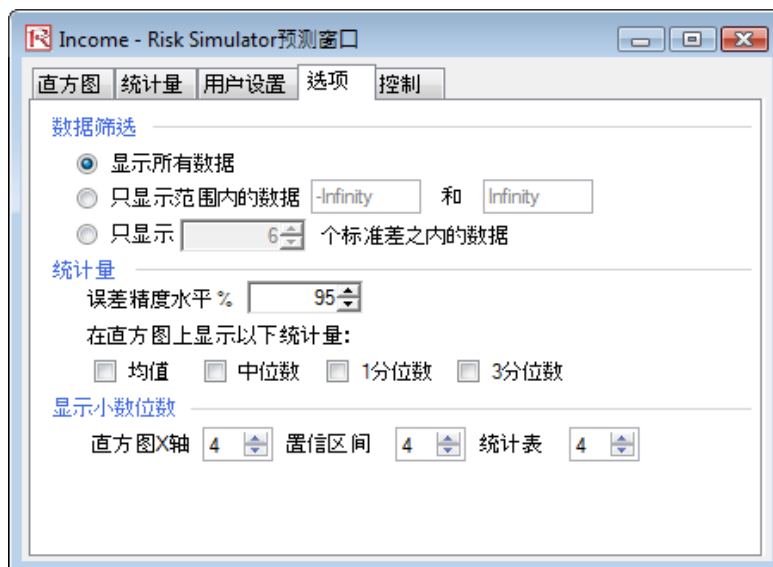


图 2.9 预测图表选项

## 使用预测图表和置信区间

在预测图表中，您可以确定事件发生的概率也就是通常所称的置信区间。给定两个值，仿真结果会落在这两个值中间区域的概率是多少？图 2.10 中可以看出最终结果（在本例中是收入水平）会落在\$0.7629 和\$1.1247 之间的概率是 90%。置信区间可以这样计算：首先类型选择双尾，然后键入所需的置信度（例如 90），按下键盘上的 TAB 键，对应此置信度的两个数值就被计算出来了。在本例中，收入低于\$0.7629 的概率是 5%，高于\$1.1247 的概率也是 5%。这意味着，双尾分布的置信区间是以中值或五十分位点数值为中心的一个对称区域。因此，双尾具有相同的概率分布。

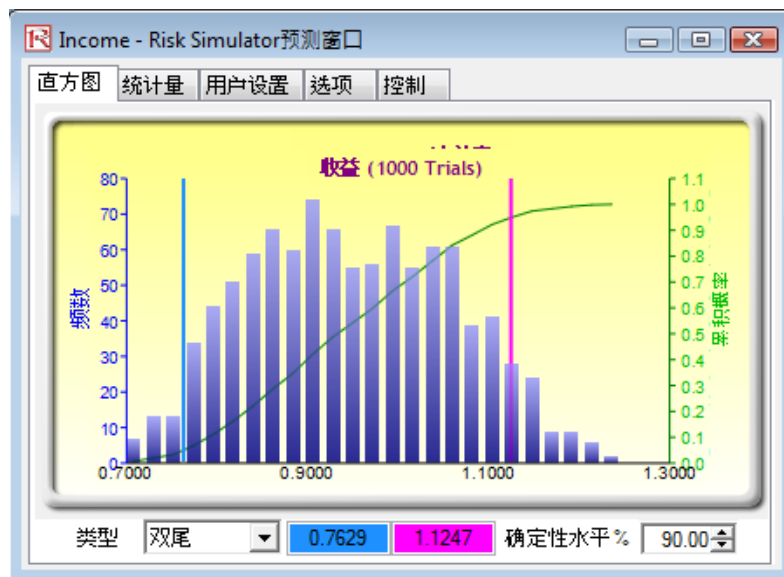


图 2.10 双尾置信区间预测图

作为选择，也可以使用单尾概率。图 2.11 是一个置信度 95% 的左尾概率分布（例如类型选择左尾，在置信度处键入 95，按下键盘上的 TAB 键）。这意味着收入低于\$1.1247 的概率是 95%，而高于\$1.1247 的概率是 5%，与图 2.10 中的情况很吻合。



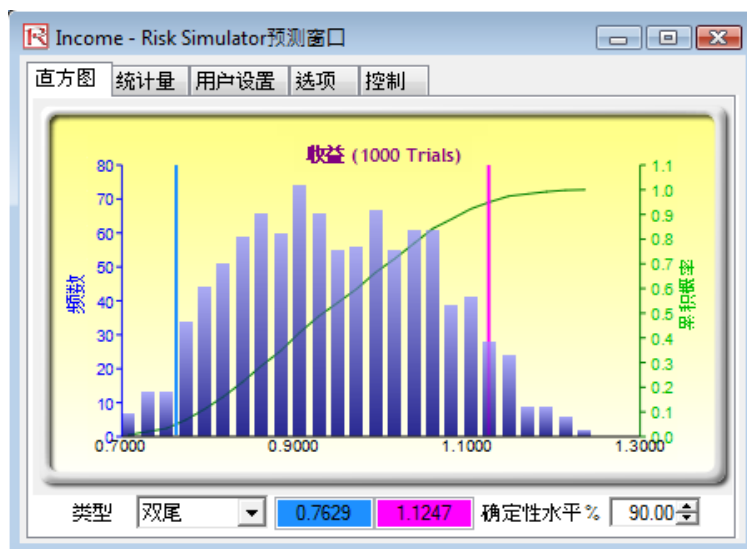


图 2.11 单尾置信区间预测图

除了可以估计预测区间（例如给出一个概率水平，确定相应的收入值），您还可以确定获得某个收益范围的概率。例如，收入小于\$1的概率是多少？为了计算这个数值，您可以在类型中选择左尾，在数值框中键入1然后敲TAB键。相应的置信度就会被计算出来（在本例中，收入低于\$1的概率为68.00%）。



图 2.12 预测图概率估计

为了完整起见，您可以在概率类型中选择右尾，在数值框中键入1然后敲TAB键。得出的结果是大于1的右尾概率，即收入大于\$1的概率（在本例中，我们可以看出此概率是32.30%）。



图 2.13 预测图概率估计

**小贴士：**可以通过拖动预测窗口的右下角来调整预测窗口的大小。通常比较明智的做法是在再次运行仿真之前，重置现有的仿真（**仿真|重置仿真**）。记住当需输入数值或左右尾数值时，敲击键盘上的 TAB 键来更新数据和结果。

## 2.3 相关性和精度控制

### 2.3.1 相关性的基础知识

相关系数测量的是两个变量之间相关性的强弱和方向，它可以是从-1.0 到+1.0 之间的任何数值。相关系数可以分解为符号（两变量是正相关还是负相关）和相关性的大小或强弱（相关系数的绝对值越大，两者之间的相关性越强）。

相关系数的计算方法有几种。第一种方法是利用以下的公式来手动计算变量 x 和 y 之间的相关系数 r:

$$r_{x,y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

第二种方法是使用 Excel 的 CORREL 函数。例如，如果 x 和 y 各有 10 个数据，分布在单元格 A1: B10 区域，可以直接使用函数 CORREL (A1: A10, B1: B10)。

第三种方法是运行 Risk Simulator 多元拟合工具来计算。

**重点注意：**相关并不意味着存在因果关系。两个完全无关的随机变量也可能呈现出一定的相关性，但是这并不意味着它们之间有任何的因果关系（例如，太阳黑子与股市有一定的相关性，但是它们之间并不存在因果关系）。

存在两种类型的相关：参数相关和非参数相关。Pearson 相关系数是最常用的相关量度标准，通常简称为相关系数。但是 Pearson 相关是一种参数测量方法，要求两个相关的变量都属于正态分布，并且变量之间的关系是线性的。但是在 Monte Carlo 仿真中，不符合这种条件的例子经常出现，此时非参数相关就显示出其重要性了。非参数相关主要有 Spearman Rank 相关和 Kendall's Tau 相关。

Spearman Rank 相关使用最普遍并且很适合应用于 Monte Carlo 仿真中——因为它不要求正态分布和线性关系，所以对不同分布变量的相关也适用。为了计算 Spearman Rank 相关，首先将变量 x 和 y 的值排序，然后再利用 Pearson 相关性来计算。

在 Risk Simulator 的例子中用到的相关大部分是非参数 Spearman Rank 相关。但是为了简化仿真过程并与 Excel 表中的相关函数保持一致，软件所要求的相关输入是 Pearson 相关系数。Risk Simulator 然后会利用自身的运算法则将它们转换成 Spearman Rank 相关，这样就简化了整个过程。但是，为了简化用户界面，我们允许使用者键入更常见的 Pearson 积差相关系数（例如，利用 Excel 表的相关函数计算），但在数学编码中，我们会将这些简单的相关转换成 Spearman Rank 相关再进行分布仿真。

### 2.3.2 在 Risk Simulator 中应用相关性

相关可以通过几种方式应用到 **Risk Simulator** 中：

- 在定义假设（**仿真|设置输入假设**）时，只需将相关系数填入分布图的相关矩阵网格中。
- 利用现有数据运行多元拟合工具（**仿真|工具|分布拟合|多元变量**）来进行分布拟合，得到成对变量之间的相关系数矩阵。如果仿真文件已经存在，那么对应的假设会自动包含这些相关系数。
- 存在假设之后，您可以在用户界面上直接点击**仿真|编辑相关性**来键入所有假设的相关系数。

注意，相关系数矩阵必须是正定矩阵。也就是相关系数在数学上必须是有效的。例如，假设您想要将以下三个变量相关联：毕业生某年的成绩，他们每周的啤酒消费量，他们每周学习的小时数。则必须假设存在以下的相关关系：

成绩和啤酒：负相关 喝的越多，成绩越低（不会在考试中表现出来）

成绩和学习时间：正相关 花在学习上的时间越多，成绩越高

啤酒和学习时间：负相关 喝的越多，花在学习上的时间越少（一直醉酒和聚会）

如果您将成绩和学习之间的关系定义为负相关，并且假设相关系数的数值很高，那么相关矩阵就是非正定的了。它不符合逻辑、相关性要求和数学上矩阵的一些常识。但是，即使逻辑性很差，小的相关系数有时还是存在的。当您键入非正定或错误的相关矩阵时，**Risk Simulator** 会自动提示您，并在维持原有相关关系整体结构的基础上为您提供调整后的半正定相关系数矩阵（同样的符号和大小）。

### 2.3.3 相关性对 Monte Carlo 仿真的影响

尽管在仿真中对相关系数变量进行的计算很复杂，但是结果却是很清晰的。图 2.14 显示的是一个简单相关性模型（示例文件夹中的**相关性影响模型**）。收入的计算很简单，用价格乘以数量。同样的模型将会同时应用于价格和数量之间不存在相关性，存在正相关（+0.9），存在负相关（-0.9）这三种情况进行比较。

相关模型

不相关	正相关	负相关	
价格	\$2.00	\$2.00	\$2.00
数量	1.00	1.00	1.00
收益	\$2.00	\$2.00	\$2.00

图 2.14 简单相关模型

图 2.15 是运行结果的统计数据。注意不存在相关性模型的标准差是 0.23，相比之下，正相关的标准差是 0.30，负相关的是 0.12。在简单模型中，负相关倾向于缩小分布的波动范围，生成的预测分布更加紧凑集中，相比较而言，正相关的波动范围更大些。但是均值大致是保持稳定的。这意味着相关性对项目的预期值影响不大，但是会降低或增加项目的风险。

Revenue Negative Correlation - Risk Simulator 预测窗口

直方图 统计量 用户设置 选项

统计量	结果
仿真次数	1000
均值	1.9976
中位数	1.9961
标准差	0.0717
方差	0.0051
变异系数	0.0359
极大值	2.2148
极小值	1.8197
极差	0.3951
偏度	0.1040
峰度	-0.3191
25分位数	1.9437
75分位数	2.0487
95%置信度的百分误	0.2224%

Revenue No Correlation - Risk Simulator 预测窗口

直方图 统计量 用户设置 选项

统计量	结果
仿真次数	1000
均值	2.0036
中位数	1.9995
标准差	0.1450
方差	0.0210
变异系数	0.0724
极大值	2.3907
极小值	1.6844
极差	0.7063
偏度	0.0304
峰度	-0.7316
25分位数	1.8945
75分位数	2.1128
95%置信度的百分误	0.4486%

Revenue Positive Correlation - Risk Simulator 预测窗口

直方图 统计量 用户设置 选项

统计量	结果
仿真次数	1000
均值	2.0020
中位数	1.9992
标准差	0.1886
方差	0.0356
变异系数	0.0942
极大值	2.4147
极小值	1.6278
极差	0.7869
偏度	0.0788
峰度	-0.9641
25分位数	1.8475
75分位数	2.1480
95%置信度的百分误	0.5839%

图 2.15 相关结果

图 2.16 所示的是运行仿真之后的结果，这里展示了提取的假设的原始数据，计算了变量之间的相关系数。仿真结果中的数据隐含了输入假设。也就是说，如果您键入+0.9 和-0.9 的相关系数，仿真出的结果会具有相同的相关系数。

#### **Spearman's Nonlinear Rank Correlation on Raw Data Extracted from Simulation**

<i>Price Negative Correlation</i>	<i>Quantity Negative Correlation</i>	<i>Price Positive Correlation</i>	<i>Quantity Positive Correlation</i>
676	145	<b>-0.90</b>	102
368	452		461
264	880		515
235	877		477
122	711		874
490	641		769
336	638		481
495	383		627
241	568		82
651	571		190
854	59		659
66	950		188
707	262		286
943	186		439
			981
			972
			569
			812

图 2.16 相关系数恢复

### 2.3.4 精度和误差控制

Monte Carlo 仿真中比较有效的工具之一是精度控制。例如，运行一个复杂的模型需要多少次试验才足够呢？精度控制会自动确定所需的试验次数，当达到预先设定的精度水平时仿真就会停止。

精度控制功能允许您自己设置想要的精度。一般来说，试验的次数越多，置信区间就越窄，统计数据也越精确。Risk Simulator 里的精度控制功能利用置信区间的特征来确定是否达到了某统计量的精度水平。对于每个预测，您都可以指定具体的相应精度的置信区间。

不要混淆这三个不同的概念：误差，精度和置信度。尽管它们听起来差不多，但是三者有很大的区别。下面是对它们的一个简单解释。假设您是一位玉米面豆卷包装制造商，想要找出平均每箱中 100 个包装里的损坏量是多少。一种方法是收集事先包装好的一箱 100 个玉米面豆卷包装，打开箱子计算有多少是破损的。您每天可以生产一百万箱（这是您的总体），您只随机打开 10 箱（这是您的样本大小，就是我们所说的仿真中的试验次数）进行检查。每箱里破损的包装数量统计如下：24, 22, 4, 15, 33, 32, 4, 1, 45, 2。计算出来的平均破损量是 18.2。基于这十个样本的试验，平均值是 18.2，80%的置信区间是从 2 到 33（也就是说在这个样本空间或试验次数条件下，破损量在 2 到 33 之间的可能性是 80%）。但是您能在多大程度上确定 18.2 就是正确的平均数呢？10 次试验足以证明这一点吗？从 2 到 33 的置信区间范围太广，不确定性太大。假设您想要一个更加精确的平均值，90%的置信度和误差控制在 ±2 个玉米面豆卷包装以内——意思是如果您打开一天生产的所有一百万箱产品，九十万箱产品的平均破损量都在某个平均值的 ±2 幅度波动。要达到这种精度水平，您需要选择多少箱玉米面豆卷包装作为样本（或试验次数）呢？这里的 2 代表误差水平，90%是精度水平。如果试验

次数足够多，由于对均值的估计更加精确，那么 90%的置信区间的估计也就更加精确。举一个例子，假设平均值是 20，那么 90%的置信区间就是 18 到 22，这个区间的精度是 90%，也就是在所有打开的一百万个箱子中有九十万箱的破损数量在 18 到 22 个之间。要达到这个精度所需的试验次数可以利用样本的误差公式  $\bar{x} \pm Z \frac{s}{\sqrt{n}}$  计算： $Z \frac{s}{\sqrt{n}}$  代表的是误差 2， $\bar{x}$  是样本均值，代表由 90%的精度水平得到的标准正态 Z 值，s 是样本标准差，n 是满足特定误差和精度水平所需的试验次数。图 2.17 和图 2.18 解释了精度控制是如何运用于 Risk Simulator 的多元仿真预测中的，这个功能使用户不需要根据自己猜测来决定试验的次数。



图 2.17 设置预测的精度水平

统计量	结果
仿真次数	1000
均值	2.0036
中位数	1.9995
标准差	0.1450
方差	0.0210
变异系数	0.0724
极大值	2.3907
极小值	1.6844
极差	0.7063
偏度	0.0304
峰度	-0.7316
25分位数	1.8945
75分位数	2.1128
95%置信度的百分误	0.4486%

图 2.18 误差的计算

### 2.3.5 对预测统计量的理解

大多数分布都可以通过四个矩的计算来描述。第一矩描述分布的位置或集中趋势（期望值），第二矩描述分布的宽度或幅度（风险），第三矩是分布的偏度方向（最可能事件），第四矩是分布的尖峰性或尾部的厚度（突发损失或收益）。在实际项目的分析过程中，所有这四个部分都要被考虑到以便提供充分全面的观点。Risk Simulator 从统计学观点出发，在预测图表中提供了所有这四部分的结果。

#### 测量分布的中心值—— 第一矩

分布的第一矩描述的是某个项目的期望回报率。它测量了项目的定位和可能的平均回报。第一部分常用的统计量包括均值（平均值），中位数（分布的中心）和众数（最可能出现的值）。图 2.19 显示了第一矩——在本例中，是以均值（ $\mu$ ）或平均数来测量的。

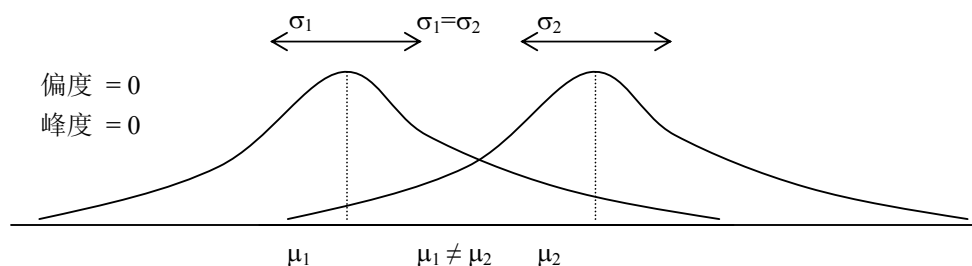


图 2.19——第一矩

#### 测量分布的范围——第 二矩

分布的第二矩描述的是分布的幅度，是风险的测量。分布的幅度或宽度描述了变量的可变性，即是变量会落在分布不同区域的可能性——也就是结果的可能范围。图 2.20 显示的是两个分部，它们有着相同的第一矩（同样的均值），但是第二矩或风险是不同的。图 2.21 中可以看得更加清楚。举例来说，假设有两只股票，第一只股票（黑色线）的波动很小，比较而言，第二只股票（虚线）的价格波动很大。因为风险大的股票与风险小的股票相比其结果不可预知性将加大，很明显投资者会认为波动更大，股票风险更大。图 2.21 的纵轴代表股票价格，因此大风险股票的波动范围更宽。在图 2.20 中这个范围是以分布的宽度（横轴）表示的，宽的分布代表风险大的资产。因此，分布的宽度或幅度代表变量的风险。

注意到在图 2.20 中，两个分布有相同的第一矩或集中趋势，但是很明显两个分布是不同的。分布宽度的差别是可以测量的。从数学和统计学的角度来说，一个变量的宽度或风险可以通过几个不同的统计变量来计算，它们包括范围，标准差，方差，变异系数和百分位数。



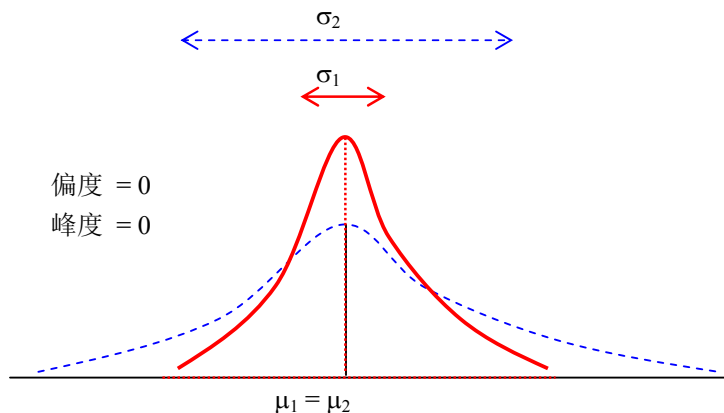


图 2.20——第二矩

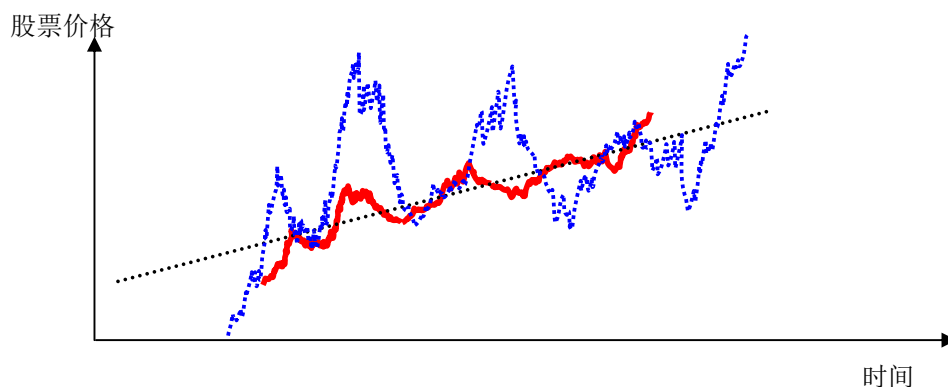


图 2.21 股票价格波动

测量分布的偏度——第三矩

分布的第三矩测量了分布的偏度，就是分布偏度和幅度。图 2.22 显示的是负偏或左偏（分布的尾部偏向左边），图 2.23 显示的是正偏或右偏（分布的尾部偏向右边）。均值一般偏向分布的尾部，而中位数是保持不变的。从图中也可以看出均值是不同的，但中位数和方差可以是相同的。如果不考虑第三矩，只考虑期望值（中位数或均值）和风险（标准差），可能会错误地选择一个正偏的项目！例如如果以横轴表示项目的净收益，那么我们会更偏好左偏分布，因为与较小值出现的可能性较高的分布（图 2.23）相比，这种分布中大值出现的可能性更高（图 2.22）。因此，在偏态分布中，中位数是一个很好的测量回报的工具，由于图 2.22 和 2.23 的中位数是相同的，风险也是相同的，那么此时左偏分布的净收益就是一个较好的选择。如果不考虑项目的分布偏度的话，可能会导致您选择错误的项目（例如两个项目可能有相同的第一个矩和第二个矩，也就是两个项目有相同的回报和风险，但是它们的分布偏度是非常不一样的）。

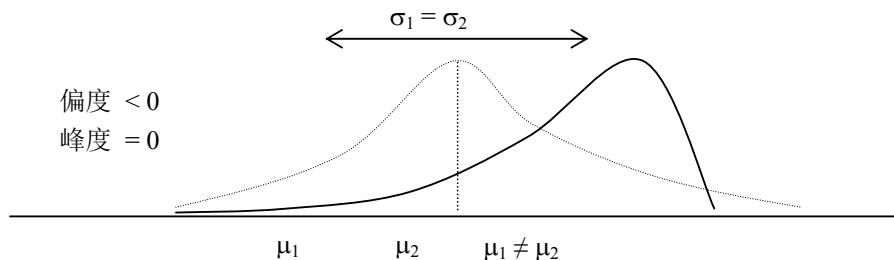


图 2.22 第三个矩（左偏）

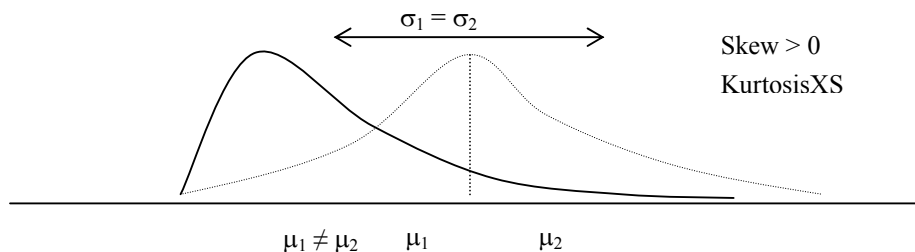


图 2.23 第三矩（右偏）

测量分布中的尾部突发事件——第四矩

第四矩测量的是分布的峰度，图 2.24 显示了这一效果。图中（虚线）是一个峰度为 3.0 的或超峰度为 0.0 的正态分布。Risk Simulator 的结果显示了超峰度的值，用 0 来作为峰度的正常水平，这意味着负的超峰度代表比较扁平的尾部（与均匀分布类似的低峰态分布），正值代表比较肥胖的尾部（与 Student's T 分布和对数正态分布类似的尖峰态分布）。粗线代表的分布有着更大的超峰度，因此曲线尾部所覆盖的区域比中心区域更多。这个因素对图 2.24 中两个分布的风险分析有重大的影响，前三个矩（均值，标准差和偏度）是相同的，但是第四个矩（峰度）是不同的。这种情况表明尽管收益和风险相同，但对于高峰度的分布来说极端和突发事件（可能出现的大的损失或收益）发生的可能性比较大（股票市场的收益就是尖峰度分布,峰度值很大）。忽视项目的峰度可能是非常危险的。一般来说，高超峰度值暗示着分布两端的的风险比较大（例如项目的在险价值（VaR）可能是非常重要的）。

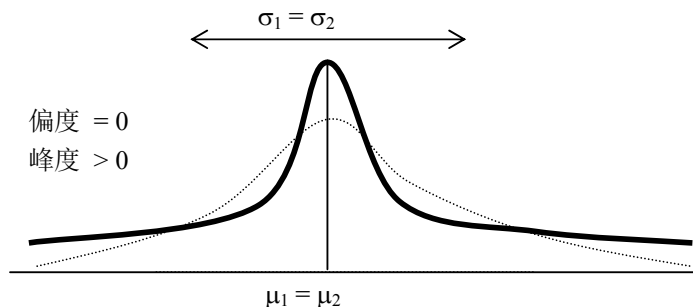


图 2.24 第四矩

### 2.3.6 了解 Monte Carlo 仿真的概率分布

本章节介绍了 Monte Carlo 仿真的作用，但是为了开始仿真，首先要了解概率分布中的一些概念。在开始介绍概率之前，先来看看这个例子：您想要查看一家大公司某个部门里未被免职人员的工资分布。第一步您要搜集原始数据——在本例中是指此部门每个未被免职员工的工资。第二步将数据组织好绘成频数分布图。为了画频数分布图，首先要将工资按照区间进行分组，并在图的横轴上列出这些区间。然后在纵轴上每个对应的区间标出员工的个数或频数。现在您可以清晰地看到此部门的未免职员工的工资分布图。

图 2.25 中显示出大部分员工（大约 180 人里有 60 人）每小时的工资是位于\$7.00 到\$9.00 之间。

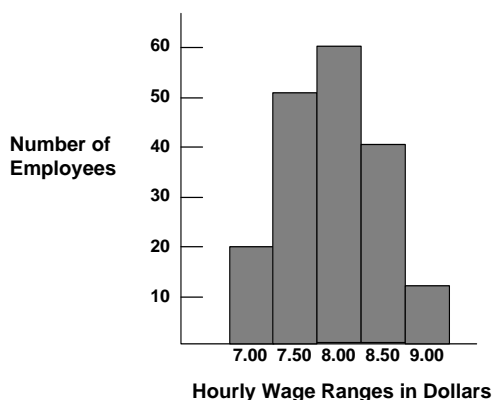


图 2.25 频数方柱图 I

您还可以将这些数据绘成概率分布图。概率分布代表的是每个区间的员工人数占总员工数比重。为了制作这个概率分布图，您可以用每个区间的员工人数除以总员工人数，将得到的结果标在图表的纵轴上。

图 2.26 显示的就是每组员工数占总员工数的比例；您可以估计从整个集合中随机抽取一位员工，他的工资位于某一区间的可能性或概率。例如，假设在同样的条件下抽取样本，那么从整个集合中随机抽取一位员工其小时收入落在\$8.00 到\$8.50 之间的概率是 0.33（三个之中有一个）。

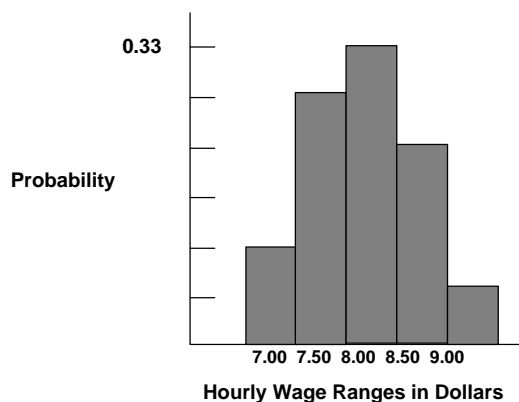


图 2.26 频数方柱图 II

概率分布可以是离散的，也可以是连续的。*离散概率分布*描述的是独立的值，通常是整数，没有中间值，用一系列直方柱来表示。例如离散分布可以用来描述四次抛硬币过程中出现头像的次数 0, 1, 2, 3 或 4。*连续分布*是一个比较抽象的数学概念，因为它假设在两个数字之间存在中间值。也就是说连续分布假设在分布的任意两点之间存在无数的数值。但是在很多情况下您可以用连续分布来近似描述一个离散分布，即使连续模型不能确切地描述这种情况。

### 选择正确的概率分布

将数据绘制成点图是选择概率分布的方法之一。以下的步骤提供了另外一种在电子表格中为不确定变量选择最佳概率分布的方法。

为了选择合适的概率分布，请参照以下步骤：

- 查看需要讨论的变量,列出已知的关于此变量的所有相关情况。可以从此未知变量的历史数据中搜集信息。如果得不到历史数据，就基于以往的经验做出判断。
- 回顾关于概率分布的描述。
- 选择可以表现变量特征的概率分布。当分布的特征与变量的特征相符时，就说这个分布可以表现该变量的特征。

### Monte Carlo 仿真

简单地来说 Monte Carlo 仿真就是一个随机数生成器，用来预测，估计和进行风险分析。仿真就是通过重复地从一个事先定义的不确定性变量的概率分布中挑选出一些数值，并将这些数值用于模型来计算一个模型的众多情景。由于所有这些情景在模型中都会产生关联性的结果，每个情景可以有一种预测结果。预测通常包含公式和函数，是模型的一个重要输出量。这些事件通常是总量、净利润和总成本等等。

简单说来，蒙特卡罗仿真就像是从一个大篮子中不断的抽取和放回高尔夫球。篮子的形状和大小取决于分布的假设（例如，均值为 100 标准差为 10 的正态分布，均匀分布或者三角分布）。由于有的篮子较深，有的篮子较为对称，让某些球被更为频繁地取到。重复取出球的次数取决于仿真的次数。对于包含相关性假设的更大的模型，就像是一个非常巨大的篮子，里面还包含很多的小篮子。每个小篮子都包含一系列的高尔夫球。有的时候小篮子和小篮子之间会缠在一起（如果变量之间包含相关性），当前一个球被取到的时候，后一个球也有机会被取到。这些小球每次被取出，它们之间的交互作用被记录和计算下来，作为仿真的预测结果。利用 Monte Carlo 仿真，Risk Simulator 为每个概率分布假设所生成的随机数都是完全独立的。换句话说，某次试验所选择的随机数对下一组随机数的生成没有任何影响。也可以使用蒙特卡罗抽样技术在电子表格中对现实世界进行情景模拟。

## 2.4 离散分布

以下是一份 Monte Carlo 仿真中可以使用的不同类型概率分布的详细介绍列表。为了便于用户参考我们在附录中也包含了此份列表。

### 伯努力分布

伯努力分布是一种离散的概率分布，满足该分布的事件只包含两种结果（例如，正面或反面，成功或失败），这就是它也被称为（0，1）分布的原因。只进行一次二项分布实验的结果满足伯努力分布。它往往是任何复杂概率分布的基础。例如：

- 二项分布：多次试验的伯努力分布，计算的是总试验次数中成功次数  $x$  的概率。
- 几何分布：多次试验的伯努力分布，计算的是第一次成功出现之前总的失败次数。
- 负二项分布：试验次数较多的伯努力分布，计算的是第  $x$  次成功出现之前总的失败次数。

伯努力分布的数学表达式如下：

$$P(n) = \begin{cases} 1-p & \text{for } x=0 \\ p & \text{for } x=1 \end{cases}$$

或者

$$P(n) = p^x(1-p)^{1-x}$$

$$\text{均值} = p$$

$$\text{标准差} = \sqrt{p(1-p)}$$

$$\text{偏度} = \frac{1-2p}{\sqrt{p(1-p)}}$$

$$\text{超峰度} = \frac{6p^2 - 6p + 1}{p(1-p)}$$

成功的概率  $p$  是唯一的分布参数。要注意伯努力分布只有一次试验，仿真值可能是 0 或 1 中的一个。

参数输入要求：

成功的概率  $>0$  并且  $<1$ （也就是  $0.0001 \leq P \leq 0.9999$ ）

### 二项分布

二项分布描述的是在固定次数的试验中某个事件发生的次数，例如在十次抛硬币的过程中头像出现的次数或是选择的 50 个物品中次品的数目。

条件

二项分布的三个先决条件是：

- 每次试验只可能出现两种结果，并且它们之间是互斥的。

- 试验之间是相互独立的——第一次试验的结果不会影响下一次试验。
- 每次试验中事件发生的概率是相同的。

二项分布的数学表达式如下：

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)} \quad n > 0; x = 0, 1, 2, \dots, n; 0 < p < 1$$

$$\text{均值} = np$$

$$\text{标准差} = \sqrt{np(1-p)}$$

$$\text{偏度} = \frac{1-2p}{\sqrt{np(1-p)}}$$

$$\text{超峰度} = \frac{6p^2 - 6p + 1}{np(1-p)}$$

分布的参数有成功的概率（p）和总试验次数（n）。试验的成功次数用 x 表示。注意成功概率值为 0 和 1 的情况价值不高，不需要进行仿真，所以在软件中不允许这两种情况。

输入假设：

成功概率 > 0 且 < 1（也就是  $0.0001 \leq P \leq 0.9999$ ）

试验次数  $\geq 1$  且  $\leq 1000$ ，必须为正整数（对于大型试验，则采用正态分布，计算出的二项均值和标准差做为正态分布的参数）。

## 离散均匀分布

离散型均匀分布是一个描述等可能发生事件的概率分布，如描述 N 个事件，每个事件发生的概率相同。分布类似于均匀分布，但是是离散的而非连续的。

离散均匀分布的数学表达式如下：

$$P(x) = \frac{1}{N}$$

$$\text{均值} = \frac{N+1}{2} \times (\text{最大值} - \text{最小值})$$

$$\text{标准差} = \sqrt{\frac{(N-1)(N+1)}{12}} \times (\text{最大值} - \text{最小值})$$

偏度 = 0 (分布是完全对称的)

$$\text{超峰度} = \frac{-6(N^2+1)}{5(N-1)(N+1)} \times (\text{最大值} - \text{最小值})$$

参数输入要求：

最小值 < 最大值，并且两者必须均是整数（包含负整数和 0）

## 几何分布

几何分布描述了直到第一次成功出现试验进行的次数，例如，在赢得转盘游戏之前所转转盘的次数。试验的次数不是固定的，试验不断进行直到成功事件第一次出现，每次试验的成功概率都是一定的。

条件

几何分布的三个先决条件是：

- 试验次数是不固定的。
- 试验要一直进行直到第一次成功出现。
- 在每次试验中事件发生的概率是相同的。

几何分布的数学表达式如下：

$$P(x) = p(1-p)^{x-1} \quad 0 < p < 1 \quad x = 1, 2, \dots, n$$

$$\text{均值} = \frac{1}{p} - 1$$

$$\text{标准差} = \sqrt{\frac{1-p}{p^2}}$$

$$\text{偏度} = \frac{2-p}{\sqrt{1-p}}$$

$$\text{超峰度} = \frac{p^2 - 6p + 6}{1-p}$$

成功概率  $p$  是唯一的分布参数。试验的成功次数用  $x$  表示，只能取正整数。

参数输入要求：

成功的概率  $>0$  并且  $<1$ （也就是  $0.0001 \leq p \leq 0.9999$ ）。注意成功概率值为  $0$  和  $1$  的情况意义不大，不需要进行仿真，所以在软件中不允许这两种情况。

## 超几何分布

超几何分布与二项分布类似，都是描述在一个固定数目的试验次数中，某一特定事件发

生的次数。区别在于二项分布中的试验是相互独立的，而超几何分布的试验会改变每一个后发试验的概率，被称为“不放回的试验”。例如，假设已知一个装有机件零件的箱子里有一些次品。您从箱子里挑选了一件后，发现是次品，就将它从箱子里移走。如果再从箱子里取出另一个零部件，与第一次相比它是次品的概率就降低了，因为已经取走了一件次品。如果放回那件次品的话，概率还是相同的，整个过程就满足二项分布的条件了。

条件

超几何分布的先决条件有：

- 总的元素或项目（样本空间）的数目是固定的，一个有限的总体。总体空间必须小于等于 1750。
- 样本空间（或试验次数）是总体的一部分。
- 在每次试验后总体中最初已知的成功概率会发生变化。

超几何分布的数学表达式如下：

$$P(x) = \frac{\frac{(N_x)!}{x!(N_x-x)!} \frac{(N-N_x)!}{(n-x)!(N-N_x-n+x)!}}{\frac{N!}{n!(N-n)!}} \quad x = \text{Max}(n - (N - N_x), 0), \dots, \text{Min}(n, N_x)$$

$$\text{均值} = \frac{N_x n}{N}$$

$$\text{标准差} = \sqrt{\frac{(N - N_x) N_x n (N - n)}{N^2 (N - 1)}}$$

$$\text{偏度} = \sqrt{\frac{N - 1}{(N - N_x) N_x n (N - n)}}$$

超峰度 = 峰度 - 3 (具体函数形式比较复杂, 在此不作具体描述)

参数输入要求:

总体空间  $\geq 2$ , 并且是整数

样本空间  $> 0$ , 并且是整数

总体成功数  $> 0$ , 并且是整数

总体空间  $>$  总体成功数

样本空间  $<$  总体成功数

总体空间  $< 1750$

## 负二项分布

负二项分布被用于模拟基于既定成功次数 (R) 的额外实验次数的概率分布。例如, 为了获得一个 10 份订单的机会, 在给定每个电话的成功几率的前提下, 在打过 10 个电话之后, 还需要打多少个电话? x 轴展示了需要打的额外的电话次数或者失败的电话次数。试验的次数并不固定, 试验继续进行直到第 R 次成功之后, 每次成功的概率一定。

条件

负二项分布的三个先决条件是:

- 试验次数是不固定的。
- 试验要一直进行直到第 R 次成功出现。
- 每次试验某类结果发生的概率是一个常数。

负二项分布的数学表达式如下:

$$P(x) = \frac{(x + r - 1)!}{(r - 1)! x!} p^r (1 - p)^x \quad x = r, r + 1, \dots; 0 < p < 1$$

$$\text{方差} = \sqrt{\frac{r(1-p)}{p^2}}$$

$$\text{均值} = \frac{r(1-p)}{p}$$

$$\text{偏度} = \frac{2-p}{\sqrt{r(1-p)}}$$



$$\text{超峰度} = \frac{p^2 - 6p + 6}{r(1-p)}$$

分布参数有成功的概率（p）和成功试验次数（R）。

参数输入要求：

需要的成功次数必须是正整数，且>0，<8000。

成功的概率>0，<1（也就是，0.0001≤P≤0.9999），这点要注意

成功概率值为 0 和 1 的情况意义不大，不需要进行仿真，所以在软件中不允许这两种情况。

## 帕斯卡分布

帕斯卡分布用于模拟在进行多少次实验后，成功事件发生的情形。例如，为了成交 10 个销售机会，在给定每个成功概率的情况下，需要总共拨打多少个销售电话？X 轴显示了总共需要拨打的电话，包括成功和失败的电话。实验的次数不固定，实验继续直到出现第 R 次成功，每次成功的概率相同。帕斯卡分布和负二项分布相关。负二项分布被用于模拟这样的概率分布，基于既定成功次数的额外实验次数（换句话说就是失败的次数），而帕斯卡分布计算了成功和失败的总数。成功数和成功概率是该分布的参数。

帕斯卡分布的数学表达式如下：

$$f(x) = \begin{cases} \frac{(x-1)!}{(x-s)!(s-1)!} p^s (1-p)^{x-s} & \text{for all } x \geq s \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} \sum_{x=1}^k \frac{(x-1)!}{(x-s)!(s-1)!} p^s (1-p)^{x-s} & \text{for all } x \geq s \\ 0 & \text{otherwise} \end{cases}$$

$$\text{均值} = \frac{s}{p}$$

$$\text{标准差} = \sqrt{s(1-p)p^2}$$

$$\text{偏度} = \frac{2-p}{\sqrt{r(1-p)}}$$

$$\text{超峰度} = \frac{p^2 - 6p + 6}{r(1-p)}$$

输入要求

成功数>0 必须是整数

0 ≤ 概率 ≤ 1

泊松分布

泊松分布描述的是在某个指定区间内事件发生的次数，例如每分钟的电话个数或是文件中每页的错误数量。

条件

泊松分布的三个先决条件是：

- 在任何范围内事件出现的次数可以是无限的。
- 事件之间是相互独立的。某一区间的事件数量不会影响其它区间的事件数量。
- 区间之间的平均事件发生概率要保持一致。

泊松分布的数学表达式如下：

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x \text{ 和 } \lambda > 0$$

$$\text{均值} = \lambda$$

$$\text{标准差} = \sqrt{\lambda}$$

$$\text{偏度} = \frac{1}{\sqrt{\lambda}}$$

$$\text{超峰度} = \frac{1}{\lambda}$$

唯一的分布参数是比率或（ $\lambda$ ）。

参数输入要求：

比率 $>0$ ， $\leq 1000$ （也就是， $0.0001 \leq \text{比率} \leq 1000$ ）

## 2.5 连续分布

### 反正弦分布

反正弦函数是 U 型的 Beta 分布的特殊形态，当尺度参数和形状参数都等于 0.5。接近最大值和最小值的数值具有较高出现的概率，在这两个值之间的数值具有较小的概率。最大值和最小值是分布的参数。

反正弦分布的数学表达式如下：

$$f(x) = \begin{cases} \frac{1}{\pi\sqrt{x(1-x)}} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{2}{\pi} \sin^{-1}(\sqrt{x}) & \text{for } 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

$$\text{均值} = \frac{\text{Min} + \text{Max}}{2}$$

$$\text{标准差} = \sqrt{\frac{(\text{Max} - \text{Min})^2}{8}}$$

偏度=0

超峰度= 1.5

输入要求：

最小值<最大值

### Beta 分布

Beta 分布非常的灵活，通常用来表现某个固定区域内的可变性。当描述的历史数据和预测随机行为为百分比形式时，也就是说数值的取值在 0 和 1 之间，您往往可以使用 Beta 分布。改变 Beta 分布的形状和大小只要改变该分布的两个重要参数：alpha 和 beta（只能取正数）。并且如果这两个参数相等，分布就是对称的；如果任意一个参数为 1，另外一个大于 1，分布就是三角形的或者 J 形的；如果 alpha 小于 beta，分布是正偏度的（大部分的值都靠近最小值）。如果 alpha 大于 beta，分布是负偏度的（大部分值都靠近最大值）。

Beta 分布的数学表达式如下：

$$f(x) = \frac{(x)^{(\alpha-1)}(1-x)^{(\beta-1)}}{\left[ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \right]} \quad \text{for } \alpha > 0; \beta > 0; x > 0$$

$$\text{均值} = \frac{\alpha}{\alpha + \beta}$$

$$\text{标准差} = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(1 + \alpha + \beta)}}$$

$$\text{偏度} = \frac{2(\beta - \alpha)\sqrt{1 + \alpha + \beta}}{(2 + \alpha + \beta)\sqrt{\alpha\beta}}$$

$$\text{超峰度} = \frac{3(\alpha + \beta + 1)[\alpha\beta(\alpha + \beta - 6) + 2(\alpha + \beta)^2]}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)} - 3$$

分布包含两个参数  $\alpha$  和  $\beta$ ， $\Gamma$  是  $\gamma$  函数

条件

Beta 分布的两个先决条件是：

- 不定变量是 0 到 1 之间的随机正值。
- 利用两个正值可以确定分布的形状。

参数输入要求：

$\alpha$  和  $\beta > 0$ ，可以取任意正值

### Beta 乘数转换分布

Beta 分布非常灵活常用来表示某个固定距离之内的变化性。用来描述经验数据预测 0 或者 1 结果的比例的随机行为。Beta 分布的形状各异会随着参数  $\alpha$  和  $\beta$  的变化而变化。Beta 转换分布可以通过 beta 分布乘以某个倍数，转换位置参数，使输出的结果超过自然准则 0 和 1 的界限。Alpha, beta 和位置参数是输入参数。

输入要求

Alpha > 0

Beta > 0

位置参数可以是任何正数或者负数包括 0

参数 > 0

### 柯西分布、Lorentzian 分布或 Breit-Wigner 分布

柯西分布，也被称为 **Lorentzian** 分布或 **Breit-Wigner** 分布，是用于描述共振行为的连续分布。它还可以用来描述与 x 轴成任意角度相交的某条斜线的水平距离的分布。

柯西分布或洛伦兹分布的数学表达式如下：

$$f(x) = \frac{1}{\pi} \frac{\gamma/2}{(x - m)^2 + \gamma^2/4}$$

柯西分布是一个特例，它不存在任何理论部分（均值，标准差，偏度和峰度），因为它们都是不明确的。

模的位置（ $\alpha$ ）和尺度（ $\beta$ ）是本分布的两个参数。位置参数描述了分布的高峰和众数的位置，而尺度参数描述了分布的宽带。此外，柯西分布或洛伦兹分布的均值和方差都是不确定的。

另外，柯西分布是自由度为 1 的学生分布。

此分布还可以由两个标准正态分布的比率构成（均值为 0，方差为 1 的正态分布），

且两者相互独立。

参数输入要求：

位置参数  $\alpha$  可以取任意值

尺度参数  $\beta$  可以是大于 0 的任何正值

## 卡方分布

卡方分布是用来进行假设检验的概率分布，和 Gamma 分布和标准正态分布类似。例如，独立的标准正态分布平方和是满足自由度为  $k$  的卡方分布：

$$Z_1^2 + Z_2^2 + \dots + Z_k^2 \sim \chi_k^2$$

卡方分布的数学表达式如下：

$$f(x) = \frac{0.5^{-k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2} \quad x > 0$$

均值 =  $k$

标准差 =  $\sqrt{2k}$

$$\text{偏度} = 2\sqrt{\frac{2}{k}}$$

$$\text{超峰度} = \frac{12}{k}$$

$\Gamma$  代表  $\gamma$  函数。自由度  $k$  是唯一的分布参数。

$\gamma$  分布可以通过一定的设置变为卡方分布：

$$\text{形状参数} = \frac{k}{2}, \quad \text{尺度} = 2S^2, \quad S \text{ 表示尺度。}$$

参数输入要求：

自由度为大于 1 小于 300 的整数

## 余弦分布

余弦函数看起来有点像逻辑分布在最大值和最小值之间的中位数具有最高的峰值或者出现最多的可能性，当接近最小值和最大值的时候概率逐渐变小。最大值和最小值是分布的两个参数。

余弦分布的数学表达式如下：

$$f(x) = \begin{cases} \frac{1}{2b} \cos\left[\frac{x-a}{b}\right] & \text{for } \min \leq x \leq \max \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where } a = \frac{\min + \max}{2} \text{ and } b = \frac{\max - \min}{\pi}$$

$$F(x) = \begin{cases} \frac{1}{2} \left[ 1 + \sin\left(\frac{x-a}{b}\right) \right] & \text{for } \min \leq x \leq \max \\ 1 & \text{for } x > \max \end{cases}$$

## Risk Simulator

$$\text{均值} = \frac{Min + Max}{2}$$

$$\text{标准差} = \sqrt{\frac{(Max - Min)^2 (\pi^2 - 8)}{4\pi^2}}$$

偏度=0

$$\text{超峰度} = \frac{6(90 - \pi^4)}{5(\pi^2 - 6)^2}$$

输入要求:

最小值<最大值

## 双对数分布

双对数分布看起来有点像柯西分布，中心具有峰态，然后迅速向最大值和最小值下降，并且以中心为轴对称分布。最大值和最小值是分布的两个参数。

双对数分布的数学表达式如下:

$$f(x) = \begin{cases} \frac{-1}{2b} \ln\left(\frac{|x-a|}{b}\right) & \text{for } \min \leq x \leq \max \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where } a = \frac{\min + \max}{2} \text{ and } b = \frac{\max - \min}{2}$$

$$F(x) = \begin{cases} \frac{1}{2} - \left(\frac{|x-a|}{2b}\right) \left[1 - \ln\left(\frac{|x-a|}{b}\right)\right] & \text{for } \min \leq x \leq a \\ \frac{1}{2} + \left(\frac{|x-a|}{2b}\right) \left[1 - \ln\left(\frac{|x-a|}{b}\right)\right] & \text{for } a \leq x \leq \max \end{cases}$$

$$\text{均值} = \frac{Min + Max}{2}$$

$$\text{标准差} = \sqrt{\frac{(Max - Min)^2}{36}}$$

偏度=0

输入要求:

最小值<最大值

## Erlang 分布

Erlang分布和Gamma分布类似要求Alpha或者形状参数为正整数。Erlang分布的一个应用案例就是用来校准粒子穿越一个系统空间的比率。该系统被广泛的用于生物学和生态学（例如，某种流行病，某个个体从健康成为疾病的携带者服从指数分布，接着以指数分布的概率从携带者成为感染者）。Alpha（形状参数）Beta（尺度参数）是该分布的参数。

Erlang 分布的数学表达式如下：

$$f(x) = \begin{cases} \frac{\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta}}{\beta(\alpha-1)!} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-x/\beta} \sum_{i=0}^{\alpha-1} \frac{(x/\beta)^i}{i!} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{均值} = \alpha\beta$$

$$\text{标准差} = \sqrt{\alpha\beta^2}$$

$$\text{偏度} = \frac{2}{\sqrt{\alpha}}$$

$$\text{超峰度} = \frac{6}{\alpha} - 3$$

输入要求：

Alpha（形状参数）>0 必须是整数

Beta（形状参数）>0

## 指数转换分布

指数分布被广泛地用于描述独立随机事件发生的时间间隔的分布，例如电子产品的寿命分布或者到达服务摊点的时间间隔。它与泊松分布相关（泊松分布用来描述在某个时间间隔内事件发生的次数）。指数分布的重要特性就是它的无记忆性，这意味着未来产品的寿命也满足相同的概率分布，不必考虑时间的存在。换句话说，时间对未来事件的发生没有影响。成功概率( $\lambda$ )是分布的唯一参数。

输入变量的要求：Lambda >0. 位置参数可以是任何正数或者负数包括0.

## 指数分布

指数分布被广泛地用于描述独立随机事件发生的时间间隔的分布，例如电子产品的寿命分布或者到达服务摊点的时间间隔。它与泊松分布相关（泊松分布用来描述在某个时间间隔内事件发生的次数）。指数分布的重要特性就是它的无记忆性，这意味着未来产品的寿命也满足相同的概率分布，不必考虑时间的存在。换句话说，时间对未来事件的发生没有影响。

指数分布的数学表达式如下：

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0; \lambda > 0$$

$$\text{均值} = \frac{1}{\lambda}$$

$$\text{标准差} = \frac{1}{\lambda}$$

偏度=2（这个值适用于所有成功率  $\lambda$ ）

超峰度=6（这个值适用于所有成功率  $\lambda$ ）

成功率（ $\lambda$ ）是唯一的分布参数。成功的试验次数用  $x$  表示。

指数分布的先决条件是：

- 指数分布描述的是事件之间的时间间隔。

参数输入要求：

成功率 > 0

## 极值分布或 Gumbel 分布

极值分布（类型 1）通常用于描述一段时间内，极大值响应的大小（例如，洪水，降雨，地震等等）。其它一些应用包括材料的断裂强度，建筑设计，飞机负荷量等。极值分布也被称为 **Gumbel** 分布。

极值分布的数学表达式如下：

$$f(x) = \frac{1}{\beta} z e^{-z} \quad z = e^{\frac{x-\alpha}{\beta}} \quad \beta > 0; \quad x \quad \text{和 } \alpha \text{ 可取任何值}$$

$$\text{均值} = \alpha + 0.577215\beta$$

$$\text{标准差} = \sqrt{\frac{1}{6}\pi^2\beta^2}$$

$$\text{偏度} = \frac{12\sqrt{6}(1.2020569)}{\pi^3} = 1.13955 \quad (\text{适用于所有的模和尺度值})$$

超峰度=5.4（适用于所有的模和尺度值）

偏度=（适用于所有的模和尺度值）

超峰度=（适用于所有的模和尺度值）

模（ $\alpha$ ）和尺度（ $\beta$ ）是分布参数。

参数的计算：

极值分布里存在两个标准参数：模和尺度。模参数是变量的极大值（概率分布的



最高点)。在您选择模参数之后，您就可以对尺度参数进行估计。尺度参数的值大于 0。尺度参数越大，变量越大。

参数输入要求：

模  $\alpha$  可以取任意值

尺度  $\beta > 0$

## F 分布

F 分布，也称作 **Fisher-Snedecor** 分布，也是一种连续的概率分布，常常用来进行假设检验。特别是，在方差分析中用来检验两个方差的统计性差异。分子自由度为  $n$ ，分母自由度为  $m$  的 F 分布与卡方分布之间的关系如下：

$$\frac{\chi_n^2 / n}{\chi_m^2 / m} \sim F_{n,m}$$

$$\text{均值} = \frac{m}{m-2}$$

$$\text{标准差} = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)} \quad \text{for all } m > 4$$

$$\text{偏度} = \frac{2(m+2n-2)}{m-6} \sqrt{\frac{2(m-4)}{n(m+n-2)}}$$

$$\text{超峰度} = \frac{12(-16+20m-8m^2+m^3+44n-32mn+5m^2n-22n^2+5mn^2)}{n(m-6)(m-8)(n+m-2)}$$

分子自由度  $n$  和分母自由度  $m$  是唯一的分布参数。

参数输入要求：

分子和分母自由度均为大于 0 的整数

## Gamma 分布 (Erlang 分布)

可以应用 Gamma 分布的物理量很多，并且其与其它一些分布都有关联：如对数正态分布，指数分布，Pascal 分布，Erlang 分布，泊松分布和卡方分布等。它可以被用于表示气象过程中的污染浓度和降水量，或是在当事件过程并不是完全随机情况下，度量事件发生的间隔，其它一些应用包括存货控制，经济理论和保险事故理论等。

条件

Gamma 分布最常见的用途是用于泊松过程中某事件第  $r$  次发生时的时间分布。应用于此种情况必须三个条件有：

- 任一计量单位里事件可能发生的次数不限定于某个固定的数目。
- 每次发生的过程是相互独立的。在某个计量单位发生的次数不影响另一计量单位的发生次数。
- 单位之间发生次数的平均值保持不变。

Gamma 分布的数学表达式如下：

$$f(x) = \frac{\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta} \quad \alpha > 0 \quad \beta > 0, \alpha, \beta \text{可取任何值}$$

$$\text{均值} = \alpha\beta$$

$$\text{标准差} = \sqrt{\alpha\beta^2}$$

$$\text{偏度} = \frac{2}{\sqrt{\alpha}}$$

$$\text{超峰度} = \frac{6}{\alpha}$$

形状参数（ $\alpha$ ）和尺度参数（ $\beta$ ）是分布参数， $\Gamma$  是 Gamma 函数。

参数  $\alpha$  为正值的 Gamma 分布被称为 Erlang 分布，被用于预测排队系统中的等待时间，Erlang 分布是一些独立同一的且服从无记忆指数分布的随机分布变量的总和。假定这些随机变量的个数为  $n$ ，那么 Erlang 分布的数学表达式如下：

$$f(x) = \frac{x^{n-1} e^{-x}}{(n-1)!} \quad x > 0, n \text{取正整数}$$

参数输入要求：

尺度参数  $\beta$  可以为大于 0 的任意数

形状参数  $\alpha$  可以为大于等于 0.5 任意数

位置数可以取任意值

## 拉普拉斯分布

拉普拉斯分布被称为双指数分布，它可以构建两个指数分布（增加了位置参数），使中间峰值格外突出。拉普拉斯分布的概率密度函数会使人联想到正态分布。但是，正态分布是以相对于均值的标准差来表示的，而拉普拉斯分布是以均值的绝对差异来表示的。使拉普拉斯分布的尾部较正态分布更为宽阔。当位置参数为 0，拉普拉斯分布的随机变量是以尺度参数逆向指数分布的。Alpha（位置参数）和 beta（尺度参数）是该分布的参数。

拉普拉斯分布的数学表达式如下：

$$f(x) = \frac{1}{2\beta} \exp\left(-\frac{|x-\alpha|}{\beta}\right)$$

$$F(x) = \begin{cases} \frac{1}{2} \exp\left[\frac{x-\alpha}{\beta}\right] & \text{when } x < \alpha \\ 1 - \frac{1}{2} \exp\left[-\frac{x-\alpha}{\beta}\right] & \text{when } x \geq \alpha \end{cases}$$

$$\text{均值} = \alpha$$

$$\text{标准差} = 1.4142\beta$$

偏度=0

超峰度=3

输入要求:

Alpha (位置参数) 可以取任何整数或者负数, 包括0

Beta (尺度参数) >0

## Logistic 分布

**Logistic** 分布通常用来描述增长现象, 例如, 人口数量随着时间变化的函数。也用来描述化学反应和群体或者个体的增长过程。

**Logistic** 分布的数学表达式如下:

$$f(x) = \frac{1}{x\sqrt{2\pi} \ln(\sigma)} e^{-\frac{[\ln(x)-\ln(\mu)]^2}{2[\ln(\sigma)]^2}} \quad x > 0; \mu > 0 \quad \sigma > 0$$

$$= \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

均值

$$\text{标准差} = \sqrt{\exp(\sigma^2 + 2\mu)[\exp(\sigma^2) - 1]}$$

$$\text{偏度} = \left[\sqrt{\exp(\sigma^2) - 1}\right](2 + \exp(\sigma^2))$$

$$\text{超峰度} = \exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 6$$

对所有均值和尺度均适用

均值 (α) 和尺度 (β) 是分布参数。

参数的计算:

**Logistic** 分布的两个标准参数为: 均值和尺度。均值参数就是平均值, 在这个分布里也就是指众数, 因为它属于对称分布。在选择均值参数之后, 您就可以估计出尺度参数。尺度参数为大于 0 的任意值。尺度参数越大, 方差越大。

参数输入要求:

尺度 β 为大于 0 的任意值

均值 α 可以取任意值

## 对数正态分布

对数正态分布被广泛地用于数值偏度为正的情形。例如, 在证券估价 (如股价等) 的财务分析中, 标的数值大小都不可能小于 0, 且通常都是正偏而非正态分布的 (对称的)。同样地, 房地产的价格也表现出正偏特性。这些含有不确定性的变量不可能小于零, 但大部分值都接近于下限。

对数正态分布的三个先决条件是:

- 不确定变量的增长没有上限, 但是不能低于 0。
- 不确定变量的分布是正偏的, 也就是大部分的值都集中在低端。
- 不确定变量的自然对数服从正态分布。

一般来说, 如果变异系数大于 30%, 就使用对数正态分布。否则, 就使用正态分布。

对数正态分布的数学表达式如下：

$$f(x) = \frac{1}{x\sqrt{2\pi \ln(\sigma)}} e^{-\frac{[\ln(x)-\ln(\mu)]^2}{2[\ln(\sigma)]^2}} \quad x > 0; \mu > 0 \quad \sigma > 0$$

$$\text{均值} = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$\text{标准差} = \sqrt{\exp(\sigma^2 + 2\mu)[\exp(\sigma^2) - 1]}$$

$$\text{偏度} = \left[\sqrt{\exp(\sigma^2) - 1}\right](2 + \exp(\sigma^2))$$

$$\text{超峰度} = \exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 6$$

均值（ $\mu$ ）和标准差（ $\sigma$ ）是分布参数。

参数输入要求：

均值和标准差均可以为大于 0 的任意正值

设定对数正态分布的参数

在通常情况下，对数正态分布使用的是算术平均值和标准差。在可以得到历史数据的情况下，我们使用对数平均值和标准差，或是几何平均数和标准差更合适。

## 正态分布

正态分布是概率里最重要的分布，因为它可以描述很多自然现象，如人们的 IQ 值或高度。决策者们也可以利用正态分布来描述诸如通货膨胀率或未来汽油价格等不定变量。

条件

正态分布的三个先决条件是：

- 不确定变量的某些值是最可能值（分布的均值）。
- 不确定变量可能高于均值，也可能低于均值（对称的分布在均值周围）。
- 不确定变量更可能分布在均值的附近而不是较远处。

正态分布的数学表达式如下：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \text{ 和 } \sigma \text{ 可取任何值; } \sigma > 0$$

$$\text{均值} = \mu$$

$$\text{标准差} = \sigma$$

偏度=0（对于所有的均值和标准差均适用）

超峰度=0（对于所有的均值和标准差均适用）

均值（ $\mu$ ）和标准差（ $\sigma$ ）是分布参数。

参数输入要求：

标准差可以为大于 0 的任意正值

均值可以取任意值

## Risk Simulator

### 对数正态分布转换分布

对数正态分布广泛地用于数值大部分左偏的情形，在金融证券分析或者地产估值，价值不可能为 0。股票的价格都是正偏斜的而非对称的正态分布。股票价格趋于该形态，具有 0 下限。相对而言，对数正态转换分布类似于正态分布，但是它的结果被转换可以取负数值。均值，标准差和转换值是该分布的参数。

输入要求：

均值>0

标准差>0

转换参数可以是正数或者负数包括 0

### 抛物线的分配

抛物线的分配是 Beta 分配的一个特别的案例当形状=规模=2。靠近最小值和最大值的值有发生的低可能性鉴于这些二个极限之间的值有更高的可能性或发生。最小值和最大值是 distributional 参数。

抛物线的分配的数学表达式如下：

$$f(x) = \frac{(x)^{(\alpha-1)}(1-x)^{(\beta-1)}}{\left[ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \right]} \quad \text{for } \alpha > 0; \beta > 0; x > 0$$

Alpha = Beta = 2

$$\text{均值} = \frac{\text{Min} + \text{Max}}{2}$$

$$\text{标准差} = \sqrt{\frac{(\text{Max} - \text{Min})^2}{20}}$$

偏度=0

超峰度=-0.8571

输入要求：

最小值<最大值

### Pareto 分布

**Pareto** 分布被广泛地用于这样的现象，例如城市的人口数量，自然资源的数量，公司大小，个人收入，股票价格的波动，通信电路的错误集中率。

**Pareto** 分布的数学表达式如下：

$$f(x) = \frac{\beta L^\beta}{x^{(1+\beta)}} \quad x > L$$

$$\text{均值} = \frac{\beta L}{\beta - 1}$$

$$\text{标准差} = \sqrt{\frac{\beta L^2}{(\beta - 1)^2(\beta - 2)}}$$

$$\text{偏度} = \sqrt{\frac{\beta-2}{\beta} \left[ \frac{2(\beta+1)}{\beta-3} \right]}$$

$$\text{超峰度} = \frac{6(\beta^3 + \beta^2 - 6\beta - 2)}{\beta(\beta-3)(\beta-4)}$$

形状 ( $\alpha$ ) 和位置 ( $\beta$ ) 是分布参数。

参数计算:

**Pareto** 分布存在两个标准参数: 位置和形状。位置参数是变量的下界。在您选择了位置参数之后, 您可以估计出形状参数。形状参数的值为大于 0 的某个数, 通常是大于 1 的。形状参数越大, 方差越小, 分布的右尾就越粗。

参数输入要求:

位置可以为大于 0 的任意正值

形状参数  $\geq 0.05$

## 佩尔森 V 分布

佩尔森 V 分布与逆 Gamma 分布相关, 正好与 gamma 分布互补。佩尔森 V 分布也用于模拟递延时间, 当存在未确定的最小递延和最大递延。例如, 紧急服务到达时间或者机器修理的递延时间。Alpha (形状参数) 以及 Beta (尺度参数) 是分布的参数。

佩尔森 V 分布的数学表达式如下:

$$f(x) = \frac{x^{-(\alpha+1)} e^{-\beta/x}}{\beta^{-\alpha} \Gamma(\alpha)}$$

$$F(x) = \frac{\Gamma(\alpha, \beta/x)}{\Gamma(\alpha)}$$

$$\text{均值} = \frac{\beta}{\alpha-1}$$

$$\text{标准差} = \sqrt{\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}}$$

$$\text{偏度} = \frac{4\sqrt{\alpha-2}}{\alpha-3}$$

$$\text{超峰度} = \frac{30\alpha-66}{(\alpha-3)(\alpha-4)} - 3$$

输入要求:

Alpha (形状参数)  $> 0$

Beta (尺度参数)  $> 0$

## 佩尔森 VI 分布

佩尔森 VI 分布与 gamma 分布相关，是基于两个 gamma 分布的双变量有理函数。Alpha1（已知形状 1），Alpha2（已知形状 2），Beta（尺度参数）是该分布的参数。

的数学表达式如下：

$$f(x) = \frac{(x/\beta)^{\alpha_1-1}}{\beta B(\alpha_1, \alpha_2)[1+(x/\beta)]^{\alpha_1+\alpha_2}}$$

$$F(x) = F_B\left(\frac{x}{x+\beta}\right)$$

$$\text{均值} = \frac{\beta\alpha_1}{\alpha_2-1}$$

$$\text{标准差} = \sqrt{\frac{\beta^2\alpha_1(\alpha_1+\alpha_2-1)}{(\alpha_2-1)^2(\alpha_2-2)}}$$

$$\text{偏度} = 2\sqrt{\frac{\alpha_2-2}{\alpha_1(\alpha_1+\alpha_2-1)}}\left[\frac{2\alpha_1+\alpha_2-1}{\alpha_2-3}\right]$$

$$\text{超峰度} = \frac{3(\alpha_2-2)}{(\alpha_2-3)(\alpha_2-4)}\left[\frac{2(\alpha_2-1)^2}{\alpha_1(\alpha_1+\alpha_2-1)}+(\alpha_2+5)\right]-3$$

输入要求：

Alpha1(形状1)>0

Alpha2(形状2)>0

Beta(尺度参数)>0

## PERT 分布

PERT 分布被广泛的用于项目管理，定义了完成项目的最差的情形，一般情形，以及最优情形。该分布与 Beta 分布和三角分布相关。PERT 分布可以用来定义项目的风险和成本模型，以满足成本目标，最优成本，常规成本和最差成本。PERT 分布可以提供最接近正态和对数正态分布分布。类似三角分布，PERT 分布需要定义最有可能值，最小值和最大值。但是与三角分布不同的是，PERT 分布可以在最有可能值左右构建平滑曲线。使该分布的情形更接近于真实世界的要求。使生成的分布更接近于正态分布的曲线。最小值，最大值和最有可能值。

## Risk Simulator

PERT 分布的数学表达式如下:

$$f(x) = \frac{(x - \min)^{A1-1} (\max - x)^{A2-1}}{B(A1, A2)(\max - \min)^{A1+A2-1}}$$

$$\text{where } A1 = 6 \left[ \frac{\min + 4(\text{likely}) + \max}{6} - \min \right] \text{ and } A2 = 6 \left[ \max - \frac{\min + 4(\text{likely}) + \max}{6} \right]$$

and  $B$  is the Beta function

$$\text{均值} = \frac{\text{Min} + 4\text{Mode} + \text{Max}}{6}$$

$$\text{标准差} = \sqrt{\frac{(\mu - \text{Min})(\text{Max} - \mu)}{7}}$$

$$\text{偏度} = \sqrt{\frac{7}{(\mu - \text{Min})(\text{Max} - \mu)}} \left( \frac{\text{Min} + \text{Max} - 2\mu}{4} \right)$$

输入要求:

最小值  $\leq$  最有可能值  $\leq$  最大值 可以是正数, 负数, 或者 0

## Power 分布

Power 分布与指数分布相关, 较小的结果概率较大, 并且随着结果值变大指数递减。Alpha (形状参数) 是分布的唯一参数。

的数学表达式如下:

$$f(x) = \alpha x^{\alpha-1}$$

$$F(x) = x^\alpha$$

$$\text{均值} = \frac{\alpha}{1 + \alpha}$$

$$\text{标准差} = \sqrt{\frac{\alpha}{(1 + \alpha)^2 (2 + \alpha)}}$$

$$\text{偏度} = \sqrt{\frac{\alpha + 2}{\alpha}} \left( \frac{2(\alpha - 1)}{\alpha + 3} \right)$$

输入要求:

Alpha (形状)  $> 0$



Power 乘数转换分布

Power 分布与指数分布相关，较小的结果概率较大，并且随着结果值变大指数递减。Alpha（形状参数）是分布的唯一参数。

输入要求：

Alpha（形状）>0

位置参数可以是任何正数或者负数包括 0

参数>0

学生 t 分布

t 分布（学生分布）是在假设检验中使用的最广泛的一种分布。它被用于在样本空间很小的情况下，估计某正态分布总体的均值，还被用于检验两样本均值的统计显著性区别或是小样本空间的置信区间。

t 分布的数学表达式如下：

$$f(t) = \frac{\Gamma[(r+1)/2]}{\sqrt{r\pi} \Gamma[r/2]} (1+t^2/r)^{-(r+1)/2}$$

均值=0（这适用于自由度 r 取所有值的情况，除非分布转移到一个非 0 为中心的位置）

$$\text{标准差} = \sqrt{\frac{r}{r-2}}$$

偏度 = 0 (适用于自由度为 r 的所有值)

$$\text{超峰度} = \frac{6}{r-4}, r > 4$$

$$t = \frac{x - \bar{x}}{s}, \Gamma \text{ 是 gamma 函数。}$$

自由度 r 是唯一的分布参数。

t 分布与 F 分布的关系如下：自由度为 r 的 t 分布的值的平方，服从自由度为 1 和 r 的 F 分布。除了它更扁幅度更广或是它的尖顶峰度（更肥胖的尾部和更尖的中部）以外，t 分布的概率密度函数的形状与均值为 0，方差为 1 的正态分布变量的形状类似。当自由度增加时（例如超过 30），t 分布就趋近于均值为 0，方差为 1 的正态分布。

参数输入要求：

自由度必须为大于 1 的整数

三角分布

三角分布描述的是已知最小值，最大值以及最大似然值时的一种情形。例如，您可以通过过去汽车销售数量的最大值、最小值、以及每周的一般销售数量来描述现在的汽车销售数量。

三角分布的三个先决条件是：

- 项目的最小值是固定的。
- 项目的最大值是固定的。
- 项目的最大似然值落在最小值和最大值之间，形成一个三角形的分布，显示在最大值和最小值附近发生的概率小于在最大似然值附近发生的概率。

三角分布的数学表达式如下：

$$f(x) = \begin{cases} \frac{2(x - \text{Min})}{(\text{Max} - \text{Min})(\text{Likely} - \text{Min})} & \text{for } \text{Min} < x < \text{Likely} \\ \frac{2(\text{Max} - x)}{(\text{Max} - \text{Min})(\text{Max} - \text{Likely})} & \text{for } \text{Likely} < x < \text{Max} \end{cases}$$

$$\text{均值} = \frac{1}{3}(\text{Min} + \text{Likely} + \text{Max})$$

标准差

$$= \sqrt{\frac{1}{18}(\text{Min}^2 + \text{Likely}^2 + \text{Max}^2 - \text{MinMax} - \text{MinLikely} - \text{MaxLikely})}$$

偏度

$$= \frac{\sqrt{2}(\text{Min} + \text{Max} - 2\text{Likely})(2\text{Min} - \text{Max} - \text{Likely})(\text{Min} - 2\text{Max} + \text{Likely})}{5(\text{Min}^2 + \text{Max}^2 + \text{Likely}^2 - \text{MinMax} - \text{MinLikely} - \text{MaxLikely})^{3/2}}$$

超峰度 = -0.6 (适用于所有的最大, 最小, 最有可能值)

最小值 (Min), 最有可能值 (Likely) 和最大值 (Max) 是分布的三个参数。

参数输入要求:

最小值 ≤ 最有可能值 ≤ 最大值, 且可以取任意值

但是, 最小值 < 最大值, 且可以取任意值

### 均匀分布

在均匀分布情况下, 所有值都落在最大值和最小值之间的范围内且发生的概率相同。条件:

- 最小值是固定的。
- 最大值是固定的。
- 在最大值和最小值范围内的值的发生概率是相同的。

均匀分布的数学表达式如下:

$$f(x) = \frac{1}{\text{Max} - \text{Min}} \quad \text{Min} < \text{Max}$$

$$\text{均值} = \frac{\text{Min} + \text{Max}}{2}$$

$$\text{标准差} = \sqrt{\frac{(\text{Max} - \text{Min})^2}{12}}$$

偏度 = 0 (此值对于任意最大值和最小值均成立)

超峰度 = -1.2 (此值对于任意最大值和最小值均成立)

此值对于任意最大值和最小值均成立

最小值 (Min) 和最大值 (Max) 是分布的两个参数。

参数输入要求: 最小值 < 最大值, 且可以取任意值

### 韦伯分布 (Rayleigh 分布)

**Weibull** 分布用于描述来自寿命和衰变测试的数据满足的概率分布。通常被用于描述可靠性分析中的出错次数，包括材料和质量控制中的强度测试。**Weibull** 分布也用于表现不同的物理量，比如音速。**Weibull** 分布是一族分布的统称，可以用于表现很多事物的特性。例如，通过选取形状参数，**Weibull** 分布可以用于模拟指数分布（形状参数=1.0）和 **Rayleigh** 分布（形状参数=2.0）等等。当形状参数=1.0 时，通过设置 **Weibull** 分布中心位置尺度参数（beta），可以建立起始点不在原点（0，0）的指数分布。当形状参数小于 1.0 时，**Weibull** 分布成削尖的曲线。制造业可以使用这个分布描述零件在烧制过程中的出错率。

韦伯分布的数学表达式如下：

$$f(x) = \frac{\alpha}{\beta} \left[ \frac{x}{\beta} \right]^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}$$

$$\text{均值} = \beta \Gamma(1 + \alpha^{-1})$$

$$\text{标准差} = \beta^2 \left[ \Gamma(1 + 2\alpha^{-1}) - \Gamma^2(1 + \alpha^{-1}) \right]$$

$$\text{偏度} = \frac{2\Gamma^3(1 + \beta^{-1}) - 3\Gamma(1 + \beta^{-1})\Gamma(1 + 2\beta^{-1}) + \Gamma(1 + 3\beta^{-1})}{\left[ \Gamma(1 + 2\beta^{-1}) - \Gamma^2(1 + \beta^{-1}) \right]^{3/2}}$$

$$\text{超峰度} = \frac{-6\Gamma^4(1 + \beta^{-1}) + 12\Gamma^2(1 + \beta^{-1})\Gamma(1 + 2\beta^{-1}) - 3\Gamma^2(1 + 2\beta^{-1}) - 4\Gamma(1 + \beta^{-1})\Gamma(1 + 3\beta^{-1}) + \Gamma(1 + 4\beta^{-1})}{\left[ \Gamma(1 + 2\beta^{-1}) - \Gamma^2(1 + \beta^{-1}) \right]^2}$$

形状（ $\alpha$ ）和中心位置尺度（ $\beta$ ）是分布的两个参数， $\Gamma$  是  $\gamma$  函数。

参数输入要求：

形状参数  $\alpha \geq 0.05$

尺度参数  $\beta$  可以为大于 0 的任意正值

### Weibull 乘数转换分布

**Weibull** 分布用于描述来自寿命和衰变测试的数据满足的概率分布。通常被用于描述可靠性分析中的出错次数，包括材料和质量控制中的强度测试。**Weibull** 分布也用于表现不同的物理量，比如音速。**Weibull** 分布是一族分布的统称，可以用于表现很多事物的特性。例如，通过选取形状参数，**Weibull** 分布可以用于模拟指数分布（形状参数=1.0）和 **Rayleigh** 分布（形状参数=2.0）等等。当形状参数=1.0 时，通过设置 **Weibull** 分布中心位置尺度参数（beta），可以建立起始点不在原点（0，0）的指数分布。当形状参数小于 1.0 时，**Weibull** 分布成削尖的曲线。制造业可以使用这个分布描述零件在烧制过程中的出错率。形状参数（ $\alpha$ ）和尺度参数（ $\beta$ ）是分布的重要参数。

输入变量要求：

形状参数  $\geq 0.05$

中心位置控制尺度  $\text{Beta} > 0$  可以是任何正值

位置参数可以是任何正数或者负数包括 0. 参数  $> 0$

### 3. 预测

预测是基于历史数据对未来做出的合理估计，或是在不存在历史数据的情况下对未来进行的揣摩。当存在历史数据时，最好使用定量方法（或统计学方法），但是当不存在历史数据时，定性法（或判断法）就是唯一的选择了。下图 3.1 列出了最常见的预测方法。

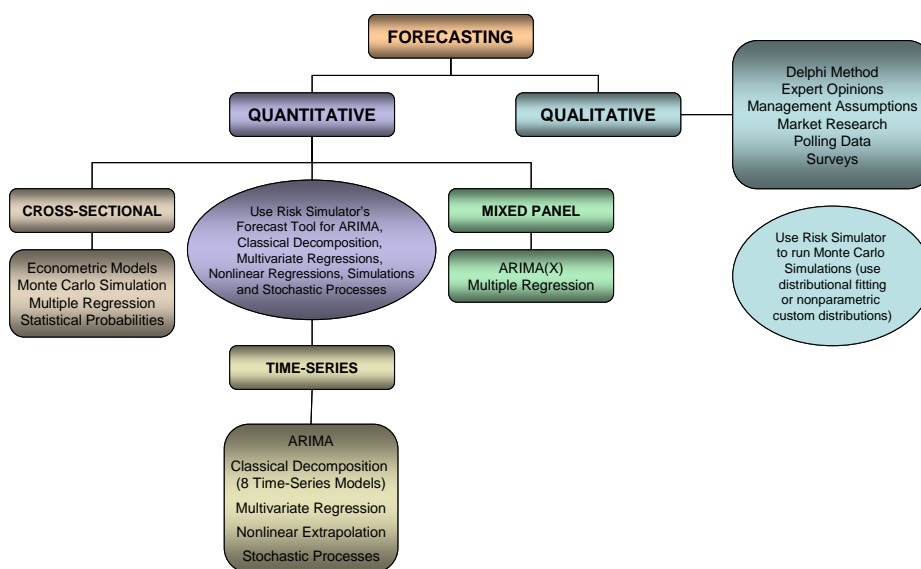


图 3.1 预测方法

#### 3.1 不同类型的预测方法

一般来说，预测的方法可以分为定性和定量两种。当几乎不存在可靠的历史数据，同期数据或可比数据时，就采用定性法。定性法包含有德尔菲法或专家意见法（行业专家、市场专家，或内部成员达成一致的预测）、管理假设法（由高级管理层设置的目标增长率），还有市场研究、外部数据、投票和调查（从第三方如行业或地区指标获得的数据，或是积极的市场调研）等。这些估计可以是单点估计值（一般意见）或是一系列预测值（预测分布）。后者可以在 Risk Simulator 中输入一个自定义分布，并对预测结果进行仿真。意思就是可以通过预测的数值本身得到一个分布来进行一次非参数仿真。

定量法中的可用数据或是需要预测的数据可以分为时间序列数据（涉及时间因素的变量，如不同年份的收益、通胀率、利率、市场份额、失败率等），截面数据（与时间无关的变量，如某一年份不同地区大二学生的平均绩点，每个学生的 SAT 分数水平，每周酒精饮料的消费量等），和混合面板数据（时间序列数据和面板数据的混合体，例如，在给定市场成本预算及市场份额的前提下，预测未来十年的销售量，这意味着销售数据是一个时间序列的外生变量，将销售成本及市场份额作为模型参数有助于仿真预测）。

**风险模拟**软件为用户提供以下的各个预测方法：

1. ARIMA（自回归和移动平均）
2. 自动 ARIMA
3. 计量经济学自动分析功能
4. 计量经济学基本功能
5. 神经网络
6. 三次样条插值法
7. GARCH（广义自回归条件异方差）
8. J-曲线
9. S-曲线
10. 组合模糊逻辑
11. 马尔可夫链
12. 最大似然法
13. 非线性外推
14. 回归分析
15. 随机过程
16. 时间序列分析
17. 趋势线

对于每个不包含在用户手册中的预测方法的分析细节，读者如果想了解更多细节，请参看由 Johnathan Mun 博士撰写的《*风险建模：应用蒙特卡罗模拟，实物期权分析，预测与优化技术*》(Wiley Finance 2006)。他还是**风险模拟**软件的开发者。然而，以下还阐述一些更加常用的方法。所有其它的预测方法可以在**风险模拟**软件中相当容易地进行应用。

### 3.2 运行 Risk Simulator 中的预测工具

一般说来，在开始预测之前先进行以下几个步骤：

- 打开 Excel 表格输入历史数据或是打开已有的历史数据表格
- 选择数据，点击**仿真选择预测**
- 选择相关项（ARIMA、多元回归、非线性外推法、随机预测、时间序列分析），然后输入相关的参数

图 3.2 显示了预测工具及多种预测方法



图 3.2 Risk Simulator 的预测方法

接下来的部分将对每种方法作一个简短的介绍，并提供了一些软件使用的例子。例子使用的数据文件可以通过以下路径获得：**开始|程序|Real Option Valuation|Risk Simulator|示例**。

### 3.3 时间序列分析

理论：

图 3.3 列出了八种最常见的时间序列模型，通过季节性和趋势性来分类。例如，如果数据没有季节性和趋势性，那么用单滑动平均模型或单光滑指数模型就足够了。但是如果存在季节性但是没有表现明显的趋势性，那么最好使用季节附加模型或是季节乘积模型。

	无季节性	有季节性
无趋势性	单滑动平均模型	季节附加模型
	单光滑指数模型	季节乘积模型
有趋势性	双滑动平均模型	Holt-Winter 附加模型
	双光滑指数模型	Holt-Winter 乘积模型

表 3.3 八种最常见的时间序列方法

步骤:

- 打开 Excel，如果需要的话，打开您的历史数据（下面的例子中使用的是示例文件夹中的时间序列预测文件）
- 选择历史数据（数据必须列在同一列中）
- 选择仿真|预测|时间序列分析
- 选择要应用的模型，输入相关的参数，点击**确定**

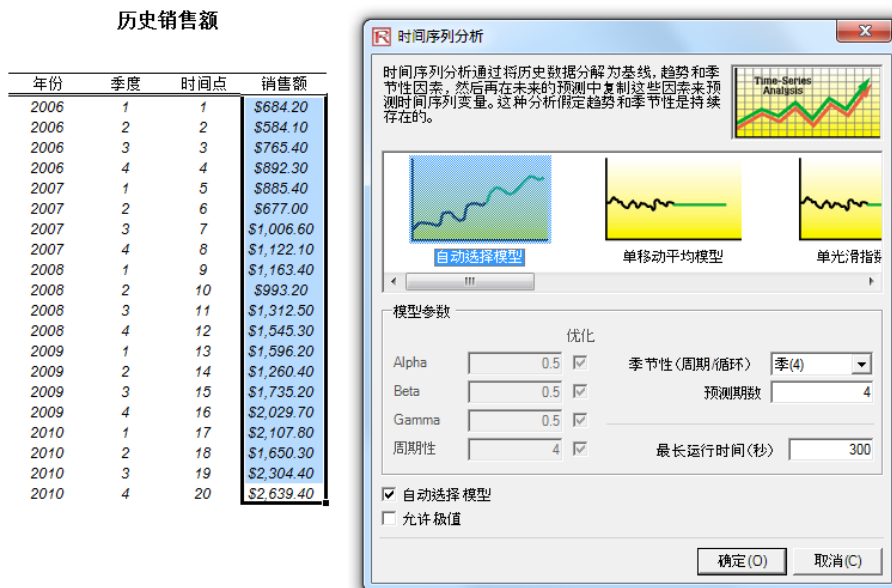


图 3.4 时间序列分析

结果解析:

图 3.5 中所示是使用预测工具生成的结果样本。使用的是 Holt-Winter 乘积模型。注意到图 3.5 中的模型拟合和预测图显示出 Holt-Winter 乘积模型很好地表现了趋势性和季节性。时间序列分析报告会提供相关的参数  $\alpha$ ， $\beta$ ， $\gamma$  的最优值，误差度量，拟合数据，预测值，以及拟合-预测图等。参数仅供参考。 $\alpha$  针对的是基准水平随时间变化的记忆影响， $\beta$  是趋势参数，衡量了趋势的强度， $\gamma$  衡量了历史数据的季节性强度。分析首先将历史数据分解为三个因素，然后又将这三个因素重组起来预测未来的数据。拟合数据利用重组模型分析了历史数据和拟合数据，表现了预测和过去的接近程度（这种技术叫倒推）。预测值可能是单点估计或假设（如果选择了自动生成假设选项，并存在一个仿真文档）。图中显示了历史值，拟合值以及预测值。图表是一种有效的交流途径，并在视觉上为我们显示了预测模型的效果。

注意:

如图 3.3 所示时间序列分析模块包含八个时间序列模型。您可以基于趋势和季节性来选择特定的模型运行，或是选择自动选择模型选项，它会自动在八种方法中重复匹配，最优化参数，最终找出这些数据的最佳拟合模型。如果您选择了八个模型的其中之一，您也可以不选择最优化这一检验栏，而是输入您自己的  $\alpha$ ， $\beta$ ， $\gamma$  参数。想要了解更多关于这些参数的技术细节，请参考 Dr. Johnathan Mun 的 *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《风险建模：应用蒙特卡洛模拟，实物期权分析，预测及最优化》，第二版，Wiley 2010。如果您选择自动选择模型或是其它任何季节性模型，您需

要输入相关的季节性周期。输入的必须是一个正整数（例如，如果数据是季节性的，就输入 4 作为一年的季节周期，如果是月度数据就输入 12）。接下来输入预测的期数。同样这个数也必须是正整数。设置的最长运行时间是 300 秒。一般来说，这个不需要修改。但是如果运行大批量历史数据时，分析需要的时间可能会比较长，如果运行时间超过了这个最长时间，过程就会中止。您也可以选择预测自动生成假设。也就是说，预测可以通过假设的分布进行而不是单点估计。最后，选择极值参数令您可以在优化  $\alpha$ ， $\beta$ ， $\gamma$  参数时，包含 0 值和 1 值。某些预测软件允许这些极值参数，但是有些不允许。Risk Simulator 允许您在这两者之间进行选择。一般情况下不必选择允许极值参数。

Holt-Winter 乘积

统计汇总

Alpha, Beta, Gamma	均方根误差	Alpha, Beta, Gamma	均方根误差
0.00, 0.00, 0.00	914.824	0.00, 0.00, 0.00	914.824
0.10, 0.10, 0.10	415.322	0.10, 0.10, 0.10	415.322
0.20, 0.20, 0.20	187.202	0.20, 0.20, 0.20	187.202
0.30, 0.30, 0.30	118.795	0.30, 0.30, 0.30	118.795
0.40, 0.40, 0.40	101.794	0.40, 0.40, 0.40	101.794
0.50, 0.50, 0.50	102.143		

假设参数  $\alpha = 0.2429$ ,  $\beta = 1.0000$ ,  $\gamma = 0.7797$ , and seasonality = 4

时间序列分析汇总

当原始数据中体现出季节性和趋势性时，更高级的模型将数据分解成三部分：数据的基本水平其权重由参数  $\alpha$  表示，趋势性部分其权重由参数  $\beta$  表示，季节性部分其权重由参数  $\gamma$  表示。诚然，在这方面有很多方法，但普遍使用的是 Holt-Winters 季节附加模型和 Holt-Winters 附加模型而言，数据基本水平、季节性部分和趋势性部分是在一块进行预测的。

寻找平均预测的最佳拟合测试利用均方根误差 (RMSE)，RMSE 计算的是拟合值与实际数据之间的均方差的平方根。

均方误差 (MSE 统计量) 是使用误差 (实际的历史数据和模型拟合的数值之差) 平方的方法来剔除误差符号的影响。得到的一种绝对误差度量。MSE 统计量使用平方，倾向于使误差大的点在总个数中所占权重更大而误差小的点所占权重更小，还可用来比较不同的时间序列模型的优劣。均方根误差 (RMSE) 也叫二次根均方差是由 MSE 统计量开平方得到，是最常用的一种误差度量。RMSE 统计量可以被看成预测误差的均值的绝对值，当预测的误差所带来的损失和预测的误差绝对值大小成比例时，这是非常有用的。RMSE 统计量被用来判断最佳时间序列拟合的一种标准。

平均绝对百分比误差 (MAPE 统计量) 计算相对历史数据的平均绝对百分比，是一种相对误差统计量度量，当预测误差所造成的损失和误差百分比更有关而不是大小时，该统计量是很有用的。最常用的一种有用的误差度量是 Theil's U 统计量。该统计量是对模型表现的检测，当 Theil's U 统计量小于 1.0 时，表示该模型预测提供的估计从统计上来讲比实际观测要好。

时间点	真实值	拟合值	误差度量
1	684.20		均方根误差 71.8132
2	584.10		均方误差 5157.1348
3	765.40		平均绝对误差: 53.4071
4	892.30		平均绝对误差: 4.50%
5	885.40	684.20	Theil's U 0.3054
6	677.00	667.55	
7	1006.60	935.45	
8	1122.10	1198.09	
9	1163.40	1112.48	
10	993.20	887.95	
11	1312.50	1348.38	
12	1545.30	1546.53	
13	1596.20	1572.44	
14	1260.40	1299.20	
15	1735.20	1704.77	
16	2029.70	1976.23	
17	2107.90	2026.01	
18	1650.30	1637.28	
19	2304.40	2245.93	
20	2639.40	2643.09	
预测 21		2713.69	
预测 22		2114.79	
预测 23		2900.42	
预测 24		3293.81	

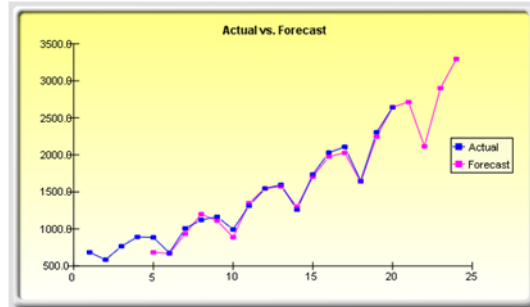


图 3.5 Holt-Winter 预测报告示例



### 3.4 多元回归

理论：

假设使用者对回归分析的基础知识已有充分的了解。一般的双变量线性回归的形式是： $Y = \beta_0 + \beta_1 X + \varepsilon$ ，其中  $\beta_0$  代表截距， $\beta_1$  代表斜率， $\varepsilon$  是误差项。之所以称其为双变量是因为它只存在两个变量，因变量  $Y$ ，自变量  $X$ ，同时  $X$  也被称为回归量（有时候双变量回归也被称为单变量回归，因为只存在一个自变量  $X$ ）。之所以取名为因变量是因为它受自变量影响，例如，销售收益与产品广告和促销带来的市场成本量有关，那么销售收益就是因变量，市场成本就是自变量。双变量回归的一个例子可以看二维空间中如图 3.6 中的左图，简单地在一系列数据点中插入一条最佳拟合直线。还有些例子是多元回归，就是存在多个或是  $n$  个自变量，此时的回归表达式为  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon$ 。在这个例子中，最佳拟合线就位于一个  $n+1$  维的空间中了。

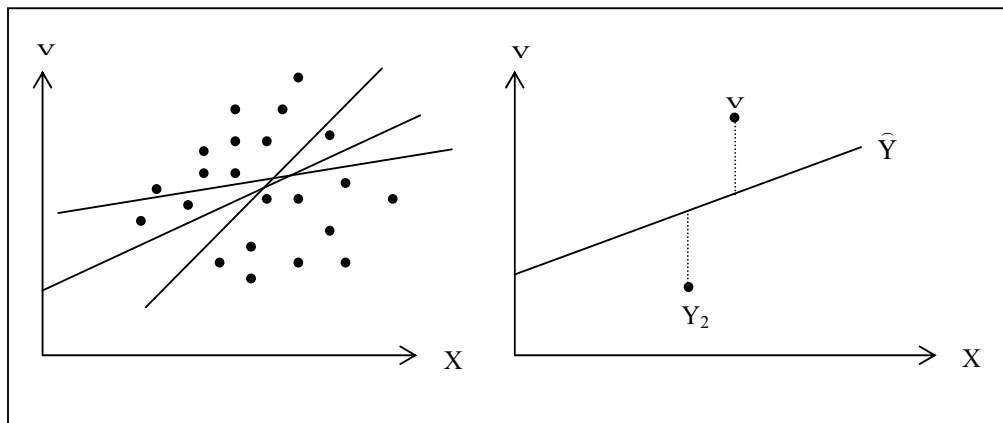


图 3.6 双变量回归

但是类似于图 3.6，对点状图中的一系列数据用直线进行拟合，可能会有很多结果。

总体垂直误差（也就是图 3.6 右图中显示的真实数据点  $(Y_i)$  和估计直线  $(\hat{Y})$  之间的距离绝对值之和）最小的那条线就被称为最佳拟合直线。为了找出让误差最小的最佳拟合直线，我们需要一种更加娴熟的方法，那就是回归分析。回归分析通过最小化总体误差来找出唯一的最佳拟合直线。这是通过计算如下方程式

$$\text{Min} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

来实现的。

只有唯一一条直线可以使误差平方和最小。我们将误差（真实数据点和预测线之间的垂直距离）平方以避免正负误差之间相互抵消。为了解决这个涉及截距和斜率的最小化问题，我们需要通过它的一阶导数值为 0 的性质，来得到方程：

$$\frac{d}{d\beta_0} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = 0 \quad \text{and} \quad \frac{d}{d\beta_1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = 0$$

这样就可以得到双变量回归的最小二乘等式：

$$\beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n X_i Y_i - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n}}{\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

对于多元回归，可以依此类推计算多个因变量，例如，

$Y_i = \beta_1 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i$ , 斜率的估计可以通过如下式子来计算：

$$\hat{\beta}_2 = \frac{\sum Y_i X_{2,i} \sum X_{3,i}^2 - \sum Y_i X_{3,i} \sum X_{2,i} X_{3,i}}{\sum X_{2,i}^2 \sum X_{3,i}^2 - \left(\sum X_{2,i} X_{3,i}\right)^2}$$

$$\hat{\beta}_3 = \frac{\sum Y_i X_{3,i} \sum X_{2,i}^2 - \sum Y_i X_{2,i} \sum X_{2,i} X_{3,i}}{\sum X_{2,i}^2 \sum X_{3,i}^2 - \left(\sum X_{2,i} X_{3,i}\right)^2}$$

在运行多元回归时，对求解过程和解释结果需要更加注意。例如，在建立一个合适的模型之前需要对计量建模的知识有很好的掌握（识别回归的缺陷如结构断点，多元共线性、异方差性、自相关、规格测试和非线性等）。关于多元回归以及如何识别回归缺陷的详情分析及讨论请参考 *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《风险建模：应用蒙特卡洛模拟、实物期权分析、预测及最优化》，第二版（Wiley 2010）。

#### 步骤：

- 打开 Excel，如果需要的话打开您的历史数据（下面的图示使用的是示例文件夹中的多元回归文件）
- 检查以确定数据都排在同一列，选择包括变量名称的整个数据区域，然后选择**仿真|预测|多元回归**
- 选择自变量并检查相关选项（滞后，逐步回归，非线性回归等），然后点击**确定**

#### 结果解析：

图 3.8 是一份多元回归的结果样本报告。这份报告很完整，包括所有的回归结果，变量结果的分析，拟合图以及假设检验结果。对于这些结果技术细节的解释不包括在本手册范围内。更多关于多元回归及回归报告的解释分析及讨论请参考 Dr. Johnathan Mun 的 *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《风险建模：应用蒙特卡洛模拟，实物期权分析，预测及最优化》，第二版（Wiley 2010）。

多元回归

Y	X1	X2	X3	X4	X5
521	18308	185	4.041	79.6	7.2
367	1148	600	0.55	1	8.5
443	18068	372	3.665	32.3	5.7
365	7729	142	2.351	45.1	7.3
614	100484	432	29.76	190.8	7.5
385	16728	290	3.294	31.8	5
286	14630	346	3.287	678.4	6.7
397	4008	328	0.666	340.8	6.2
764	38927	354	12.938	239.6	7.3
427	22322	266	6.478	111.9	5
153	3711	320	1.108	172.5	2.8
231	3136	197	1.007	12.2	6.1
524	50508	266	11.431	205.6	7.1
328	28886	173	5.544	154.6	5.9
240	16996	190	2.777	49.7	4.6
286	13035	239	2.478	30.3	4.4
285	12973	190	3.685	92.8	7.4
569	16309	241	4.22	96.9	7.1
96	5227	189	1.228	39.8	7.5
498	19235	358	4.781	489.2	5.9
481	44487	315	6.016	767.6	9
468	44213	303	9.295	163.6	9.2
177	23619	228	4.375	55	5.1
198	9106	134	2.573	54.9	8.6
458	24917	189	5.117	74.3	6.6
108	3872	196	0.799	5.5	6.9
246	8945	183	1.578	20.5	2.7
291	2373	417	1.202	10.9	5.5
68	7128	233	1.109	123.7	7.2
311	23624	349	7.73	1042	6.6
606	5242	284	1.515	12.5	6.9
512	92629	499	17.99	381	7.2
426	28795	231	6.629	136.1	5.8
47	4487	143	0.639	9.3	4.1
265	48799	249	10.847	264.9	6.4
370	14067	195	3.146	45.8	6.7
312	12693	288	2.842	29.6	6
222	62184	229	11.882	265.1	6.9



1. 选择数据范围包括数据的标题(B5:G55)
2. 点击Risk Simulator | 预测 | 多元回归
3. 选择因变量(在本例中就是变量Y)然后进行一系列对模型的优化(滞后回归, 非线性回归, 逐步回归)然后点击确定。查看分析报告作为分析的结果。



图 3.7 运行一个多元回归

回归分析报告

回归统计量	
R-方 (决定系数)	0.3272
调整后的R平方	0.2508
多元R (多元相关系数)	0.5720
估计的标准差 (SEy)	149.6720
样本数	50

R方或可决系数表明在此次回归分析中自变量对因变量的解释程度。但是, 在多元回归中, 调整的R方还考虑到了某些额外的自变量或回归量的存在, 将R方调整到一个更加精确的水平, 以增加回归的解释说明能力。因此, 只有因变量才能被回归量解释。

多元相关系数 (多元R) 度量的是真实因变量 (已知的历史结果) 与基于回归方程式的估计或拟合变量 (拟合的结果) 的相关性。它是可决系数 (R方) 的平方根。

估计的标准差描述的是数据点向上和向下偏离回归线或平面的程度。这个值稍后可用于计算和估计置信区间。

回归结果	截距	学士学位	人均警察支出 (百万)	人口 (人/平方米)	失业率
系数	57.9555	-0.0035	0.4644	25.2377	-0.0086
标准差	108.7901	0.0035	0.2535	14.1172	0.1016
t统计量	0.5327	-1.0066	1.8316	1.7877	-0.0843
p值	0.5969	<b>0.3197</b>	<b>0.0738</b>	<b>0.0807</b>	<b>0.9332</b>
下限95%	-161.2966	-0.0106	-0.0466	-3.2137	-0.2132
上限95%	277.2076	0.0036	0.9753	53.6891	0.1961

自由度	假设检验	
回归自由度	5	临界t值 (99%的置信度, 自由度为44)
残差自由度	44	临界t值 (95%的置信度, 自由度为44)
总自由度	49	临界t值 (90%的置信度, 自由度为44)
		2.6923
		2.0154
		1.6802

系数提供了回归模型的截距和斜率。举例来说, 这些系数提供了对真实值的一种估计: 因变量 (人口数 b) 的值可通过回归方程  $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$  来估计。标准差反映了预测的系数的精确程度, t统计量是预测的系数和它的标准差的比例。

t统计量被用在假设检验中。这里我们设置零假设: 系数的均值为0, 备择假设: 系数的均值不为0。运用t检验并将计算的t统计量和残差自由度的临界值进行比较。t检验是非常重要的, 它的计数值反映了当存在额外的回归量时, 每个系数是否在统计上是有效的。也就是说, t检验在统计上说明了一个回归量或独立变量是否应该保留在回归模型中还是从模型中剔除。

如果系数的t统计量超过了相关自由度t统计量的估计临界值, 这些系数在统计上就是有效的。这里, 三种常用的置信水平是90%, 95%和99%。如果系数的t统计量超过了临界值, 就认为该系数是统计上有效的。作为选择, p值计算了取到t统计量的概率, 也就是说t值越小, 所对应的系数就越有效。P值对应90%, 95%, 99%三个置信水平最常用的有效临界值选为0.01, 0.05和0.10。

变量分析						
	平方和	算术平方	F统计量	p值	假设检验	
回归	479388.4898	95877.6980	4.2799	0.0029	F统计量临界值 (99%置信度, 自由度为4和3)	3.4651
残差	985675.1902	22401.7089			F统计量临界值 (95%置信度, 自由度为4和3)	2.4270
总和	1465063.6800				F统计量临界值 (90%置信度, 自由度为4和3)	1.9828

方差分析表 (ANOVA) 提供了对回归模型总体统计有效性的F检验。与t检验关注单个回归量不同的是, F检验关注所有估计系数的统计属性。F检验计算的是回归的均值的平方和残差的均值的平方的比例。分子表明回归模型对预测值的解释程度, 分母表明有多少程度模型不能解释的。因此, F-Statistic越大表明我们的模型就越有效。相应的p值也被计算用来假设检测, 判断回归模型的整体有效性。这里零假设为: 所有的系数都为0, 备择假设: 所有系数不同时为0。如果p值在0.05或0.10的alpha有效水平下, 比0.01小, 则回归模型是有效的。同样的方法在对比计算的F-Statistic值和在各种有效水平下的临界值的F值时也能应用。

预测			
时间点	真实值 (Y)	预测值 (F)	误差 (E)
1	521	299.5124	221.4876
2	367	487.1243	(120.1243)
3	443	353.2789	89.7211
4	365	276.3296	88.6704
5	614	776.1336	(162.1336)
6	385	298.9993	86.0007
7	286	354.8718	(68.8718)
8	397	312.6155	84.3845
9	764	529.7550	234.2450
10	427	347.7034	79.2966
11	153	266.2526	(113.2526)
12	231	264.6375	(33.6375)
13	524	406.8009	117.1991
14	328	272.2226	55.7774
15	240	231.7882	8.2118
16	286	257.8862	28.1138
17	285	314.9521	(29.9521)
18	569	335.3140	233.6860
19	96	282.0356	(186.0356)
20	498	370.2062	127.7938
21	481	340.8742	140.1258
22	468	427.5118	40.4882
23	177	274.5298	(97.5298)
24	198	294.7795	(96.7795)
25	458	295.2180	162.7820
26	108	269.6195	(161.6195)
27	246	195.5955	50.4045
28	291	364.5004	(73.5004)

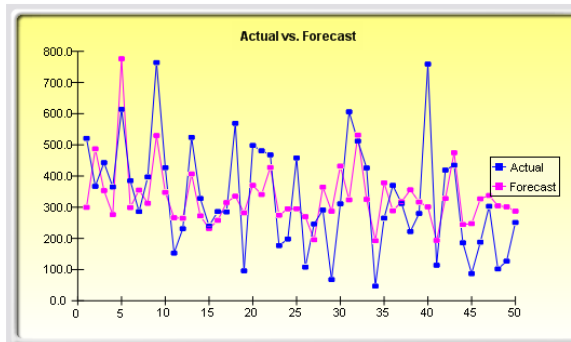


图 3.8 多元回归的结果

### 3.5 随机预测

理论:

随机过程其实是一个随时间的变化而产生一系列结果的数学方程式,从本质上说,它的结果是不确定的。这个过程或方程式不遵循任何简单的可辨规则,例如价格每年上升 X 个百分点,或由于这个因素收益每年上升 X 乘以 Y 个百分点,这里的 X, Y 都可以看成随机过程。随机过程在定义上是不确定的,您可以在随机过程方程式中随意输入参数并且每次得到不同的结果。例如,股票价格的路径(股票的走势)在本质上是随机的,没有人可以确定预测股票价格的路径。但是随时间变化的价格却是来源于某个过程。这个过程是事先确定的,但是产生的结果不确定。因此通过随机仿真,我们可以创造出多条价格路径,得到这些仿真的一个统计取样,然后在给定生成时间序列的随机过程的参数和特性的情况下,推断出真实价格可能遵循的潜在路径。Risk Simulator 的预测工具包含三个最基本的随机过程:布朗运动或随机游走,均值回复过程,跳跃扩散过程。布朗运动由于其简单性和广泛应用性成为最流行的随机过程。

在随机过程仿真中比较有趣的是历史数据并不是必需的。也就是使用这个模型并不要求对历史数据进行拟合。只需要计算出期望值和历史数据的波动率,或通过可比较的外部数据来得到估计值,以及对变量进行假定后,就能使用该模型。参见 Dr.Johnathan Mun 的 *Modeling Risk, Second Edition: Applying Monte Carlo*

*Simulation, Real Options Analysis, Forecasting, and Optimization* 《风险建模：应用蒙特卡洛模拟，实物期权分析，预测及最优化》，第二版，Wiley 2010 来查看关于如何计算每个输入量的详细说明（例如均值回复率，跳跃概率，波动率等等）。

步骤：

- 通过**仿真|预测|随机过程**按钮开启模型
- 选择想要的过程，输入参数，点击更新图表来保证过程与您的预期一致，然后点击**确定**（见图 3.9）

结果解析：

图 3.10 中显示的是一个随机过程示例结果。图表显示了迭代结果，报告中阐述了随机过程的基本知识。此外，还提供了每阶段的预测值（均值和标准差）。通过这些值您可以决定哪些时间段与您的分析是相关的，并用这些正态分布的均值和标准差设定输入假设。然后可以使用这个输入假设在您自定义的模型中进行仿真。

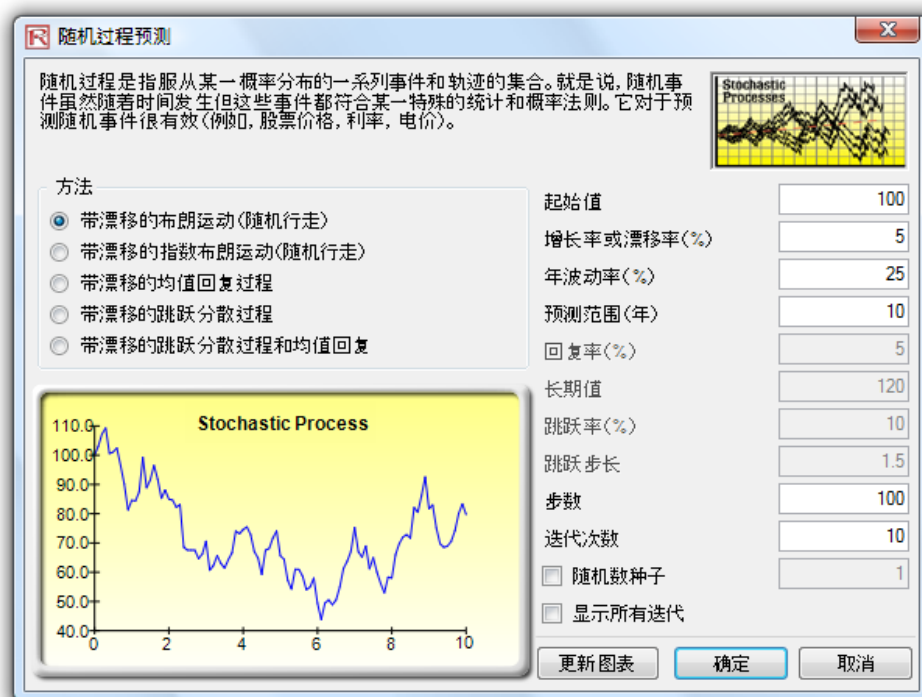


图 3.9 随机过程预测

随机过程预测

统计汇总

所谓随机过程是指从某一概率分布的一系列事件和轨迹的集合。就是说，随机事件虽然随时间发生但这类事件都符合某一特殊的统计和概率法则。我们主要提到的随机过程包括随机游走（布朗运动）、均值回复、跳跃-扩散过程。这些过程能用来预测大多数体现随机倾向但又服从概率分布的变量的变化。

随机游走（布朗运动）能用来预测股票的价格、商品的价格和任何沿带漂移路径有带漂移或增长和波动率的随机时间序列数据。均值回复可以通过长期目标水平来减小随机游走的波动。该过程可以用来预测如利率，通胀这些有长期目标水平（这些长期目标水平由权威机构和市场价格）的时间序列变量。跳跃-扩散过程可以用来预测如石油价格、电力价格（个别外部事件的发生能使价格往上跳跃或下降）这些偶尔伴有随机跳跃的时间序列数据。最后，这三类随机过程可根据需要精工配合使用。

右边的结果显示的是在每次时间步长迭代后的均值和标准差。如果您选中了显示所有迭代按钮，那么每一个迭代产生的路径将会显示在一个单独的工作簿中。在下面的图表里给出了几个生成的迭代路径。

时间	均值	标准差
0.0000	100.00	0.00
0.1000	99.10	7.47
0.2000	96.03	7.22
0.3000	94.37	13.59
0.4000	97.39	15.57
0.5000	99.50	17.01
0.6000	97.79	20.92
0.7000	102.23	25.54
0.8000	106.54	26.54
0.9000	102.34	21.16
1.0000	102.77	20.86
1.1000	103.30	22.41
1.2000	103.27	19.23
1.3000	103.02	23.61
1.4000	97.78	19.65
1.5000	96.84	20.53
1.6000	100.92	25.22
1.7000	105.18	26.90
1.8000	100.75	30.33
1.9000	101.20	29.71
2.0000	103.67	36.95
2.1000	108.09	42.76
2.2000	111.58	42.61
2.3000	111.25	41.54
2.4000	108.47	35.22
2.5000	107.13	32.56
2.6000	108.95	32.95
2.7000	114.64	38.78
2.8000	114.13	36.61
2.9000	114.37	35.91
3.0000	114.33	39.90
3.1000	112.63	39.94
3.2000	115.11	39.89
3.3000	117.64	42.82
3.4000	114.70	39.91
3.5000	115.52	43.45
3.6000	117.60	49.89
3.7000	120.21	51.94
3.8000	116.64	53.52
3.9000	118.70	56.12
4.0000	113.19	56.71
4.1000	109.09	58.33
4.2000	103.70	52.23
4.3000	108.41	53.12
4.4000	108.67	56.30
4.5000	105.96	52.42
4.6000	106.12	55.80
4.7000	107.70	55.11
4.8000	109.43	58.43
4.9000	114.50	59.64
5.0000	110.44	53.91
5.1000	109.68	53.96

随机过程：带漂移的布朗运动（随机游走）

初始值	100	步长	100.00	跳跃率	N/A
漂移率	5.00%	迭代次数	10.00	跳跃大小	N/A
波动率	25.00%	回复率	N/A	随机数种子	1431155157
水平	10	逾期值	N/A		

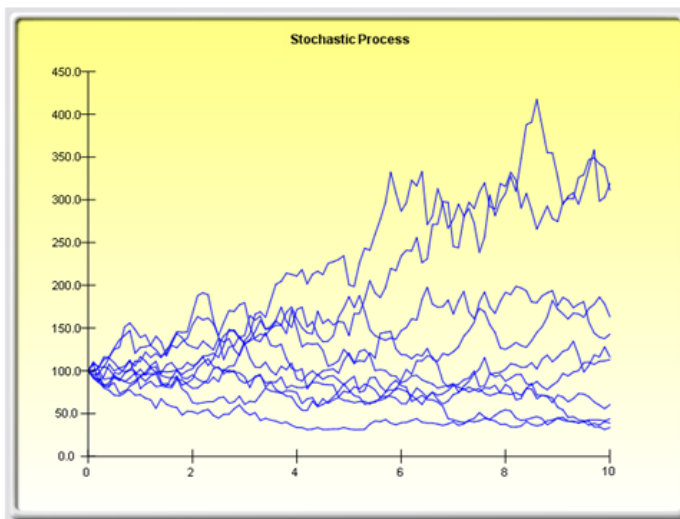


图 3.10 随机预测结果

3.6 非线性外推法

理论：

外推法是通过历史数据的趋势用统计学投影的方法来预测未来。它只能被用于时间序列预测中。对于截面数据或混合平板数据（时间序列和截面数据）来说，多元回归更加适用。这种方法在不期望发生较大的变化，即期望因果关系保持不变时，或某一情景下的因果关系没有明确被确定时比较有用。它还可以防止将个人偏见引入过程之中。外推法相对比较可靠，简单，所付的代价也相对较小。但是由于外推法是基于近期的发展趋势将和历史趋势保持一致的假设之上的，所以如果在投影期间内发生了中断，那么将会产生大量的预测误差。也就是说，时间序列的纯粹外推假设我们所需要的知道的所有信息都包含在被预测序列的历史数据之中。如果我们假设过去的行为是未来行为的一个好的预报器，那么外推法就是适用的。故当所需要知道的所有信息是一些短期预测时，外推法不失为一种有效的方法。

外推法估计了任意 x 值的 f(x)函数，通过在所有 x 值间插入一条光滑的非线性曲线，利用这条曲线，外推出一个基于此历史数据集之外的未来的 x 处 f(x)的值。

外推法可以采用多项式函数的形式或是有理数函数的形式（两多项式之比）。一般情况下，对于性质好的数据多项式函数形式就足够了，但是有时候使用有理数函数形式会更加精确（尤其是对于极值函数来说，例如，分母趋近于 0 的函数）。

步骤:

- 打开 Excel 表格，如果需要的话，打开您的历史数据（下面的例子使用的是示例文件夹
- 中的 *非线性外推* 的例子）
- 选择时间序列数据，并选择**仿真|预测|非线性外推**
- 选择外推类型（自动选择，多项式函数或有理数函数），并输入所需要预测时间段（图 3.11），点击**确定**

结果解析:

图 3.12 中的结果报告显示了外推预测值，测量误差和外推结果的图示。误差度量法应当被用于检测预测的有效性，在比较外推法和时间序列分析的预测质量和精度时尤为重要。

注意:

当历史数据很光滑，并遵循某种非线性模式和曲线时，外推法比时间序列分析法更有效。但是，当数据模式显示出季节性周期和趋势的时候，时间序列分析则可以提供更好的结果。

历史销售收入 多项式增长率				历史净收益 正弦增长率			
年份	月份	时期	销售	年份	月份	时期	收益
2004	1	1	\$1.00	2004	1	1	\$84.15
2004	2	2	\$6.73	2004	2	2	\$90.93
2004	3	3	\$20.52	2004	3	3	\$14.11
2004	4	4	\$45.25	2004	4	4	(\$75.68)
2004	5	5	\$83.59	2004	5	5	(\$95.89)
2004	6	6	\$138.01	2004	6	6	(\$27.94)
2004	7	7	\$210.87	2004	7	7	\$65.70
2004	8	8	\$304.44	2004	8	8	\$98.94
2004	9	9	\$420.89	2004	9	9	\$41.21
2004	10	10	\$562.34	2004	10	10	(\$54.40)
2004	11	11	\$730.85	2004	11	11	(\$100.00)
2004	12	12	\$928.43	2004	12	12	(\$53.66)
				2005	1	13	\$42.02
				2005	2	14	\$99.06
				2005	3	15	\$65.03
				2005	4	16	(\$28.79)
				2005	5	17	(\$96.14)
				2005	6	18	(\$75.10)

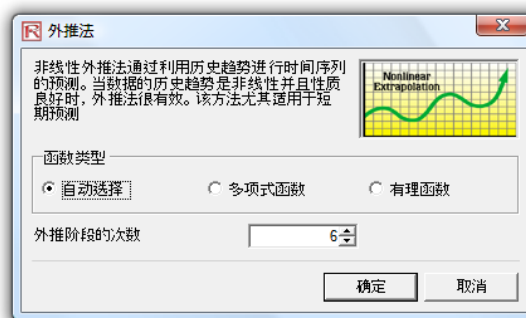


图 3.11 运行非线性外推

非线性外推

统计汇总

外推法运用了利用一段历史的发展方向来投影未来的发展的统计投影法。这只能被用于时间序列预测。对于区域交叉或者混合（时间序列和区域交叉）的数据，多变量回归法更为恰当。这个方法当主要变化是难以预测的时候，也就是说引起变化的因素为常数或者引起变化的因素所处的情况不能完全被了解的情况下很有效。它也有助于消除在整个过程中个人偏见的引入。外推法是很可以信赖的，相对简单的以及耗资小的。尽管如此，外推假设了现在和历史的发展方向将继续，在时间投影不连续的情况下这将导致很大的预测误差。那就是说，纯的时间序列外推法假定我们用来预测所需要的所有信息都包括在历史数据中。如果我们假定过去的行为是未来的良好预测者，外推法是很好的。它是对于所有预测都为短期的时候的一种很有效的途径。

外推法估计了任意 $x$ 值的 $f(x)$ 函数，通过在所有 $x$ 值间插入一条光滑的非线性曲线，利用这条曲线，外推出一个基于此历史数据集之外的未来的 $x$ 处 $f(x)$ 的值。外推法可以采用多项式函数的形式或是有理函数的形式（两多项式之比）。一般情况下，对于性质好的数据多项式函数形式就足够了，但是有时候使用有理函数形式会更加精确（尤其是对于极值函数来说，例如，分母趋近于零函数）。

时间点	真实值	预测拟合值	估计误差	误差度量
1	1.00			均方根误差 19.6799
2	6.73	1.00		均方误差 387.2974
3	20.52	-1.42	-8.15	平均绝对误差 10.2095
4	45.25	99.82	119.36	平均绝对误差 31.56%
5	83.59	55.92	-46.67	Theil's U 1.1210
6	138.01	136.71	14.39	
7	210.87	211.96	1.69	
8	304.44	304.43	-0.41	函数类型: 有理
9	420.89	420.89	0.01	
10	562.34	562.34	0.00	
11	730.85	730.85	0.00	
12	928.43	928.43	0.00	
预测13		1157.03	0.00	
预测14		1418.57	0.00	
预测15		1714.95	0.00	
预测16		2048.00	0.00	
预测17		2419.55	0.00	
预测18		2831.39	0.00	

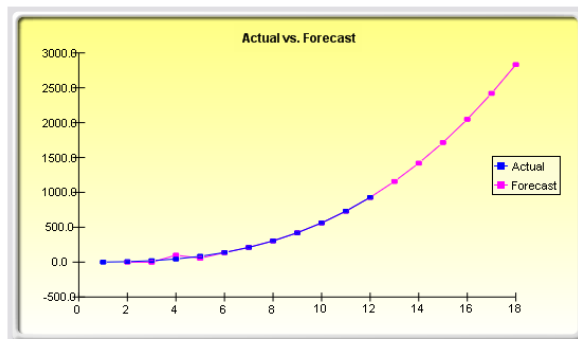


图 3.12 非线性外推的结果

### 3.7 Box-Jenkins ARIMA 高级时间序列

理论:

一种非常有效的高级时间序列预测的工具是 ARIMA 或自回归求和滑动平均过程。ARIMA 预测是由三种独立的工具集成一种综合性模型。第一个工具是自回归或“AR”项，它对应的是无条件预测模型中残差的滞后值。本质上，这个模型抓住了预测模型真实数据的历史方差，然后利用这个方差或残差来生成一个更好的预测模型。第二个工具是求和法或“I”项。这里求和法指对时间序列数据差分的总和。这个因素可以解释存在于数据中的任何非线性增长。第三个工具是滑动平均或“MA”项，它实际上就是滞后预测误差的滑动平均。通过引入这个滞后预测误差，模型可以通过滑动平均的计算来修正预测误差或错误。ARIMA 模型遵循 Box-Jenkins 方法，其中的每一项都应对于模型构建中所采用的步骤，直到只剩下随机因素。同时 ARIMA 模型在生成预测时还使用了相关性技术。ARIMA 模型可用于模拟那些在数据绘图中并不显著的样品。除此以外，ARIMA 模型还可与外生变量混合，但是要确保外生变量有足够多的数据点来覆盖需要预测的额外期数。最后需要注意鉴于模型的复杂性，可能所需的运行时间也比较长。

关于 ARIMA 模型优于普通时间序列分析和多元回归的原因有很多。时间序列分析和多元回归的普遍发现是残差项与它们自身的滞后值是相关的。这种序列相关违背了回归理论的标准假设，即扰动项之间是互不相关的。由序列相关所引发的主要问题有：



- 对各种不同的线性估计量，回归分析和基本的时间序列分析并不有效。但是，由于残差项可用于预测现有残差，我们可以利用这一信息优势，使用 ARIMA 模型来预测因变量。
- 利用回归和时间序列公式计算的标准差是不正确的，通常都被低估了，如果有滞后因变量被设置为回归量，回归估计就是有偏和不一致的，但是使用 ARIMA 就可以将它们固定。

自回归求和滑动平均过程或 ARIMA (p,d,q) 模型是 AR 模型的延伸，使用 3 个成分在

时间序列数据中模拟序列相关性。第一个组成成分是自回归 (AR) 项。AR (p) 模型在等式中使用了时间序列的 p 阶滞后。AR (p) 模型采用如下这种形式： $y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t$ 。第二个组成成分是和法 (I)。每一种和法对应于时间序列数据的不同差分的求和。如，I (1) 意味着数据的一次差分的求和。I (d) 意味着差分数据 d 次差分的求和。第三个成分是滑动平均 (MA) 项。MA (q) 模型利用预测误差的 q 阶滞后来改进预测结果。MA (q) 模型的形式： $y_t = e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}$ ，最后 ARIMA (p,d,q) 模型的综合形式是：

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}$$

步骤：

- 打开 Excel 表格，输入数据或是打开工作表中现有的用于预测的历史数据（下图中使用的是示例文件夹中的*时间序列 ARIMA* 例子）
- 选择时间序列数据，选择**仿真|预测|ARIMA**
- 输入相关的参数 P, D 和 Q（只能为正整数），然后输入需要预测的期数，点击**确定**

结果解析：

对于 ARIMA 模型结果解释的大部分说明与多元回归分析结果的解释一致（请参见《模拟风险》，Dr.Johnathan Mun 第二版来获得更多关于多元回归分析及 ARIMA 模型结果解释的技巧）。但是如图 3.14 中，有几条额外的关于 ARIMA 分析的解释。第一条是赤池信息准则 (AIC) 和史瓦兹信息准则 (SC)，它们通常被用于 ARIMA 模型选择和识别。也就是说 AIC 和 SC 被用于确定某个含 p, d 和 q 参数值的具体模型是否是一个好的统计拟合。SC 提供了对是否添加额外变量更强有力的判断，通常应该选择那些 AIC 值和 SC 值最低的模型。最后，ARIMA 报告里还提供了一份额外的自相关 (AC) 和部分自相关 (PAC) 的统计结果。

例如，如果自相关 AC (1) 为非零值，意味着序列是一阶序列相关的。如果 AC 随着滞后的增加呈几何下降趋势，这意味着序列遵循一个低阶自回归过程。如果经过几次滞后之后 AC 值趋于 0，这意味着序列遵循一个低阶滑动平均过程。相反的，PAC 衡量了在移除了滞后干扰后的 k 阶相关值。如果自相关模式可以用小于 k 阶的自回归描述，那么 k 阶滞后的部分自相关值趋近于 0。报告里同时还提供了 Ljung-Box 的 k 阶滞后 Q 统计值和 p 值，此时被检验的零假设是 k 阶时不存在自相关。自相关图里的虚线近似表示了两个标准差的区间界限。如果自相关值在此范围之内，那么在 5% 的显著性水平内它不显著区别于 0。寻找到合适的 ARIMA 模型需要尝试和经验。AC, PAC, SC 和 AIC 都是识别正确模型的有效诊断工具。



图 3.13A Box-Jenkins ARIMA 预测工具

ARIMA (自回归和滑动平均)

回归统计量

R方 (可决系数)	0.9999	赤池信息准则 (AIC)	4.6213
调整后的R方	0.9999	史瓦兹信息准则 (SC)	4.6632
多元R方 (多元相关系数)	1.0000	最大似然值	-1005.1340
估计的标准差	297.5246	杜宾 (DW) 统计量	1.8588
样本个数	435	迭代次数	5

自回归和滑动平均过程简称ARIMA (p, d, q) 模型扩展了AR模型, 使用了三个部分来描述时间序列数据的关系。第一部分是自回归 (AR) 项。在AR (p) 模型中使用p来延迟时间序列。AR (p) 的形式为:  $y(t) = a(1) * y(t-1) + \dots + a(p) * y(t-p) + e(t)$ 。第二部分是求和 (d) 阶数项。每个求和阶数对应时间序列的差分阶数。l(1) 为使用一阶差分。l(d) 为使用d阶差分。第三部分是移动平均 (MA) 项。MA (q) 模型使用加上q项预测误差的延迟来改进预测。MA (q) 模型的形式为:  $y(t) = e(t) + b(1) * e(t-1) + \dots + b(q) * e(t-q)$ 。最后, ARMA (p, q) 的形式为:  $y(t) = a(1) * y(t-1) + \dots + a(p) * y(t-p) + e(t) + b(1) * e(t-1) + \dots + b(q) * e(t-q)$ 。

可决系数R方描述了本回归中自变量对因变量的解释程度。但是在多元回归中, 调整的R方也考虑了其它自变量和回归量的存在, 使得调整后的R方对回归的解释能力增强。但是在某些ARIMA仿真情况中 (例如, 非收敛模型), R方值就不可靠了。

多元相关系数 (多元R) 度量了真实因变量 (Y) 和基于回归等式的估计或拟合值 (Y) 的相关程度。它是决定系数 (R方) 的平方根。

估计标准差描述的是数据在回归线或平面周围上下偏移的离散。这个值稍后可用于计算和估计置信区间。

AIC和SC通常用于模型选择中。SC提供了对是否添加额外变量更强有力的判断。通常使用者应该选择那些AIC值和SC值最低的模型。

Durbin-Watson统计量用来判断回归分析中的残差项是否存在自相关性。总体上来说, DW小于2表明含有正自相关性。

回归结果

	截距	AR(1)	MA(1)
系数	-0.0626	1.0055	0.4936
标准差	0.3108	0.0066	0.0420
t统计量	-0.2013	1691.1373	11.7633
p值	0.8406	0.0000	0.0000
5%下限	0.4498	1.0065	0.5628
95%上限	-0.5749	1.0046	0.4244

自由度

	自由度	假设检验	
回归自由度	2	临界t值 (95%置信度, 自由度为432)	2.5873
残差自由度	432	临界t值 (99%置信度, 自由度为433)	1.9655
总体自由度	434	临界t值 (90%置信度, 自由度为434)	1.6484

系数提供了回归模型的截距和斜率。举例来说, 这些系数提供了对真实值的一种估计; 变量 (人口 b) 的值可通过回归方程  $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$  来估计。标准差反映了预测的系数的精确程度, t统计量是预测的系数和它的标准差的比例。

t统计量被用在假设检验中。这里我们设置零假设: 系数的均值为0。各备假设: 系数的均值不为0。运用t检验并将计算的t统计量和残差自由度的临界值进行比较。t检验是非常重要的, 它的数值反映了当存在额外的回归量时, 每个系数是否在统计上是有效的。也就是说, t检验在统计上说明了一个回归量或独立变量是否应该保留在回归模型中还是从模型中剔除。

如果系数的t统计量超过了相关自由度t统计量的估计临界值, 这些系数在统计上就是有效的。这里, 三种常用的置信水平是90%, 95%和99%。如果系数的t统计量超过了临界值, 就认为该系数是统计上有效的。作为选择, p值计算了取到t统计量的概率, 也就是说p值越小, 所对应的系数就越有效。p值对应90%, 95%, 99%三个置信水平最常用的有效临界值选为0.01, 0.05和0.10。

蓝色的p值表明在90%的置信水平或0.10 alpha水平下, 其对应的系数是统计上有效的。而红色的p值表明其对应的系数在任何 alpha水平下都不是统计上有效的。

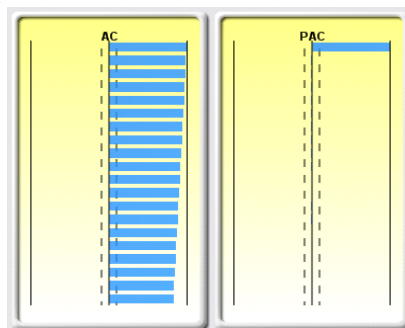
方差分析

	平方和	均方	F统计量	p值	假设检验	
回归	38415447.5277	19207723.7638	3171851.1034	0.0000	临界t值 (99%置信度, 自由度为432)	4.6546
残差	2616.0549	6.0557			临界t值 (95%置信度, 自由度为433)	3.0166
总和	38418063.5826	19207729.8195			临界t值 (90%置信度, 自由度为434)	2.3149

方差分析表 (ANOVA) 提供了对回归模型总体统计有效性的F检验。与t检验关注单个回归量不同的是, F检验关注所有估计系数的统计属性。F检验计算的是回归的均值的平方和残差的均值的平方的比例。分子表明回归模型对预测值的解释程度, 分母表明有多少程度模型不能解释的。因此, F-Statistic越大表明我们的模型就越有效。相应的p值也被计算用来假设检测, 判断回归模型的整体有效性。这里零假设为: 所有的系数都为0。各备假设: 所有系数不同时为0。如果p值在0.05或0.10的alpha有效水平下, 比0.01小, 则回归模型是有效的。同样的方法在对比计算的F-Statistic值和在各种有效水平下的临界F值时也能应用。

自相关

时间延迟	AC	PAC	下届	上届	Q统计量	概率
1	0.9921	0.9921	(0.0958)	0.0958	431.1216	-
2	0.9841	(0.0105)	(0.0958)	0.0958	856.3037	-
3	0.9760	(0.0109)	(0.0958)	0.0958	1,275.4818	-
4	0.9678	(0.0142)	(0.0958)	0.0958	1,688.5499	-
5	0.9594	(0.0098)	(0.0958)	0.0958	2,095.4625	-
6	0.9509	(0.0113)	(0.0958)	0.0958	2,496.1572	-
7	0.9423	(0.0124)	(0.0958)	0.0958	2,890.5594	-
8	0.9336	(0.0147)	(0.0958)	0.0958	3,278.5669	-
9	0.9247	(0.0121)	(0.0958)	0.0958	3,660.1152	-
10	0.9156	(0.0139)	(0.0958)	0.0958	4,035.1192	-
11	0.9066	(0.0049)	(0.0958)	0.0958	4,403.6117	-
12	0.8975	(0.0068)	(0.0958)	0.0958	4,765.6032	-
13	0.8883	(0.0097)	(0.0958)	0.0958	5,121.0697	-
14	0.8791	(0.0087)	(0.0958)	0.0958	5,470.0032	-
15	0.8698	(0.0064)	(0.0958)	0.0958	5,812.4256	-
16	0.8605	(0.0056)	(0.0958)	0.0958	6,148.3694	-
17	0.8512	(0.0062)	(0.0958)	0.0958	6,477.8620	-
18	0.8419	(0.0038)	(0.0958)	0.0958	6,800.9622	-
19	0.8326	(0.0003)	(0.0958)	0.0958	7,117.7709	-
20	0.8235	0.0002	(0.0958)	0.0958	7,428.3952	-



如果自相关项AC(k)不为0, 意味着该数据序列是一阶自相关的。如果AC(k)随着延迟的增加越来越小, 则表明该数据序列服从一个低阶自回归过程。如果当延迟较小时, AC(k)为0, 则意味着该序列数据服从一个低阶移动平均过程。偏相关PAC(k)度量的是带有k个时间点延迟的数据序列和不带延迟的数据序列之间的相关性。如果它们的自相关性能被阶数小于k的自回归过程捕捉, 那么在延迟k上的自相关性将会趋向于0。Ljung-Box statistics统计量和它们的p值有零假设: 当阶数大于k时, 不存在自相关性。左图的两条虚线是关于自相关性的边界(通过近似标准差的值来确定边界)。如果自相关性在这个边界里, 则说明在5%有效性水平下, 自相关性可近似视为0。

预测

时间点	真实值(Y)	预测值(F)	误差(E)
2	139.4000	139.6056	(0.2056)
3	139.7000	140.0069	(0.3069)
4	139.7000	140.2586	(0.5586)
5	140.7000	140.1343	0.5657
6	141.2000	141.6948	(0.4948)
7	141.7000	141.6741	0.0259
8	141.9000	142.4339	(0.5339)
9	141.0000	142.3587	(1.3587)
10	140.5000	141.0466	(0.5466)
11	140.4000	140.9447	(0.5447)
12	140.0000	140.8451	(0.8451)
13	140.0000	140.2946	(0.2946)
14	139.9000	140.5663	(0.6663)
15	139.8000	140.2823	(0.4823)
16	139.6000	140.2726	(0.6726)
17	139.6000	139.9775	(0.3775)
18	139.6000	140.1232	(0.5231)
19	140.2000	140.0513	0.1487
20	141.3000	140.9862	0.3138
21	141.2000	142.1738	(0.9738)
22	140.9000	141.4377	(0.5377)
23	140.9000	141.3513	(0.4513)
24	140.7000	141.3939	(0.6939)
25	141.1000	141.0731	0.0270
26	141.6000	141.8311	(0.2311)
27	141.9000	142.2065	(0.3065)
28	142.1000	142.4709	(0.3709)
29	142.7000	142.6402	0.0598
30	142.9000	143.4561	(0.5561)
31	142.9000	143.3532	(0.4532)
32	143.5000	143.4040	0.0960
33	143.8000	144.2784	(0.4784)
34	144.1000	144.2966	(0.1966)
35	144.8000	144.7374	0.0626
36	145.2000	145.5692	(0.3692)
37	145.2000	145.7582	(0.5582)
38	145.7000	145.6649	0.0351
39	146.0000	146.4605	(0.4605)
40	146.4000	146.5176	(0.1176)
41	146.8000	147.0891	(0.2891)
42	146.6000	147.4066	(0.8066)
43	146.5000	146.9501	(0.4501)
44	146.6000	147.0255	(0.4255)

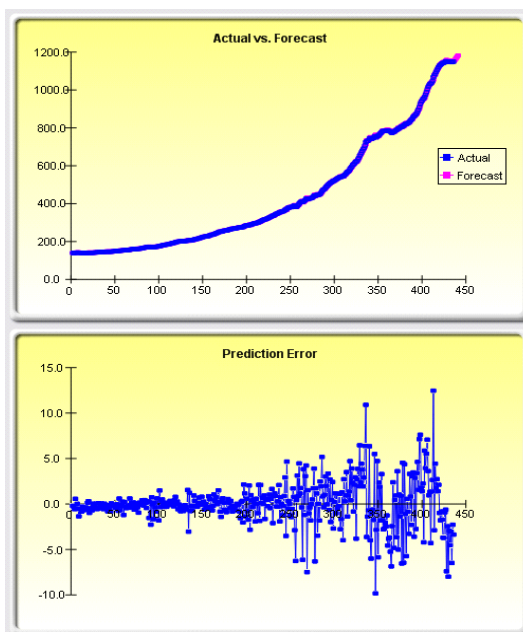


图 3.13B Box Jenkins ARIMA 预测报告

### 3.8 自回归求和移动平均模型(Box-Jenkins ARIMA 高级时间序列)

理论:

该工具和 ARIMA 模块一样提供了相同的模型，不同点就是自回归求和移动平均模型可以通过自动检验传统模型中多个参数的排列方式，从而找到最优拟合的模型。运行自回归求和移动平均模型的方式和普通的 ARIMA 预测方式相同。不同点就是不再需要输入模型的参数 P,D,Q，而且对这些参数的组合与比较都是自动完成的。

操作过程描述:

- 运行 Excel 输入数据或者打开包含历史数据需要预测的工作簿（如图 3.14 显示的示例文件）。[高级预测模型](#)在 Risk Simulator 的[示例文件](#)中)
- 在 ARIMA 预测 工作簿中,选择 Risk Simulator | 预测 |自回归求和移动平均
- 点击链接图标选择现有的时间序列数据输入需要预测的期数点击 **确定**



图 3.14 自回归求和移动平均模型

### 3.9 基本计量经济学模型

理论:

计量经济学涉及一系列的商业分析，建模和预测技术用于模拟或者预测某个商业或者经济变量。运行基本的计量经济学模型类似于一般的回归分析，而基本的计量经济学模型中的变量和因变量在进行回归前需要被修正。生成的报告类似于前面的多元回归部分，对于报告的解释也类似于多元回归部分的解释。

操作过程描述:

- 运行 Excel 输入数据或者打开包含历史数据需要预测的工作簿 (如图 3.15 显示的示例文件)。**高级预测模型**在 Risk Simulator 的**示例文件**中
- 选择基本计量经济学数据 工作簿中的数据列, 点击 Risk Simulator |预测 |基本计量经济学模型
- 输入自变量和因变量 (如图 3.15 所示) 点击**确定**运行模型和生成报告, 或者点击**显示结果**, 在生成报告之前来预览结果, 以免有对模型需要改进的地方。

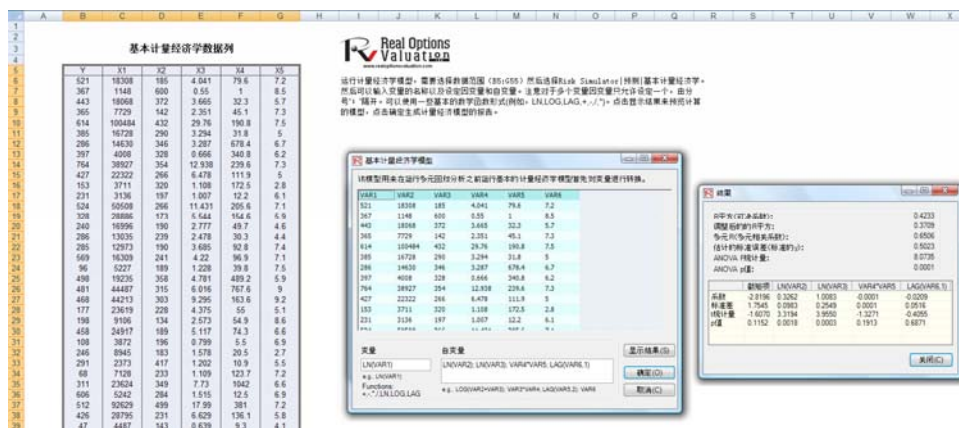


图 3.15 基本计量经济学模块

### 3.10 J-S 曲线预测

理论:

J-曲线或者指数增长曲线描述了这样的增长现象, 下一期的增长取决于本期增长的水平, 而下一期将呈指数式地增长。这意味着随着时间的增长, 从一段时间到另一段时间, 数值将很快地增加。该模型一般用于预测随着时间变化生物的增长和化学的裂变。

操作过程描述:

- 运行 Excel 然后选择 Risk Simulator |预测 |JS 曲线
- 选择 J 或者 S 曲线的类型输入需要的输入假设(如图 3.16 和 3.17)点击**确定**运行模型和生成报告

指数J增长曲线

在数学上，指数增长的数量意味着那些增长率与现有的规模总是成一定的比例。这个增长率遵循几何增长的法则。这说明对于任何指数增长的数量，数量越大，增长的越快。但是这也说明变量的大小和增长率都遵循这个严格的定理，即某个固定的比例。指数增长背后的原理是，数值越大，增长的越快。任何指数增长的数量在某个固定的增长时间内都以某个固定的比例增长。这种预测方法称之为J曲线，因为曲线的形状很像字母J。其它的增长曲线包括S曲线和马尔可夫链。

生成J曲线预测，需要遵循下面的步骤：

1. 点击 Risk Simulator | 预测 | JS 曲线
2. 选择指数J曲线输入需要的输入参数  
(例如，初始值为100，增长率为5%，终止期为100)
3. 点击确定运行预测，然后查看生成的报告

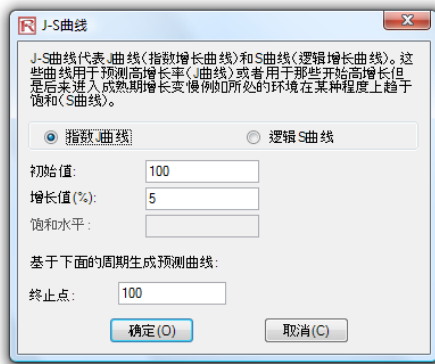
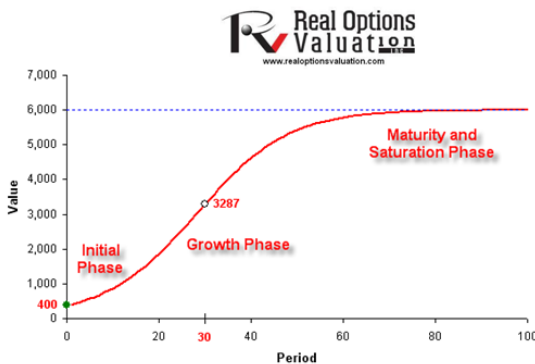


图 3.16 J-曲线预测

S-曲线或者逻辑增长曲线开始启动的时候类似于 J-曲线，具有指数式地增长。随着时间地增长，环境不断地饱和（例如，市场的饱和，竞争的加剧，堵塞），增长变得缓慢，最后预测值终止于饱和值或者最大水平。这个模型一般用于预测市场份额或者新产品从引入期到衰退期的销售增长变化。图3.17显示了示例的S-曲线。

逻辑S曲线

逻辑函数或逻辑曲线可对某些变量的s增长曲线建模。在初始极端的的增长相当于指数增长，随着竞争不断升级，增长率不断地降低，当处于饱和期时停止增长。这些函数可在生物学到经济学的很多领域广泛运用。举例来说，在胚胎生长阶段，受精卵分裂，细胞数快速增长：1，2，4，8，16，32，64等等。这是指数增长。但是胎儿只能长到子宫所能容纳的大小：这样别的因素就会减小细胞数的增长，增长率就下降了（当然，胎儿仍在成长）。经过一段时间后，胎儿被产出并且保持成长。最后，细胞数达到稳定；个体的高度变为常数，在成熟期，成长停止了。同样的原理可被用到人类的人口增长和动物的种群增长理论中，也可用到市场渗透理论和产品的收益当中。在市场渗透中有初始的增长率，但随着竞争的加强，市场的增长率下降，在成熟期市场达到饱和并保持它的规模。



1. 点击 Risk Simulator | 预测 | JS 曲线
2. 选择需要的输入参数(看下面的例子)
3. 确定然后查看预测报告

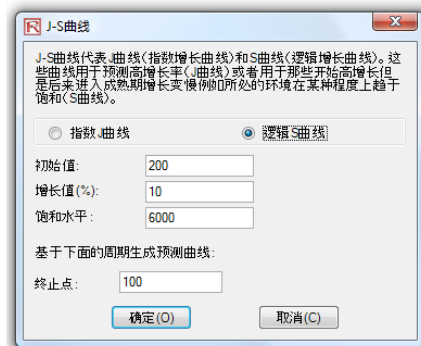


图 3.17 S-曲线预测

### 3.11 GARCH 波动率预测

理论:

GARCH 模型可用在计算流动资产和可买卖资产的波动率上，比如金融期权里的股票的波动率；该模型有时也用在另外的可买卖资产上，比如石油和电力的价格。其缺点是需要大量数据，需要高级金融模型专家的意见并且易受人为操作的影响。该模型的优点是使用了严格的统计分析来得到最优拟合的波动率曲线，提供了随时间变化的波动率估计。更多对于 GARCH 模型的讨论超出了用户手册的范围。关于 GARCH 模型 的描述请参考 Johnathan Mun 博士所编著的“Advanced Analytical Models,” (Wiley 2008)一书。

操作过程描述:

- 运行 Excel 然后打开示例文件高级预测模型，转到 广义自回归条件异方差模型 (GARCH) 工作簿选择 Risk Simulator | 预测 | GARCH
- 点击链接图标选择数据范围区域，输入需要的输入假设（如图 3.18）点击 **确定**运行模型和生成报告。

注意：一般的波动率预测要求  $P = 1, Q = 1$ ，周期数=每年的期数（12 用于月数据，52 用于周数据，252 或者 365 用于日数据），基数=1 到周期数之间的某个值，预测期=想要预测的波动率的期数。



(GARCH)广义自回归条件异方差模型(GARCH)

日期	输入变量
1	459.11
2	460.71
3	460.34
4	460.68
5	460.83
6	461.68
7	461.66
8	461.64
9	465.97
10	469.38
11	470.05
12	469.72
13	466.95
14	464.78
15	465.81
16	465.86
17	467.44
18	468.32
19	470.39
20	468.51
21	470.42
22	470.4
23	472.78
24	478.64
25	481.14
26	480.81
27	481.19
28	480.19
29	481.46

运行GARCH模型，输入相应的时间序列数据，然后点击Risk Simulator|预测|GARCH然后点击数据链接图标，选择历史数据范围(例如，C8:C2428)。输入要求的输入变量(例如，P 1，Q 1，日交易周期为252，预测基数为 1，预测期为10)然后点击确定。然后查看生成的预测报告。



图 3.18 GARCH 波动率预测



Risk Simulator

	$z_t \sim \text{Normal}$	$z_t \sim \text{T}$
GARCH-M	$y_t = c + \lambda \sigma_t^2 + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$	$y_t = c + \lambda \sigma_t^2 + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$
GARCH-M	$y_t = c + \lambda \sigma_t + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$	$y_t = c + \lambda \sigma_t + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$
GARCH-M	$y_t = c + \lambda \ln(\sigma_t^2) + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$	$y_t = c + \lambda \ln(\sigma_t^2) + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$
GARCH	$y_t = x_t \gamma + \varepsilon_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$
EGARCH	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\ln(\sigma_t^2) = \omega + \beta \cdot \ln(\sigma_{t-1}^2) +$ $\alpha \left[ \frac{\varepsilon_{t-1}}{\sigma_{t-1}} - E( \varepsilon_t ) \right] + r \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$ $E( \varepsilon_t ) = \sqrt{\frac{2}{\pi}}$	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\ln(\sigma_t^2) = \omega + \beta \cdot \ln(\sigma_{t-1}^2) +$ $\alpha \left[ \frac{\varepsilon_{t-1}}{\sigma_{t-1}} - E( \varepsilon_t ) \right] + r \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$ $E( \varepsilon_t ) = \frac{2\sqrt{\nu-2} \Gamma((\nu+1)/2)}{(\nu-1)\Gamma(\nu/2)\sqrt{\pi}}$
GJR-GARCH	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 +$ $r \varepsilon_{t-1}^2 d_{t-1} + \beta \sigma_{t-1}^2$ $d_{t-1} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}$	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 +$ $r \varepsilon_{t-1}^2 d_{t-1} + \beta \sigma_{t-1}^2$ $d_{t-1} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}$

### 3.12 马尔可夫链

理论:

当未来的状态的概率水平取决于前期状态的概率水平，从长远的角度将它们彼此链接就行成了一条链，这条链就被称为马尔可夫链。该方法一般被用来预测具有市场份额的两个竞争对手。需要输入输入变量为顾客在第一家商店（第一种状态）的概率值，顾客下次可能会去同一家商店，也可能转向竞争对手的商店。

操作过程描述:

- 运行 Excel 然后选择 Risk Simulator | 预测 | 马尔可夫链
- 输入需要输入的假设（如图 3.19）然后点击 **确定** 运行模型和报告。

#### 马尔可夫链预测

马尔可夫过程在连续时间段的多次和重复试验演化系统的研究中非常有用。该系统在特定时间的状态是未知的，而我们感兴趣的是出现这种特定状态的概率。举例来说，马尔可夫链可用于计算特定的机器或设备将在下个时期继续工作的概率，或者计算客户购买了A产品将在下个时期继续购买A产品或选中购买B品牌产品的竞争性产品的概率。这些品牌或产品记作处于某种状态。通过输入经过一段时间个体仍在状态1的概率和现在在状态2的个体在1期后仍在状态2的概率，可以使用马尔可夫链方法来确定未来的概率或所处的状态和长期稳定的状态（比如说最后的市场占有率）。

生成一个马尔可夫过程，需要进行以下的操作：

1. 点击 **Risk Simulator | 预测 | 马尔可夫链**
2. 输入相应状态的概率水平(例如，90%和80%) 然后点击确定。
3. 查看生成的报告

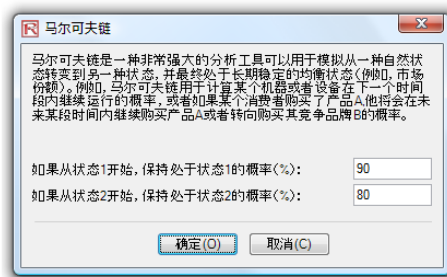


图 3.19 马尔可夫链（转换体系）

### 3.13 最大似然估计模型（MLE）：Logit, Probit, Tobit

理论:

有限因变量描述了这样的情形因变量包含的数据在范围上都是有限的，例如二元响应数据（0 或者 1），截取的数据，有序或者截断数据。例如，假设有一系列独立变量（例如，信用卡或者按揭贷款人的年龄，收入，教育水平），可以通过最大似然模型（MLE）模拟违约发生率。响应或者因变量 Y 是二元的，即，只有两种可能性发生，以 1 和 0 表示（例如，Y 表示有/没有的条件，贷款违约/没有违约，一些设备成功/失败，问卷调查是/否，等等），这样就有回归独立变量 X，假定会对 Y 产生影响。典型的普通最小二乘法对于此类问题变得失去效应，因为回归误差项具有非正态性和异方差性，估计的结果会失去意义，大于 1 或者 0。MLE 分析解决了这些问题，在因变量是有限的情况下，使用优化原则最大化对数可能性函数。

.Logit 或者逻辑回归通过将数据拟合成逻辑曲线来预测事件的发生概率。它是一个广义线性模型用于二元回归，同大多数形式的回归分析类似，借助于大量预测变量。MLE 应用二元多变量逻辑分析，该分析用于模拟因变量来决定某个可能组的期望成功概率。估计的系数都是对数比例的形式，不能直接解释为概率。一个简单的调整计算可以估计出最后的概率。

逻辑模型被定义为  $Y = \text{LN}[\text{Pi}/(1 - \text{Pi})]$  或者相应的,  $\text{Pi} = \text{EXP}(Y)/(1 + \text{EXP}(Y))$  系数  $\beta_i$  是对数比例，通过反对数可以得  $\text{EXP}(\beta_i)$  得到比率  $\text{Pi}/(1 - \text{Pi})$ 。可能性水平的变动

率为  $dP/dX = \beta_i P_i(1 - P_i)$ 。标准误差估计了预测的精度系数，t 统计量是每个预测系数与标准误差的商，是典型假设检验的参数用来估计每个估计量的显著性水平。为了预测每个特定组的成功概率（例如，基于一个吸烟者每年吸烟的数量来预测一个吸烟者是否会有胸部的并发症），可以通过使用 MLE 系数来简单计算估计的 Y 值。例如，假如模型是  $Y = 1.1 + 0.005 \times (\text{香烟})$ ，那么某人每年吸烟 100 盒就有一个估计的 Y 等于  $1.1 + 0.005 \times (100) = 1.6$ 。因此，计算可能比率的反对数  $\text{EXP}(\text{估计 } Y) / [1 + \text{EXP}(\text{估计 } Y)] = \text{EXP}(1.6) / (1 + \text{EXP}(1.6)) = 0.8320$ ，可得，这个人在他的一生中有 83.20% 的可能会感染胸部并发症。

Probit 模型（也称为 Normit 模型）作为对于二元响应模型的替代模型，应用 probit 函数估计使用最大似然估计和 probit 回归。Probit 和逻辑回归模型都是类似的预测模型，逻辑回归的参数估计比 Probit 模型高出 1.6 到 1.8 倍。对于两种模型的选择都是十分方便的，主要的不同就是逻辑分布都是具有高偏性的（肥尾效应）用来描述极值现象。例如，要模拟房屋的所有权，响应变量是二元的（购房或者不购房），影响变量  $X_i$  为收入，年龄，等等，例如， $I_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_n X_{in}$ ， $I_i$  的值越大，购房的可能性越高。对于每个家庭，关键的 I\* 门槛值存在，如果超过，就购买住房，否则将不购买住房，影响概率 (P) 假设是正态分布的，例如  $P_i = \text{CDF}$  (1) 使用标准正态累积概率密度分布函数 (CDF)。因此使用正态分布估计系数（可以使用 Excel 中 NORMSDIST 函数或者 RiskSimulator 中分布分析工具，选择正态分布设定均值为 0，标准差为 1）。最后，获得 Probit 或者概率单位估计，设定  $I_i + 5$ （这是因为只要概率  $P_i < 0.5$ ，估计都是为负，这是因为正态分布是在均值 0 左右对称分布的）。

Tobit 模型（截段模型）是计量经济学或者生物统计模型用来描述非负因变量  $Y_i$  和一个或者多个自变量  $X_i$  之间的关系。所谓截断模型，是因为因变量由于小于 0 不被观测到而被截断。Tobit 模型假设存在潜在的未观测到的因变量  $Y^*$ 。这个变量与自变量  $X_i$  线性相关，相关系数  $\beta_i$  的向量决定他们内部的关系。此外，存在正态分布误差项  $U_i$ ，对他们的关系产生随机影响。观测变量  $Y_i$  等于潜在变量，当潜在变量大于零的时候或者  $Y_i$  假设等于零的时候。即， $Y_i = Y^*$ , if:  $Y^* > 0$  以及  $Y_i = 0$ , if:  $Y^* = 0$ 。如果关系参数  $\beta_i$  通过最小二乘回归得到，结果的回归估计量就不一致，向下的偏移斜率和向上的偏移截距。只有 MLE 可以符合 Tobit 模型，存在辅助的统计量称为  $\sigma$ ，等同于标准最小二乘法的标准误差的估计，估计的系数类似于回归方法得到。

#### 操作过程描述:

- 运行 Excel 然后打开示例文件高级预测模型，转到二元最大似然预测 工作表，选择包含标题的数据列 Risk Simulator | 预测 | 最大似然估计
- 从下拉菜单学者因变量（参看图 3.20）然后点击 **确定** 运行模型和生成报告



### 三次样条多项式内插和外推法

三次样条多项式内插和外推法模型是用来“填满”缺失值的缺口以及预测时间序列的数据，模型可以用来内插时间序列数据（例如，收益曲线，利率，宏观经济变量如通胀率，商品价格或者市场回报率）也可以在已知的或给定的范围内外推数据，使之更好地用于预测。

年份	收益点	
0.0833	4.55%	这里是已知的收益率 作为三次样条多项式 内插和外推模型的已 知输入变量
0.2500	4.47%	
0.5000	4.52%	
1.0000	4.39%	
2.0000	4.13%	
3.0000	4.16%	
5.0000	4.26%	
7.0000	4.38%	
10.0000	4.56%	
20.0000	4.88%	
30.0000	4.84%	

运行三次样条预测，需要点击 Risk Simulator|预测|三次样条插值然后点击链接图标选择C15:C25作为已知的X变量(时间序列图的X轴)然后选择D15:D25作为已知的Y变量(确定已知的变量X和变量Y的数据长度相同)。输入想要预测的期数(例如，开始1，完成50，步长0.5)。点击确定然后查看生成的预测结果和预测图。

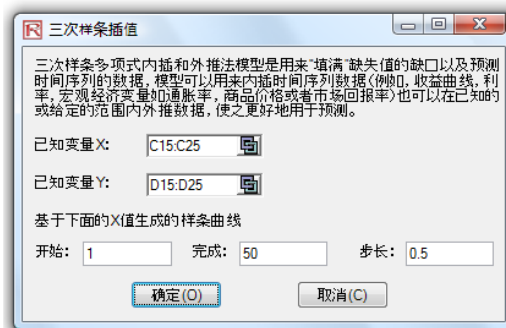


图 3.21 三次样条模块

#### 操作过程描述:

- 运行 Excel 然后打开示例文件高级预测模型，转到三次样条多项式内插和外推法工作表，选择包含标题的数据列 Risk Simulator | 预测 | 三次样条插值
- 点击链接图标链接已知的 X 值和已知的 Y 值（参看图 3.21），输入开始值和完成值。点击 **确定** 运行模型和报告（参看图 3.22）。

### 三次样条插值预测

三次样条多项式内插和外推法模型是用来“填满”缺失值的缺口以及预测时间序列的数据，模型可以用来内插时间序列数据（例如，收益曲线，利率，宏观经济变量如通胀率，商品价格或者市场回报率）也可以在已知的或给定的范围内外推数据，使之更好地用于预测。

#### 三次样条内插和外推结果

X	拟合的Y	注意
1.0	4.39%	内插
1.5	4.21%	内插
2.0	4.13%	内插
2.5	4.13%	内插
3.0	4.16%	内插
3.5	4.19%	内插
4.0	4.22%	内插
4.5	4.24%	内插
5.0	4.26%	内插
5.5	4.29%	内插
6.0	4.32%	内插
6.5	4.35%	内插
7.0	4.38%	内插
7.5	4.41%	内插
8.0	4.44%	内插
8.5	4.47%	内插
9.0	4.50%	内插
9.5	4.53%	内插
10.0	4.56%	内插
10.5	4.59%	内插
11.0	4.61%	内插
11.5	4.64%	内插
12.0	4.66%	内插



下面是三次样条内插和外推模型的已知变量：

观测值	已知的X	已知的Y
1	0.0833	4.55%
2	0.2500	4.47%
3	0.5000	4.52%
4	1.0000	4.39%
5	2.0000	4.13%
6	3.0000	4.16%
7	5.0000	4.26%
8	7.0000	4.38%
9	10.0000	4.56%
10	20.0000	4.88%
11	30.0000	4.84%

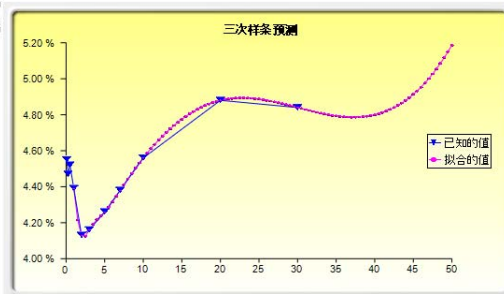


图 3.22 样条预测的结果

## 4

## 4. 优化

本章节我们会更详细地了解优化的过程和方法，它也是属于 **Risk Simulator** 的功能之一。进行优化的方法包括连续型优化和离散型优化，以及静态，动态和随机优化。

### 4.1 优化方法

运行优化时存在很多运算法则，当优化与 **Monte Carlo** 仿真一起使用时存在很多不同的步骤。在 **Risk Simulator** 中，存在三种不同的优化过程、优化类型以及不同的决策变量类型。例如，**Risk Simulator** 可以处理**连续决策变量**（1.2535，0.2215 等等），**离散决策变量**（例如，1，2，3，4 或 1.5，2.5，3.5 等等），**二元决策变量**（用于停/走决策的 1 和 0 变量），混合决策变量（整数变量和连续变量）。**Risk Simulator** 还可以处理**线性优化**（例如，当目标和约束条件都是线性公式和函数时），以及**非线性优化**（例如，当目标和约束条件是线性和非线性函数和公式的时候）。

对于整个优化过程，**Risk Simulator** 首先可用于运行**离散优化**，也就是基于离散或静态模型的优化，此时不运行仿真。换句话说，模型里的所有输入量都是静态的和确定的。当模型包含的因素都是确定的时就可以使用这种优化类型。再则，在运行更高级的优化程序之前可以首先运行一个离散优化来确定可选方案及相关的决策变量的可选分布。例如，在运行随机优化问题之前，我们可以先运行一个离散优化以确定在进行更深入的分析之前是否存在其它解决优化问题的方案。

当 **Monte Carlo** 仿真和优化一起使用时，我们就要运用**动态优化**。此过程的另一个名称叫做**仿真-优化**。也就是说，先运行一次仿真，接着在 **Excel** 模型中运用这些仿真结果，并对这些仿真结果进行优化。换句话说，在 **N** 次试验中运行仿真，然后在 **M** 次迭代中运行优化程序直到得到一个最佳结果或发现一个不可行集。利用 **Risk Simulator** 的优化模型时您可以选择要使用哪些预测和假设统计量，在运行仿真后的模型中替换。然后，这些预测统计量可以应用于优化程序中。在模型很大，里面包含很多相互关联的假定和预测，以及优化中要求某些预测统计量的情况下，这种方法很有效。例如，如果在优化模型中需要假定或预测标准差（利用均值除以资产组合的标准差来计算资产分配和优化问题的夏普指数（**Sharpe-Ratio**）），此时就应该使用这个方法。

除了整个动态优化过程要重复  $T$  次以外，**随机优化**过程与动态优化类似。也就是说，在  $N$  次试验中运行仿真，然后在  $M$  次迭代中运行优化程序直到得到一个最佳结果。然后将上述过程重复  $T$  次。最后会生成一个每个决策变量有  $T$  个值的预测表。换句话说，运行仿真以及在优化模型中使用预测或假设统计量来寻找决策变量的最佳分配。然后，再运行一次仿真，生成不同的预测值，优化这些新更新值。因此最终每个决策变量都会有一个单独的预测表，说明最佳决策变量的范围。例如，您现在可以在动态优化程序中获得决策变量的分布，进而得到每个决策变量的最佳值范围。该过程也被称为随机优化，它不仅仅是得到单点估计值。

最后，优化中还包括一种有效前沿优化过程，它在优化中应用了边际增量和影子定价理论。意思就是，如果放宽其中一个约束条件，优化的结果会发生什么变化？举个例子来说，如果假定预算约束是 100 万美元。那么当它变为 150 万，200 万等时，资产结果和最佳决策会发生什么变化？这就是金融投资领域的马可维茨前沿概念，也就是当资产的标准差被允许有轻微的增大时，资产会产生哪些额外收益？除了只允许改变一个约束条件，并且在每次变化时，仿真和优化程序仍在运行以外，该过程与动态优化过程类似。运行这个过程最好利用 Risk Simulator 手动进行。也就是先运行一个动态或随机优化，再在一个约束条件下重新运行另一次优化，最后重复几次这个过程。这个手动过程是非常重要的，因为分析者可以在改变约束条件的情况下，发现结果是类似还是不同，以此来确定是否需要进其它额外的分析，或是确定为了得到目标和决策变量的一个显著变化，约束所需的边际增加程度。

有一点值得注意。有其它一些软件在表面上可以进行随机优化，但实际上它们是不可行的。例如，它们可能是在运行一次仿真后，生成了一个优化过程的迭代，再运行另一次仿真，生成第二个优化迭代，依此类推，这样非常浪费时间和资源。在优化过程中，模型要经过一系列严格运算法则的检验，此时我们就需要运用多重迭代（范围从几次到上万次迭代）来得到最佳的结果。因此，每一次只生成一次迭代非常浪费时间和资源。与这种需要几个小时的方法相比，运用 Risk Simulator 可以在一分钟内得到同样的资产组合结果。同时这种仿真-优化方法一般会得到不好的结果，并不是一种随机的优化方法。当在我们的模型中应用优化方法时一定要非常注意这类方法。

接下来我们来看看两个优化问题的例子。一个例子使用的是连续型决策变量，另一个使用的是离散整数型决策变量。两个模型均可应用离散优化，动态优化，随机优化，甚至影子定价的最佳边界等方法。因此为了简便起见，我们仅仅介绍一下模型的设置，由使用者自己决定到底运行哪种优化程序。另外，连续型模型使用非线性优化方法（因为用于计算的资产组合风险是一个非线性函数，所以用资产回报除以资产风险表示的目标也是非线性函数），第二个例子中的整数型优化则使用线性优化模型（它的目标和所有约束条件都是线性的）。因此，这两个例子囊括了上述所有的方法。



## 4.2 连续型决策变量的优化

图 4.1 中是连续型优化模型的例子。可以依次点击**开始|程序|Real Option Valuation|Risk Simulator|示例**找到连续型优化文件。在本例中一共存在 10 种典型的资产类型（例如，不同种类的共同基金，股票或资产），目标在于如何最有效的分配这些资产以取得最佳的效果。也就是说给定每种资产的固有风险，如何得到可能的最佳资产收益。为了更好的理解优化的概念，我们必须进一步研究这个示例模型来学习如何更好的利用优化程序。

模型中显示了十种资产的分类及每种资产的年收益率和年波动率。这些收益和风险都是以年均值来表示的，以便不同的资产类型之间可以相互进行比较。收益率用相关收益的几何均值来计算，而风险则使用相关股票收益对数法来计算。更多关于股票或其它资产年波动率和年收益率计算的详情请参见本章的附录。

资产配置优化模型										
资产描述	年收益率	风险波动率	分配权重	最小分配额	最大分配额	风险收益率	收益评级 (高-低)	风险评级 (高-低)	风险收益评级 (高-低)	分配评级 (高-低)
资产1	10.54%	12.36%	10.00%	5.00%	35.00%	0.8524	9	2	7	1
资产2	11.25%	16.23%	10.00%	5.00%	35.00%	0.6929	7	8	10	1
资产3	11.84%	15.64%	10.00%	5.00%	35.00%	0.7570	6	7	9	1
资产4	10.64%	12.35%	10.00%	5.00%	35.00%	0.8615	8	1	5	1
资产5	13.25%	13.28%	10.00%	5.00%	35.00%	0.9977	5	4	2	1
资产6	14.21%	14.39%	10.00%	5.00%	35.00%	0.9875	3	6	3	1
资产7	15.53%	14.25%	10.00%	5.00%	35.00%	1.0898	1	5	1	1
资产8	14.95%	16.44%	10.00%	5.00%	35.00%	0.9094	2	9	4	1
资产9	14.16%	16.50%	10.00%	5.00%	35.00%	0.8584	4	10	6	1
资产10	10.06%	12.50%	10.00%	5.00%	35.00%	0.8045	10	3	8	1
组合总量	12.6419%	4.58%	100.00%							
风险收益率	2.7596									
优化模型的特定要求										
目标：最大风险收益率 (C18)										
决策变量：分配权重 (E6:E15)										
对决策变量的限定：最大化和最小化要求 (F6:G15)										
限制条件：总资产配置权重100% (设定单元格E17为100%)										
其他的特定要求：										
1、可以最大化组合收益或者最小化组合风险。										
2、结合模型中的Monte Carlo仿真技术，通过仿真-优化技术仿真每个资产的收益率和波动率。										
3、可以不通过Monte Carlo技术运行静态的优化技术。										

图 4.1 连续型优化模型

列的权重对决策变量有要求，需要检查变量的总份额为 100%（单元格 E17）。一般情况下在开始优化时，我们会将这些单元格设成统一的值，在本例中，从单元格 E6 到 E15 我们均设为 10%。除此以外，每个决策变量的变化范围还有具体的限制。在本例中，我们从 F 和 G 列中可以看出最低权重和最高权重分别为 5%和 35%。这意味着每类资产都有自己的权重边界。H 列代表的是风险收益率，即收益率除以风险率，这个值越大，花同样的钱取得的效果就越大。模型的余下部分显示了个别资产的收益率、风险、风险收益率和份额排名。换句话说，这些排名告诉我们哪类资产拥有低风险或高收益。

投资的资产总体收益率在单元格 C17 中用命令 Sumproduct (C6: C15, E6: E15) 来计算，那也就是份额乘以每种资产的年收益率的和。我们可以用如下等式表示：

$$R_p = \omega_A R_A + \omega_B R_B + \omega_C R_C + \omega_D R_D$$

其中  $R_p$  代表资产总的收益率， $R_{A,B,C,D}$  是项目的个体收益率， $\omega_{A,B,C,D}$  代表每个项目的权重或资金份额。

另外，单元格 D17 中资产组合的风险通过等式：

$$\sigma_p = \sqrt{\sum_{i=1}^n \omega_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{j=1}^m 2\omega_i \omega_j \rho_{i,j} \sigma_i \sigma_j}$$

来计算。这里， $\rho_{ij}$  代表资产种类之间的相关系数——因此，如果相关系数为负，那么它们就有分散风险的作用，资产的风险就降低了。但为了简化计算，我们假设在计算资产风险时，资产之间的相关性为 0，只在稍后的收益仿真中假设存在相关性。这样就不用在这些不同的资产收益之间运用静态相关性，而是只在仿真假设时应用相关性，在仿真的收益值之间创造出一种更为动态的关系。

最后还要计算资产的风险收益率或夏普指数（Sharpe-Ratio）。这个值位于 C18 单元格，代表本优化模型中的最大化目标。总结一下，本模型包含如下一些属性：

目标：最大化风险收益率（C18）

决策变量：权重（E6：E15）

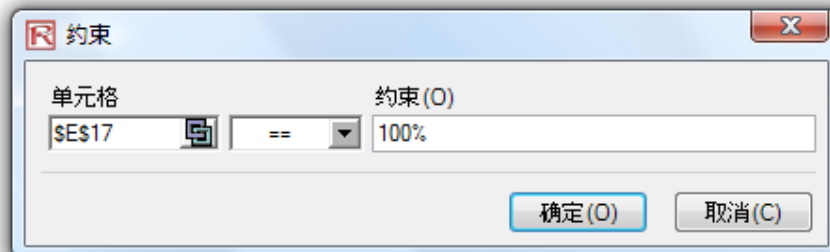
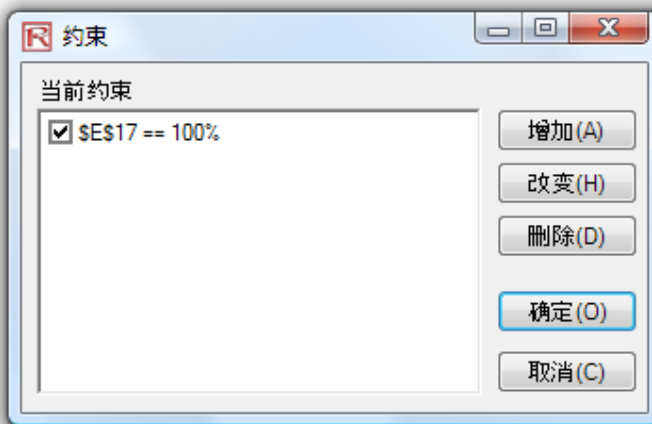
决策变量的限制：最小值和最大值（F6：G15）

约束条件：资产总权重为 100%（E17）

步骤：

- 打开案例文件通过点击仿真|新建仿真文档新建一个新文件，并命名。
- 优化的第一步就是设置决策变量。选定单元格 E6 开始设置第一个决策变量（**仿真|优化|设置决策变量**），点击连接图表选择命名单元格（B6），同时在单元格 F6 和 G6 中选择上限和下限。然后，使用 Risk Simulator 的复制功能，复制单元格 E6 的决策变量，并将其粘贴到剩下的单元格 E7 至 E15 中。
- 优化的第二步是设置约束条件。这里只有一个约束条件，就是资产的总份额之和为 100%。点击**仿真|优化|约束条件...**选择增加添加一个新约束。然后选中单元格 E17，输入=100%。完成后点击**确定**。
- 优化的最后一步是设置目标函数，选择目标单元格 C18，依次点击**仿真|优化|运行优化**，然后选择优化类型（静态优化，动态优化或者随机优化），最后就可以开始仿真了。开始可以选用静态优化。再检查确保已经在单元格 C18 中设置了目标，并选择最大化。如果需要的话还可以检查一下决策变量和约束条件，或是直接点击**确定**运行静态优化。
- 优化完成之后，您可以选择**恢复**来回到决策变量和目标的原始值，或是选择**替换**直接利用优化后的决策变量。一般情况下，我们会在优化完成后选择**替换**功能。

图 4.2 是以上这些步骤的图示。可以在模型的收益和风险（C 列和 D 列）处增加仿真假定，然后使用动态优化和随机优化。



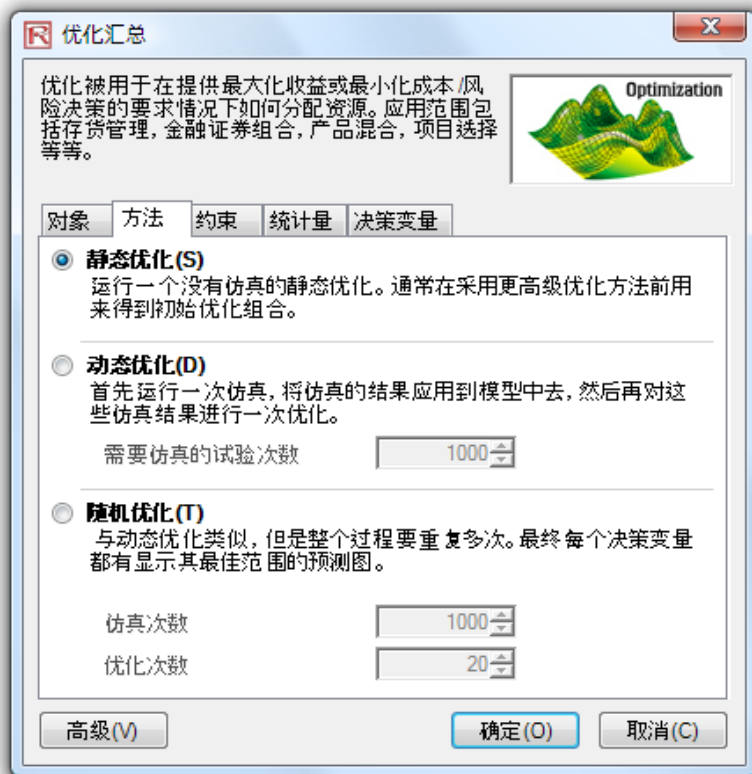


图 4.2 在 Risk Simulator 中运行连续型优化

## 结果解析:

图 4.3 中显示了优化的最终结果，资产的最佳组合参见单元格 E6 至 E15。也就是说给定每种资产份额在 5%到 35%之间波动以及份额总和为 100%这两个限制条件，图 3.4 中显示了使得风险收益率最大的资产组合。

在回顾优化过程和结果之前一些值得注意的事项：

- 运行优化的正确方式是最大化风险收益率（夏普指数（Sharpe-Ratio））或最大化花同样钱获得的收益。
- 如果我们不要求最大化整体组合的收益，那么最佳组合结果就没有太大意义，我们也就不需要优化这一步骤了。也就是说，收益最低的 8 类资产的份额为 5%（允许的最小值），最高收益资产的份额设为 35%，剩下的 25% 分配给收益次优的资产。不需要进行优化。但是，当这样投资分配时，其风险比最大化风险收益率时的风险高很多，尽管此时资产组合的收益更高些。
- 相反的，您可以使得整体资产组合的风险最小，此时对应的收益也会小些。

表格 4.1 是三种不同目标情况下运行优化的结果

目标	资产组合收益	资产组合风险	资产组合的风险收益率
最大化风险收益率	12.69%	4.52%	2.8091
最大化收益	13.97%	6.77%	2.0636
最小化风险	12.38%	4.46%	2.7754

表 4.1 优化结果

从表中我们可以看出，最佳的方法是最大化风险收益率，也就是，在同等风险下，这种

组合的收益最大。相反的，在同等收益下，这种组合的可能风险是最小的。这种花同样的钱获得收益或风险收益率方法是现代投资组合理论中马可维茨有效前沿理论的基石。意思就是如果我们限制投资组合整体的风险水平，并让其随时间逐渐增大，我们会得到几个不同风险水平的最佳投资组合。因此，对风险的不同偏好会得到不同的最佳投资组合。

**资产配置优化模型**

资产描述	年收益率	风险波动率	分配权重	最小分配额	最大分配额	风险收益率	收益评级 (高-低)	风险评级 (高-低)	风险收益评级 (高-低)	分配评级 (高-低)
资产1	10.54%	12.36%	11.09%	5.00%	35.00%	0.8524	9	2	7	4
资产2	11.25%	16.23%	6.87%	5.00%	35.00%	0.6929	7	8	10	10
资产3	11.84%	15.64%	7.78%	5.00%	35.00%	0.7570	6	7	9	9
资产4	10.64%	12.35%	11.22%	5.00%	35.00%	0.8615	8	1	5	3
资产5	13.25%	13.28%	12.08%	5.00%	35.00%	0.9977	5	4	2	2
资产6	14.21%	14.39%	11.04%	5.00%	35.00%	0.9875	3	6	3	5
资产7	15.53%	14.25%	12.30%	5.00%	35.00%	1.0898	1	5	1	1
资产8	14.95%	16.44%	8.90%	5.00%	35.00%	0.9094	2	9	4	7
资产9	14.16%	16.50%	8.37%	5.00%	35.00%	0.8584	4	10	6	8
资产10	10.06%	12.50%	10.35%	5.00%	35.00%	0.8045	10	3	8	6
<b>组合总里</b>	<b>12.6919%</b>	<b>4.52%</b>	<b>100.00%</b>							
<b>风险收益率</b>	<b>2.8091</b>									

图 4.3 连续型优化结果

### 4.3 离散整数型优化

有时候，决策变量是离散型（如 0 和 1）而不是连续的。这意味着我们可以使用像转换开关或执行/不执行来决策那样的优化。图 4.4 是一个包含 20 个项目的项目选择模型。本例使用的是**开始程序|Real Option Valuation|Risk Simulator|示例中的离散优化文件**。跟以往一样，每个项目都有自己的收益（ENPV 扩展净现值和 NPV 净现值——ENPV 就是 NPV 与任何战略实物期权值之和），运作成本和风险等。如果需要的话，还可以对模型进行修改，加入所需的全工时评量法（FTE），其它各种函数资源，以及针对这些附加资源的附加约束。本模型中的输入量基本都是从其它表中的模型链接而来的。例如，每个项目的投资模型都会有自己的折现现金流或收益大小。此处是应用在一定的预算下最大化投资组合的夏普指数 (Sharpe-Ratio)，或是最小化风险，或是选择的项目总数不超过 10 个的情况下增加额外的约束等等。所有这些情况都可以利用这个现有的模型。

步骤:

- 打开案例文件，通过点击**仿真|新建仿真**生成一个新文档，并命名。
- 优化的第一步就是设置决策变量。选定单元格 J4 开始设置第一个决策变量（**仿真|优化|设置决策变量**），点击连接图标选择命名单元格（B4），选择**二元变量**。然后，使用 Risk Simulator 的复制功能，复制单元格 J4 的决策变量，并将其粘贴到剩下的单元格 J5 至 J23 中。在只存在几个决策变量的情况下，这是最佳的方法，可以为每个变量提供一个特殊的名字以方便以后的查找。
- 优化的第二步是设置约束条件。这里有两个约束条件：资产组合的总预算要小于\$5000，项目的总数不能超过 6 个。点击**仿真|优化|约束条件...**选择**增加**，添加一个新约束。然后选中单元格 D17，输入小于等于 (<=) 5000。

通过  $J17 \leq 6$  来重复设置。

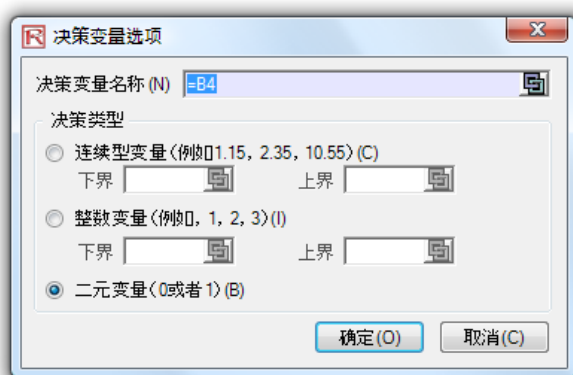
- 优化的最后一步是设置目标函数，通过选择目标单元格 C19（或 C17），**仿真|优化|运行优化**，然后选择优化类型（静态优化，动态优化或者随机优化），最后就可以开始仿真了。开始可以选用静态优化。再检查确保目标单元格中是夏普指数(Sharpe-Ratio)或风险收益率，然后选择**最大化**。如果需要的话还可以回顾一下决策变量和约束条件，或是直接点击**确定**运行静态优化。

图 4.5 是以上这些步骤的图示。您可以在模型的 ENPV 和风险（C 列和 F 列）处增加仿真假设，然后练习使用动态优化和随机优化。

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
3		备选项目	扩展净现值	成本	风险	风险	回报/风险比	盈利指数			选择	
4		项目1	\$458.00	\$1,732.44	\$54.96	12.00%	8.33	1.26			1.00	
5		项目2	\$1,954.00	\$859.00	\$1,914.92	98.00%	1.02	3.27			1.00	
6		项目3	\$1,599.00	\$1,845.00	\$1,551.03	97.00%	1.03	1.87			1.00	
7		项目4	\$2,251.00	\$1,645.00	\$1,012.95	45.00%	2.22	2.37			1.00	
8		项目5	\$849.00	\$458.00	\$925.41	109.00%	0.92	2.85			1.00	
9		项目6	\$758.00	\$52.00	\$560.92	74.00%	1.35	15.58			1.00	
10		项目7	\$2,845.00	\$758.00	\$5,633.10	198.00%	0.51	4.75			1.00	
11		项目8	\$1,235.00	\$115.00	\$926.25	75.00%	1.33	11.74			1.00	
12		项目9	\$1,945.00	\$125.00	\$2,100.60	108.00%	0.93	16.56			1.00	
13		项目10	\$2,250.00	\$458.00	\$1,912.50	85.00%	1.18	5.91			1.00	
14		项目11	\$549.00	\$45.00	\$263.52	48.00%	2.08	13.20			1.00	
15		项目12	\$525.00	\$105.00	\$309.75	59.00%	1.69	6.00			1.00	
16												
17		加总:	\$17,218.00	\$8,197.44	\$7,007.00	40.70%					12.00	
18		目标:	最大化	<=\$5000							<=6	
19		Sharpe比率:	2.46									
20												
21												

扩展净现值是备选项目的预期净现值，然而可以使管理和需要持有来管理总备选项目的总成本，风险是备选项目扩展净现值的变异系数。

图 4.4 离散优化模型



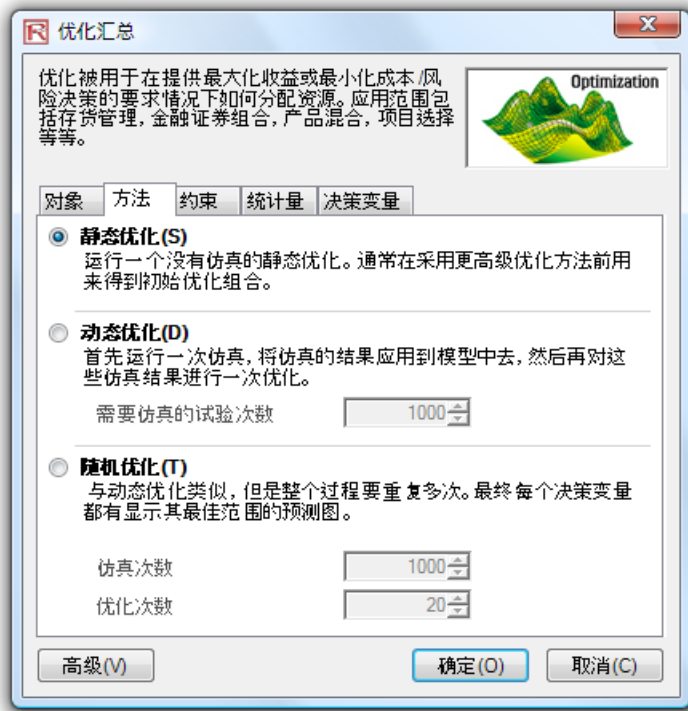


图 4.5 在 Risk Simulator 中运行离散型优化

结果解析:

图 4.6 中显示了通过最大化夏普指数 (Sharpe-Ratio) 的方法来选择最佳项目的一个示例。您也可以选择最大化总体收益, 并简单地从清单中不断选择最高收益的项目直到资金用完或是超出预算约束, 这将是一个非常繁琐的过程。但是这样可能会选到一些理论上不欢迎的项目, 因为收益越高, 风险越大。使用 Risk Simulator 将会使这一切变得非常方便。现在如果需要的话, 您可以通过在输入假设中加入 ENPV 和风险值, 然后利用随机或动态优化来重复优化过程。

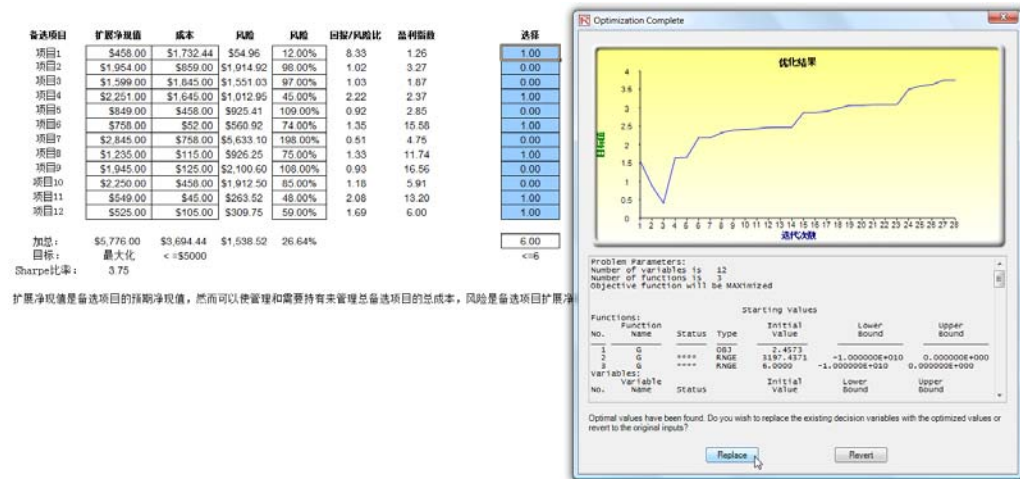


图 4.6 最大化夏普指数 (Sharpe-Ratio) 条件下的最佳项目选择

如果想要了解更多关于优化的例子，可以参考《实物期权分析：工具和方法》，第二版一书的第十一章综合风险分析，(Wiley Finance, 2005)。这个案例学习解释了如何得到一个有效前沿，以及如何将预测，仿真，优化和实物期权融入到分析过程之中。

#### 4.4 有效边际和高级优化选项

图 4.7 的第二个图表显示了优化的限制条件。这里，如果在设定某些限制条件之后，点击了**有效边际**按钮，现在可以改变这些限制因素。即，每个限制因素可以在最大值和最小值之间。例如，限制因素单元格  $J17 \leq 6$  可以设定在 4 和 8 之间运行优化（图 4.7）。即将运行五组优化，每组都有以下的限制条件： $J17 \leq 4$ ,  $J17 \leq 5$ ,  $J17 \leq 6$ ,  $J17 \leq 7$  和  $J17 \leq 8$ 。最优的结果将绘制成有效边际然后创建报告（图 4.8）。需要指出的是，下面显示了创建变动限制的步骤：

- 在一个优化模型中（例如，包含目标值，决策变量，和设定好的限制条件）点击 **Risk Simulator | 优化 | 限制** 然后点击 **有效边际**。
- 选择想要变动的限制条件（例如， $J17$ ）然后输入参数作为最大值，最小值和优化步数（图 4.7）然后点击 **添加** 然后 **确定**。
- 运行优化（Risk Simulator | 优化 | 运行优化）。可以选择静态，动态或者随机。
- 结果将在用户手册中显示（图 4.8）。点击 **创建报告** 生成报告工作簿包含所有的详细信息和优化结果。



图 4.7



# Risk Simulator

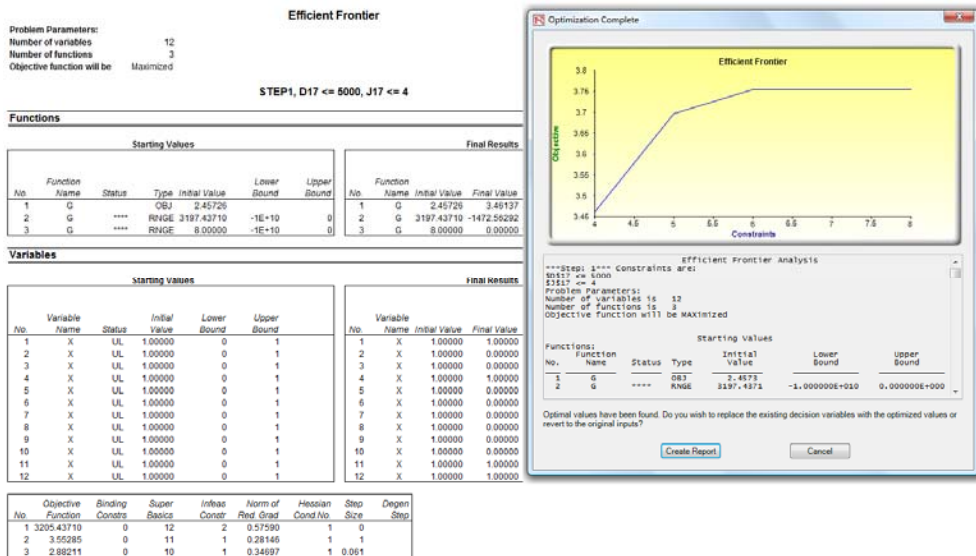


图 4.8



## 5. 风险仿真分析工具

本章涉及的是 Risk Simulator 的分析工具。我们会通过逐步分析的方法来讨论 Risk Simulator 软件中的一些案例应用，以确定分析工具的作用。这些工具对于风险分析领域的分析者来说是非常有价值的。本章将会详细介绍每种工具的应用。

### 5.1 仿真中的飓风和敏感性分析工具

理论：

仿真中的重要工具之一就是飓风图分析——它描述的是每个变量对模型结果的静态作用。这个工具会自动扰动模型中每个变量的预设值，记录模型预测或最终结果的波动，并按照扰动重要性的高低对结果进行排序。图 5.1 至图 5.6 是飓风分析应用的图示。例如，图 5.1 是一个折现现金流模型，它的输入假设如图所示。问题是哪一种关键因素对模型输出的影响最大？也就是，什么控制着\$93.63 的净现值或哪种输入变量对这个值的影响最大？

通过依次点击**仿真|工具|飓风分析**就可以使用飓风图工具了。继续第一个例子，打开案例文件夹中的飓风和敏感性图（线性）文件。如图 5.2 中所示的模型，单元格 G6 中的净现值是需要被分析的目标结果。模型中目标单元格的引用部分被用于生成飓风图。引用变量是所有影响模型结果的输入量和中间变量。例如，如果模型中包含  $A=B+C$ ，并且  $C=D+E$ ，那么 B, D, E 就是 A 的引用变量（C 是中间计算值，不是引用变量）。图 5.2 中还显示了用于估计目标结果的每个引用变量的检验范围。如果引用变量仅仅是输入量，那么检验范围将会是基于所选择范围（例如，误差范围在  $\pm 10\%$  之间）的一个简单扰动。每个引用变量都可以根据需要在不同的百分比范围内扰动。大范围比较重要，因为相对于围绕期望值的小幅度扰动来说，它可以涉及极值的检验。在某些情况下，极值可能会造成比较大，比较小，或是不稳定的影响（例如，对于一个较大或较小的变量值，当它们的规模经济效应和范围经济效应增加或减少时就容易发生非线性事件），只有在大范围情况下才能反映出这种非线性影响。

Discounted Cash Flow 模型

基年	2005	总现值收益	\$1,896.63
贴现率	15.00%	总现值投入	\$1,800.00
风险中性概率	5.00%	净现值	\$96.63
销售增长额	2.00%	内部收益率	18.80%
价格侵蚀	5.00%	投资回报	5.37%
税率	40.00%		

	2005	2006	2007	2008	2009
产品A的价格	\$10.00	\$9.50	\$9.03	\$8.57	\$8.15
产品B的价格	\$12.25	\$11.64	\$11.06	\$10.50	\$9.98
产品C的价格	\$15.15	\$14.39	\$13.67	\$12.99	\$12.34
产品D的价格	50.00	51.00	52.02	53.06	54.12
产品E的价格	35.00	35.70	36.41	37.14	37.89
产品F的价格	20.00	20.40	20.81	21.22	21.65
<b>总利润</b>	<b>\$1,231.75</b>	<b>\$1,193.57</b>	<b>\$1,156.57</b>	<b>\$1,120.71</b>	<b>\$1,085.97</b>
已售商品成本	\$184.76	\$179.03	\$173.48	\$168.11	\$162.90
<b>总利润</b>	<b>\$1,046.99</b>	<b>\$1,014.53</b>	<b>\$983.08</b>	<b>\$952.60</b>	<b>\$923.07</b>
运营成本	\$157.50	\$160.65	\$163.86	\$167.14	\$170.48
管理及办公室费用	\$15.75	\$16.07	\$16.39	\$16.71	\$17.05
<b>营业收入 (EBITDA)</b>	<b>\$873.74</b>	<b>\$837.82</b>	<b>\$802.83</b>	<b>\$768.75</b>	<b>\$735.54</b>
折旧	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00
摊销	\$3.00	\$3.00	\$3.00	\$3.00	\$3.00
<b>EBIT</b>	<b>\$860.74</b>	<b>\$824.82</b>	<b>\$789.83</b>	<b>\$755.75</b>	<b>\$722.54</b>
利息费用	\$2.00	\$2.00	\$2.00	\$2.00	\$2.00
<b>EBT</b>	<b>\$858.74</b>	<b>\$822.82</b>	<b>\$787.83</b>	<b>\$753.75</b>	<b>\$720.54</b>
税金	\$343.50	\$329.13	\$315.13	\$301.50	\$288.22
<b>净收入</b>	<b>\$515.24</b>	<b>\$493.69</b>	<b>\$472.70</b>	<b>\$452.25</b>	<b>\$432.33</b>
成本减值	\$13.00	\$13.00	\$13.00	\$13.00	\$13.00
净营运资本的改变量	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00
资本支出	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00
<b>自由流动资金</b>	<b>\$528.24</b>	<b>\$506.69</b>	<b>\$485.70</b>	<b>\$465.25</b>	<b>\$445.33</b>

投资	\$1,800.00				
----	------------	--	--	--	--

<b>财务分析</b>					
自由现金流的现值	\$528.24	\$440.60	\$367.26	\$305.91	\$254.62
投资应付的现值	\$1,800.00	\$0.00	\$0.00	\$0.00	\$0.00
<b>净现金流</b>	<b>(\$1,271.76)</b>	\$506.69	\$485.70	\$465.25	\$445.33

表 5.1: 样本模型

步骤:

- 在 Excel 模型中选择一个单一的输出单元 (例如, 一个有函数或方程的单元格) (本例中的单元格 G6)
- 选择仿真|工具|飓风图分析
- 检查一下引用变量然后重新加以适当命名 (使用更加简短的名称以便得到飓风图和蛛网图好的视觉效果), 点击**确定**

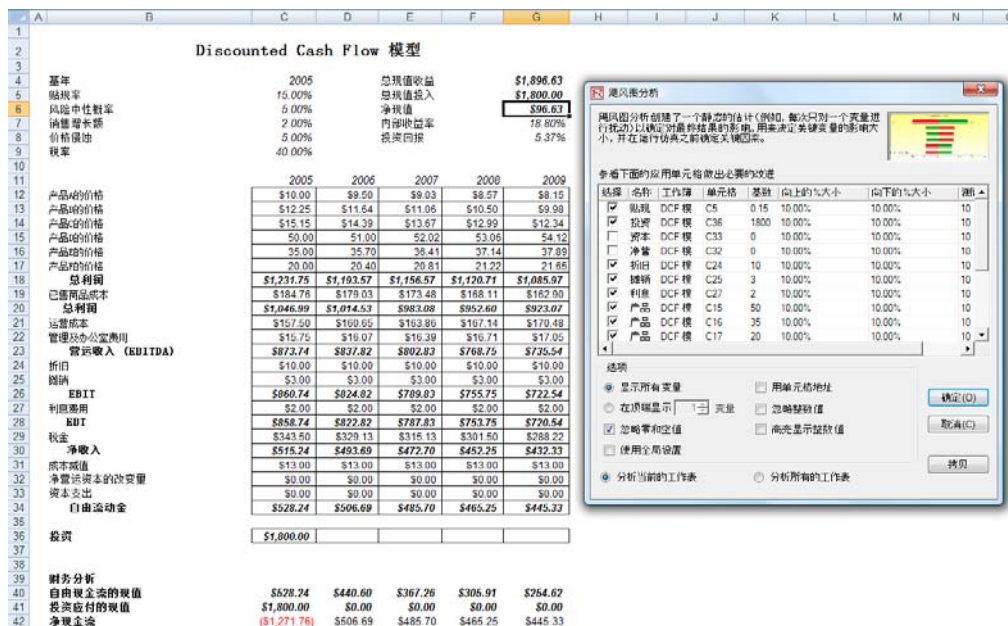


图 5.2——运行飓风分析

结果解析:

图 5.3 是飓风分析的结果报告，图中显示资本投资对净现值的影响最大，其次是税率，生产线所需的平均销售价格和数量等等。整个报告包含四个基本因素：

- 统计概述列出了运行的步骤。
- 敏感性表（图 5.4）中显示开始的 NPV 基准值为 96.63，以及每个输入量的变化（例如，投资以 10%的升幅从 1800 美元增加到 1980，以 10%的降幅从 1800 降到 1620）。结果最低值和最高值分别为-83.37 和 276.63，变化总量为 360，是对 NPV 影响最大的变量。引用变量是按照影响力大小来排列的。
- 蛛网图（图 5.5）用图来解释了这些影响。Y 轴代表净现值的目标值，X 轴代表每个引用变量变动的百分比（中心点是位于 96.63 的基准值，对于每个引用变量的偏离为 0 个百分点）。一条正斜率的直线意味着正相关的关系，负斜率的直线意味着负相关的关系（例如，投资图是负斜率的意味着投资水平越高，净现值越低）。斜率的绝对值代表影响的力度（一条比较陡的直线意味着对给定 X 轴上引用变量一定百分比的变化，Y 轴上净现值的变化较大）。
- 飓风图从另外一种图形角度进行了解释，影响最大的引用变量被置于顶部。X 轴代表净现值，中心值为图形的基准情况。图中的绿色条代表正的影响，红色条代表负的影响。因此对于投资来说右边的红色条意味着投资对高净现值的负作用——换句话说，资本投资和净现值是负相关的。反之产品 A 到 C 的价格和数量的作用亦然（图右边的绿色条）。

飓风图和蛛网图

统计汇总

你真的很有效工具之一是飓风图——它分析了每个变量对模型结果的静态影响。既然是，这个工具会让模型中的每个引用变量按照用户自定义的方式进行扰动，进而记录模型预测结果或最终结果的波动。最后按照波动性大小对扰动结果进行排序。引用变量是有可能对模型结果造成影响的输入变量和中间变量。例如，如果模型中包含A=5\*C，并且C=0.2，那么5，D，E就是A的引用变量（C是中间变量，不是引用变量）。变量的范围和数量是用户自定义的，可以用于模拟的检验，而不仅仅是高波动提供的小范围的扰动。在某些情况下，模拟可能会造成比较大，比较小，或是不稳定的影响（例如，对于一个较大或较小的变量值，当它们的波动幅度和范围波动增加或减少时能容易引发非线性事件），只有在大幅度情况下才能反映出这种非线性影响。

飓风图中列出了所有影响模型的输入变量，从对结果影响最大的输入变量开始排列。通过将每个引用变量每次在一定范围内（例如，基准的±10%之间）扰动，并将得到的结果与基准进行比较就可以得到飓风图了。除网图的状态图外还提供一只蜘蛛的中间部分和它产生的量，正负部量代表着正负相关性，负部量代表着负相关性。同时，除网图还可以帮助我们显示波动性和非线性关系。飓风图和除网图有助于找出输出单元的关键影响因素以便应用于你其他的输入变量。识别出的那些关键的不确定的变量是高风险而或被你其他的变量，不要浪费时间在你其他那些确定的或是对结果影响很小的变量。

结果

引用变量单元	基准值: 96.6261638665219			输入量变化		
	输出下限	输出上限	有效范围	输入下限	输入上限	基准值
投资	\$216.63	(\$83.37)	360.00	\$1,620.00	\$1,980.00	\$1,800.00
税率	\$219.73	(\$26.47)	246.20	36.00%	44.00%	40.00%
A价格	\$3.43	\$189.83	186.40	\$3.00	\$11.00	\$10.00
B价格	\$16.71	\$176.55	159.84	\$11.03	\$13.46	\$12.25
A数量	\$23.18	\$170.07	146.90	45.00	55.00	50.00
B数量	\$30.53	\$162.72	132.19	-31.50	36.50	35.00
C价格	\$40.15	\$153.11	112.96	\$13.64	\$16.61	\$15.15
C数量	\$48.05	\$145.20	97.16	18.00	22.00	20.00
价格侵蚀	\$138.24	\$57.00	81.24	13.50%	16.50%	15.00%
销售增长	\$116.80	\$76.64	40.16	4.50%	5.50%	5.00%
折旧	\$30.53	\$102.63	12.10	1.80%	2.20%	2.00%
利息	\$35.08	\$38.17	3.08	\$3.00	\$11.00	\$10.00
摊销	\$37.09	\$36.16	0.93	\$1.80	\$2.20	\$2.00
报摊	\$36.16	\$37.09	0.93	\$2.70	\$3.30	\$3.00
资本支出	\$36.63	\$36.63	0.00	\$0.00	\$0.00	\$0.00
运营成本	\$36.63	\$36.63	0.00	\$0.00	\$0.00	\$0.00

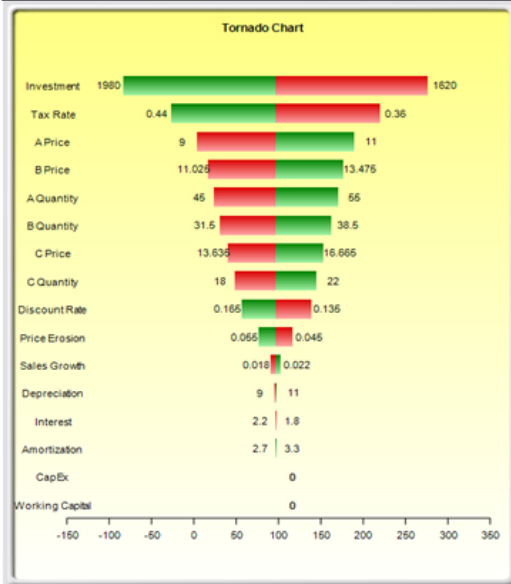
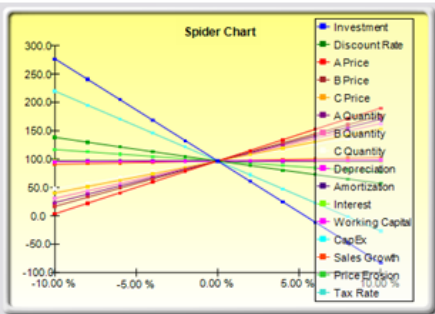


图 5.3——飓风图分析报告

注意：

记住飓风分析是一种静态的敏感性分析，应用于模型中的每一个输入变量——也就是说，每个变量都被单个扰动，其影响结果被制成表格。在运行仿真之前运行飓风分析是一个关键的步骤。风险分析中非常重要的第一步就是确定哪些是模型中对结果最重要的影响因素。下一步就是找出这些关键影响因素中哪些是不确定的。这些不确定的影响因素是项目成功的关键因素，模型的结果就取决于这些因素。这些变量是需要被仿真的。不要浪费时间去仿真那些确定的或是对结果的影响很小的变量。飓风分析可以很快很容易的帮助我们找出那些关键影响因素。在假设投资和有效税率已知并保持不变的前提下，本例中需要仿真的变量应该是价格和数量。

引用变量单元格	基准值: 96.6261638553219			输入量变化		
	输出下限	输出上限	有效范围	输入下限	输入上限	基准值
投资	\$276.63	(\$83.37)	360.00	\$1,620.00	\$1,980.00	\$1,800.00
水里	\$219.73	(\$26.47)	246.20	36.00%	44.00%	40.00%
A 价格	\$3.43	\$189.83	186.40	\$9.00	\$11.00	\$10.00
B 价格	\$16.71	\$176.55	159.84	\$11.03	\$13.48	\$12.25
A 数量	\$23.18	\$170.07	146.90	45.00	55.00	50.00
B 数量	\$30.53	\$162.72	132.19	31.50	38.50	35.00
C 价格	\$40.15	\$153.11	112.96	\$13.64	\$16.67	\$15.15
C 数量	\$48.05	\$145.20	97.16	18.00	22.00	20.00
贴现率	\$138.24	\$57.03	81.21	13.50%	16.50%	15.00%
价格侵蚀	\$116.80	\$76.64	40.16	4.50%	5.50%	5.00%
销售额增长	\$90.59	\$102.69	12.10	1.80%	2.20%	2.00%
折旧	\$95.08	\$98.17	3.08	\$9.00	\$11.00	\$10.00
利息	\$97.09	\$96.16	0.93	\$1.80	\$2.20	\$2.00
摊销	\$96.16	\$97.09	0.93	\$2.70	\$3.30	\$3.00
资本支出	\$96.63	\$96.63	0.00	\$0.00	\$0.00	\$0.00
运营成本	\$96.63	\$96.63	0.00	\$0.00	\$0.00	\$0.00

图 5.4——敏感性表格

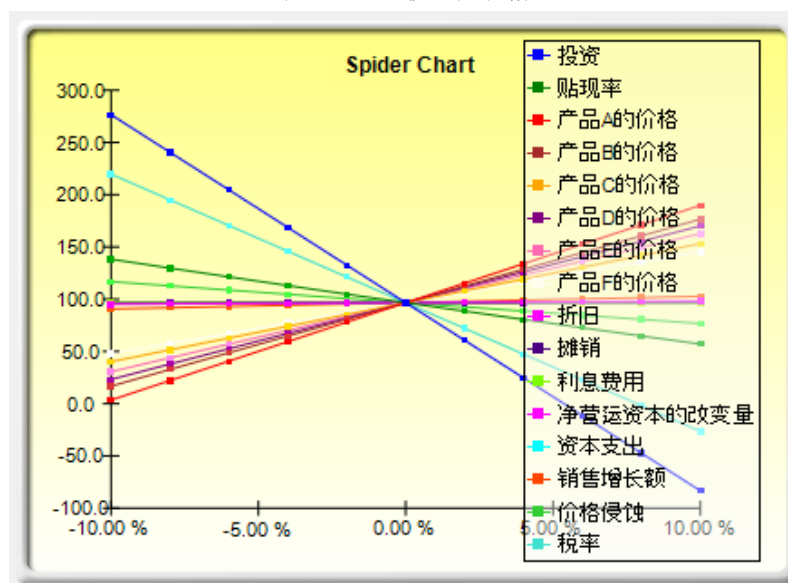


图 5.5——蛛网图

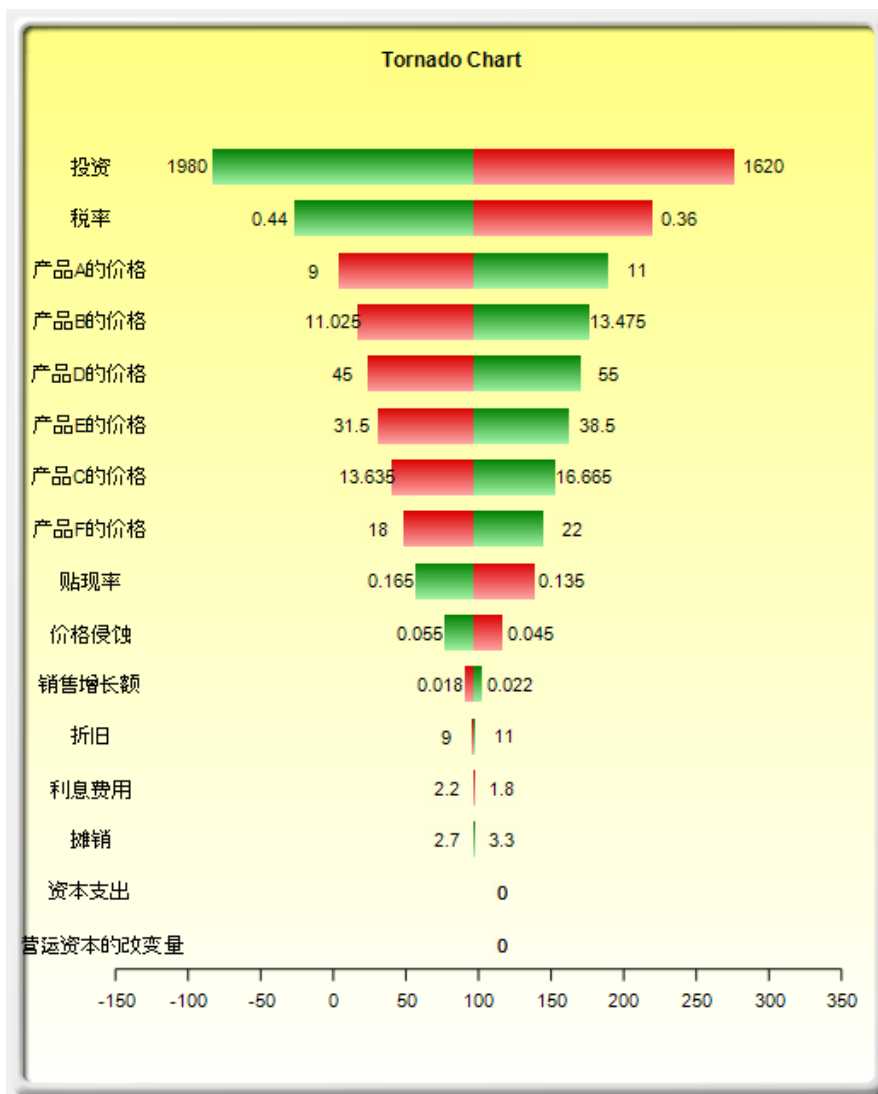


图 5.6——飓风图

尽管飓风图很容易理解,但是蛛网图有一个重要的优势就是确定模型是否存在非线性关系。例如,图 5.7 是蛛网图的另一个例子,其中有很明显的非线性关系(表中的线条不是直线而是曲线)。使用的示例文件是*飓风和敏感性表(非线性)*,它使用了 Black-Scholes 期权定价模型作为示范。飓风图不能确定这些非线性关系,虽然它们可能是模型中的重要信息或者可以让决策者更深入地观察模型的动态趋势的因素。

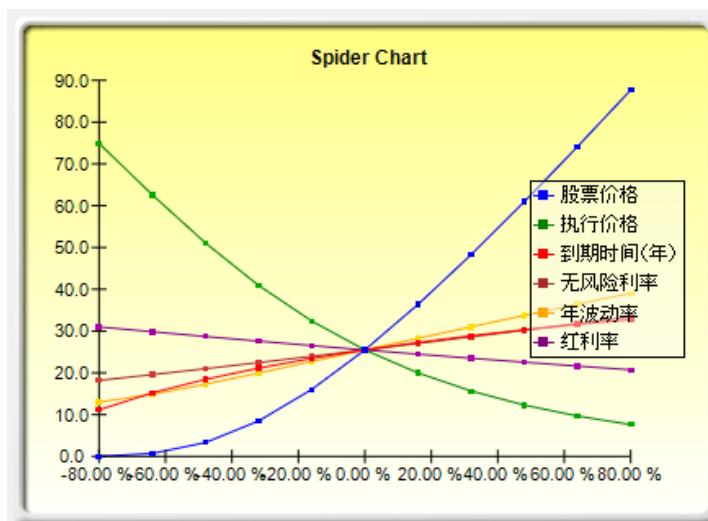


图 5.7——非线性蛛网图

更多关于飓风图的注释:

图 5.2 显示了飓风图分析工具的用户界面。注意在 Risk Simulator v4 或者更高的版本开始会有很多的增强功能。下面是关于运行飓风图分析和有关增强功能的贴士。

- 飓风图分析不应该仅仅运行一次。这就意味着，作为模型分析工具理想状态下应该运行多次。例如，在一个很大的模型中，飓风图分析可以使用默认选项运行第一次，然后显示所有的引用单元（选择**显示所有的变量**）。这样就会出现一个很大的报告和很长的飓风图（不是很利于分析和显示）。尽管如此，它却为选择多少个引用单元提作为关键因素供了一个参考（例如，飓风图可以显示 5 个具有重大影响的变量，而剩下的 200 个变量的影响则可能很小），这样飓风图运行第二次显示更少的变量（例如，选择**显示前 10 个变量**如果前 5 个是关键的，这样就创建了一个更为美观的报告和飓风图，显示了关键因素和次关键因素之间的比较，即显示不包含次关键因素的飓风图）。最后，默认设置点可以增加 $\pm 10\%$ 到某个大一些的值测试用于检验非线性（蜘蛛图可以显示非线性的线，而如果引用变量的影响是非线性的，飓风图将偏向一边）。
- **使用单元格地址**对于一个较大的模型是一个很好的办法，可以让你确定引用变量的位置（工作簿名称和单元格地址）。如果未选该选项，软件应用自身的模糊逻辑找到每个引用变量的名称（有时候名称可能会发生冲突或者过长，影响飓风图的美观）。
- **分析当前工作簿和分析所有的工作簿**选项允许客户控制引用变量是否是现有工作簿的一部分或者所有的工作表包含在同一个工作簿中。这个选项可以用在只分析当前表格中的输入输出变量还是进行全员搜索所有的引用变量。
- **使用全员设置**当分析一个较大的模型的时候十分有用，可以检验所有的引用变量，比如 $\pm 50\%$ 而非默认的 10%。不改变每个引用变量的大小，用户可以选择这个选项，改变一个设置然后**点击别的位置**，然后整个引用变量的列表都会变化。未选中这个选项允许用户每次更改一个引用变量的测试点。
- **忽略零或者空值**是一个默认选项如果引用变量包含零值或者空值则不会



在飓风图分析中考虑。这是一个典型设置。

- **着重指出可能的整数值**选项可以快速确定所有包含整数的可能引用变量单元格。这个选型有时候十分重要（例如，函数 IF 在单元格显示中可能会出现 1，或者类似的整数 1, 2, 3，这些值可能不需要进行检验）。例如， $\pm 10\%$ 对于标记 1 的结果可能会是 0.9 或者 1.1，实际上这对于模型来说是不正确的，Excel 会显示公式是错误的。这个选项一旦选择，将着重指出飓风图分析可能包含潜在问题的区域，可以手动选择或者不选这个引用变量，或者可以使用**忽略可能的整数值**同时关闭他们。

## 5.2 敏感性分析

理论：

另一个相关的特征就是敏感性分析。飓风分析（飓风图和蛛网图）是用在运行仿真之前的静态扰动，而敏感性分析则是用在运行仿真之后运行的动态扰动。飓风图和蛛网图是静态扰动后的结果，每个引用变量或假设变量每次按照一个事先设置的值进行扰动，生成的扰动结果被制成表格。相反的，敏感性分析是动态扰动的结果，多个变量假设同时被扰动，它们在模型中的相互作用和变量之间的相关性都在结果的波动中反映出来。飓风图用于识别对结果影响最大，最适合仿真的因素；敏感性分析则是确定多个变量在模型中同时被仿真时对结果的影响。图 5.8 详细的解释了这一效果。注意到关键影响因素的排名和上例中飓风图里的很类似。但是，如果再加上变量之间的相关性，图 5.9 中出现完全不同的情况。比如我们注意到价格侵蚀对净现值的影响很小，但是如果某些输入量之间相关的话，它们之间的相互作用可能会增大价格侵蚀的影响力。

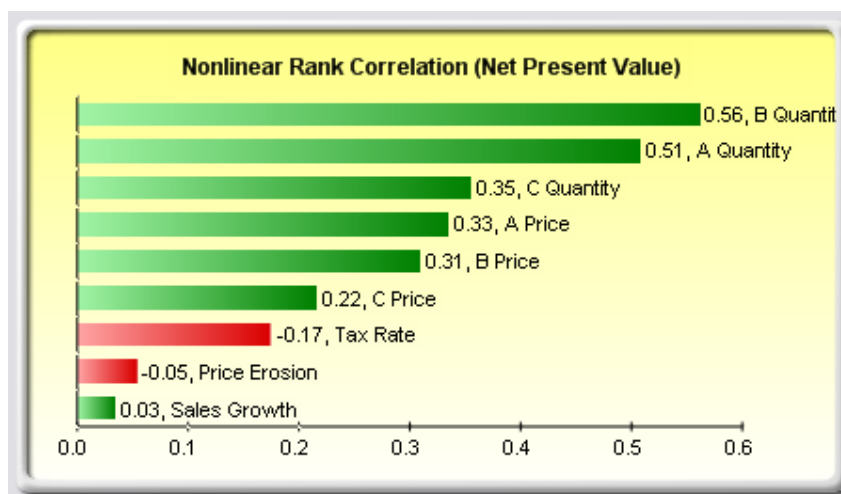


图 5.8——不存在相关性的敏感性图

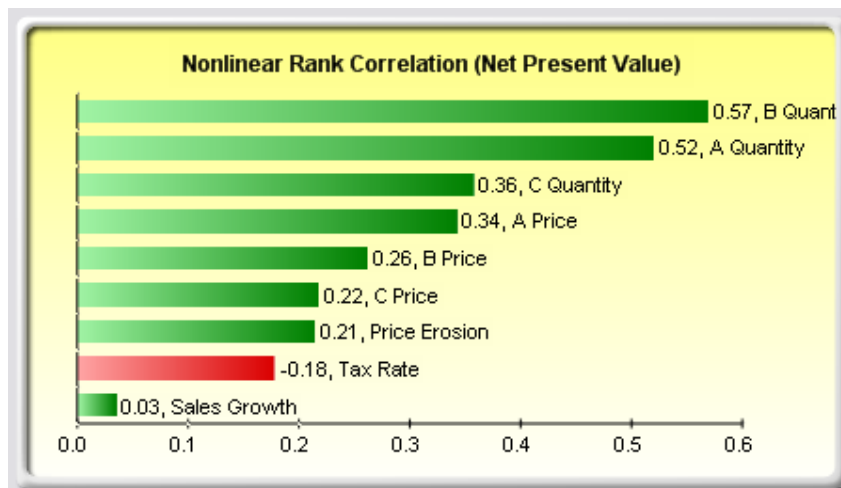


图 5.9——存在相关性的敏感性图

步骤:

- 打开或新建一个模型，定义输入和预测，运行仿真（本例使用的是**飓风**和**敏感性图（线性）**示例文件）
- 选择仿真|工具|敏感性分析
- 选择用于分析的预测点击**确定**（图 5.10）

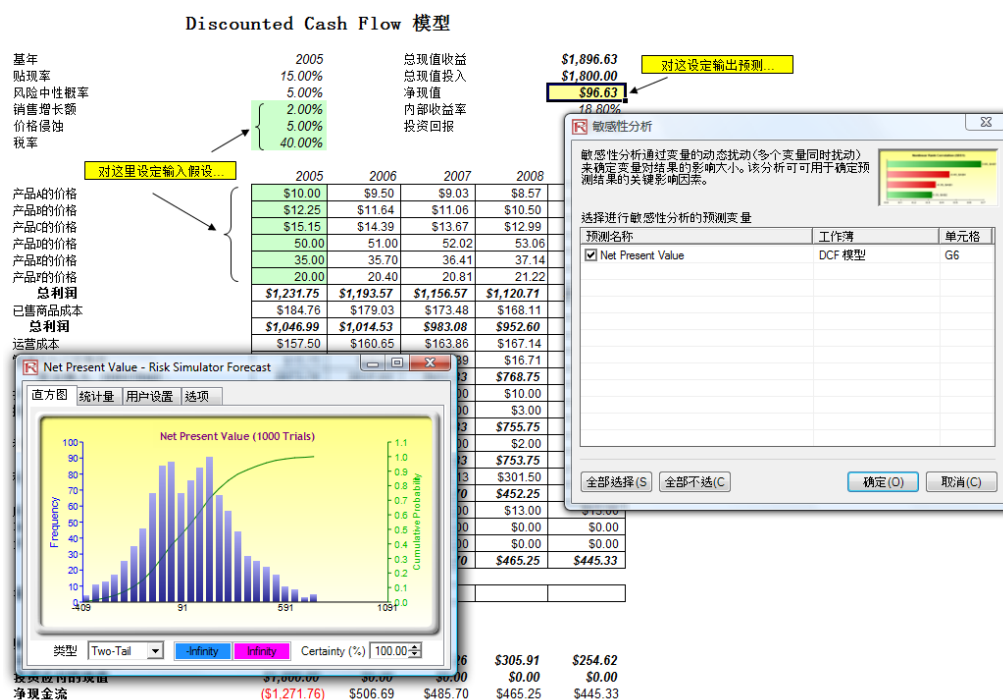


图 5.10——运行敏感性分析

结果解析:

敏感性分析的结果包括一份报告和两个关键图。首先是非线性秩相关（图 5.11），按照假设量-预测相关性由高到低来排列。这些相关关系是非线性和非参数的，这

样它们就不用遵守任何分布要求（例如，一个服从韦伯分布的假设量可以和另一个服从 $\beta$ 分布的变量相比较）。除了一点以外，从图中得出的结果与之前飓风分析的结果很类似（当然不包括资本投资这个变量，因为我们假定它是已知的，所以不需要仿真）。另外，与飓风图（图 5.6）相比，税率在敏感性分析图（图 5.11）中的位置相对低一些。这是因为如果仅是看税率，它对结果影响很大，但是一旦模型中的其它变量有相互作用，税率的影响就明显降低（这是因为由于历史税率数据的波动很小，导致其分布也较小，同时税率是税前收入的一定百分比值，其它引用变量对税前收入有很大影响）。这个例子证明敏感性分析在运行仿真之后来确定模型中是否存在相互作用以及这些作用是否会持续具有重要作用。第二个图（图 5.12）说明了变异的百分比解释。也就是说，给出预测的波动，并考虑到变量之间的相互作用之后，每个假设变量对变异的解释程度如何？注意通常所有变异的解释总和都会接近 100%（有时会有其它因素对模型产生影响，但是不能被直接观察到），如果存在相关性，那么总和有时可能会超过 100%（这取决于累积的相互作用）。

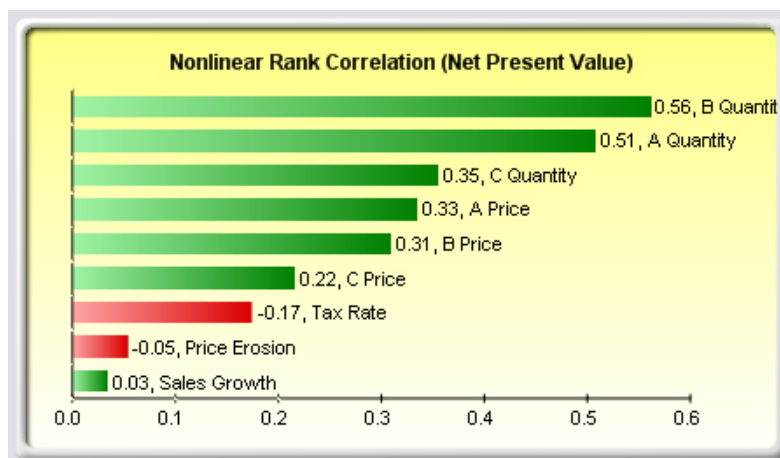


图 5.11——相关性排列图

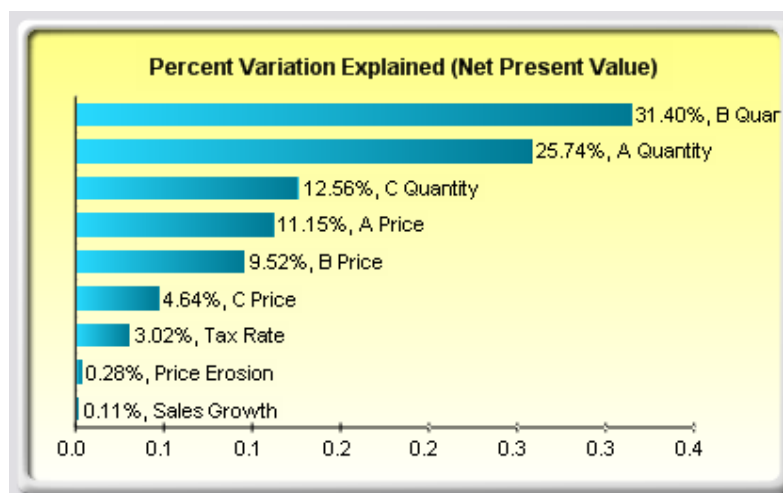


图 5.12——方差贡献图

**注意：** 飓风分析是在运行仿真之前，而敏感性分析是在运行仿真之后。飓风分析中的蛛网图可以处理非线性情况，敏感性分析中的相关性排列图可以处理非线性和自由分布情况。

### 5.3 分布拟合：单变量和多元变量

**理论：** 另一个有效的仿真工具是分布拟合。对于模型中的一个具体输入变量分析者该使用哪种分布呢？分布的相关参数有哪些？如果变量不存在历史数据，那么分析者必须对未知变量作一些相关假定。其中一种方法就是利用 Delphi 法，也就是一组专家来估计每个变量的变化。例如，一组机械工程师要通过严格的试验或推测来估计螺旋弹簧在极端情况下的直径。这些值可被看作变量的输入参数（例如，极值为 0.5 和 1.2 的均匀分布）。当试验不能进行时（如市场份额和收益增长率），管理层依然可以对可能的结果作一些估计以提供最佳案例情景，最可能案例情景和最差案例情景。

然而，如果可以得到历史数据，那我们就可以进行分布拟合。假设这种历史趋势自身是不断重复的，那么就可以利用历史数据来找到最佳拟合分布及其相关参数，以便更好的定义用于仿真的变量。图 5.13 到 5.15 是一个分布拟合的例子。使用的是示例文件夹中的数据拟合模型。

**步骤：**

- 打开一个包含数据的工作簿
- 选择想要拟合的数据（数据必须在同一列）
- 选择仿真|工具|分布拟合（单变量）
- 选择希望拟合的分布类型或是接受默认值，选择所有的分布，点击**确定**（图 5.13）
- 查看拟合的结果，选择想要的相关分布点击**确定**（图 5.14）

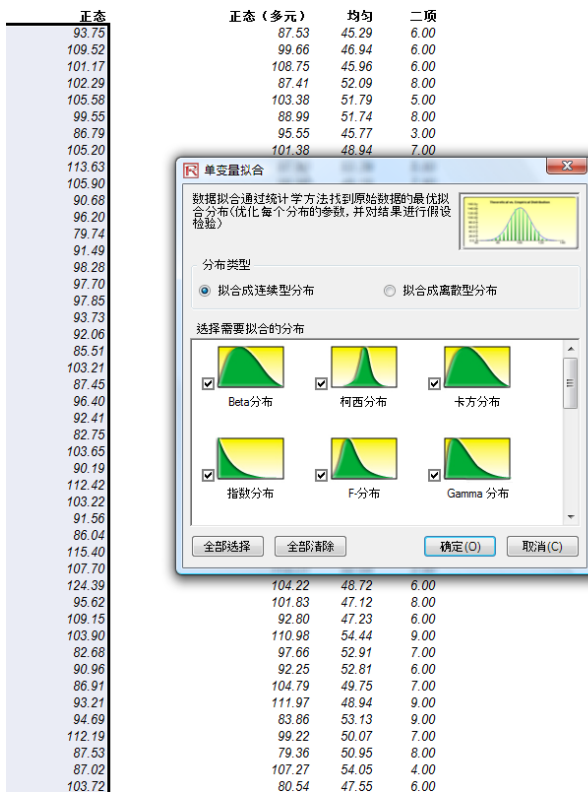


图 5.13——单变量分布拟合

结果解析:

用于检验的零假设是：样本的拟合分布与总体的分布是一致的。因此，如果计算出的 p 值小于临界  $\alpha$  水平（一般为 0.10 或 0.05），那么这个拟合的分布是不可信的。反之，p 值越大，分布拟合的越合理。概略的，您可以将 p 值看作是解释百分比，也就是说，如果 p 值为 0.9727（图 5.14），那么设置一个均值为 99.28，标准差为 10.17 的正态分布可以解释 97.27% 的数据变异，那么说明这是一个很好的拟合。结果（图 5.14）和报告（图 5.15）中都有检验量，p 值，理论统计量（基于所选择的分布），经验统计量（基于原始数据），原始数据（用于记录使用的数据），以及假设变量的相关分布参数（例如，如果选择了自动生成假定选项并且已经存在仿真文件）。结果中还对所有选择的分布按照它们的拟合情况进行了排名。

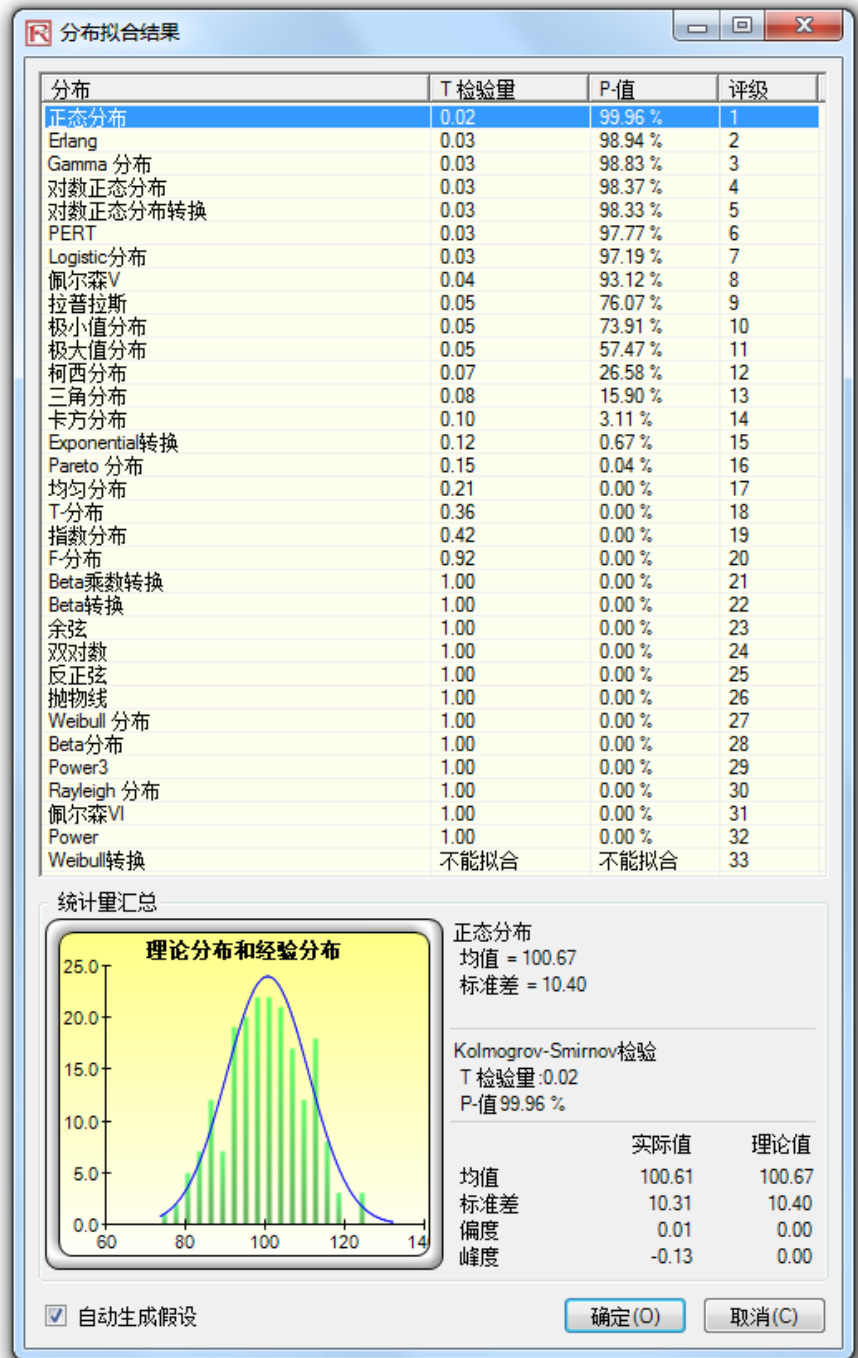


图 5.14: 分布拟合结果

单变量分布拟合

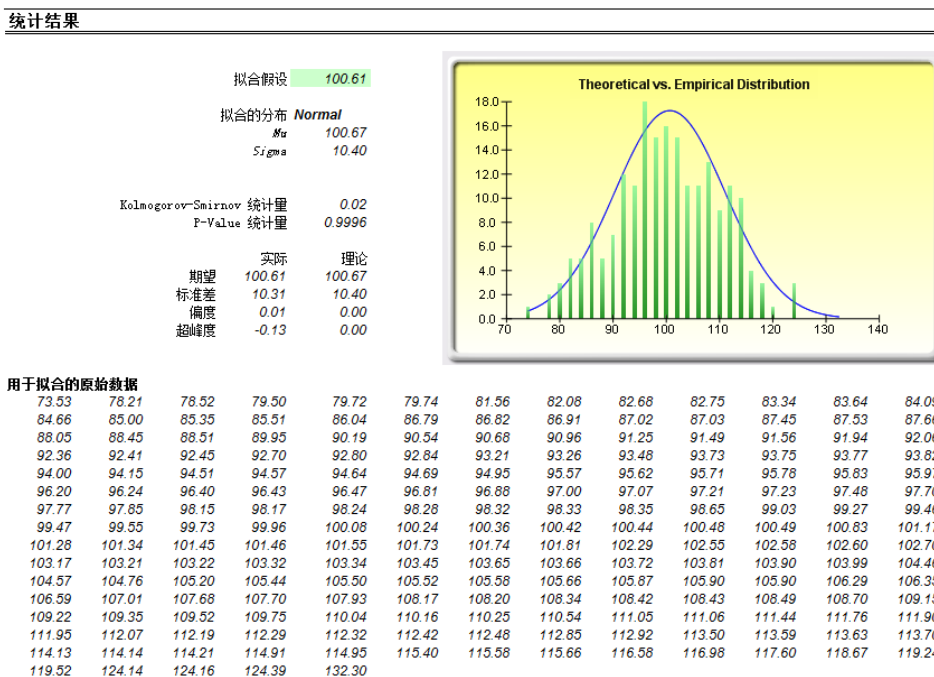


图 5.15: 分布拟合报告

多元变量的拟合过程与单个变量的拟合非常类似。但是，数据必须按列排列（每个变量被分配在一列）并且所有变量一次只能拟合一个概率分布。

步骤:

- 打开一个包含数据的工作簿
- 选择您想要拟合的数据（数据必须是同一列的）
- 选择仿真|工具|分布拟合（多元变量）
- 查看拟合的结果，选择您想要的相关分布点击**确定**

注意:

注意到分布拟合程序中所使用的统计排名方法是卡方检验和柯尔莫诺夫-斯米尔诺夫检验。前者用于检验离散分布，后者用于检验连续分布。简单来说，具有内部优化规则的假设检验被用于寻找每种被检验分布的最优拟合参数，并将结果由好到坏排列。

## 5.4 Bootstrap 仿真（拔靴法）

理论：

Bootstrap 仿真用来估计统计预测量或其它样本原始数据可靠性或精确度的一种简单方法，一般来说 Bootstrap 仿真被用于假设检验中。过去传统的方法都是依靠数学公式来描述样本统计量的精确度。这些方法假定样本统计量的分布接近正态分布，这样统计量的标准误差或置信区间的计算就相对容易一些。但是，当统计样本的分布不属于正态分布或不容易发现时，这些传统的方法就无法使用了。相反，Bootstrap 通过反复取样并从每次取样的不同样本中创造分布来对样本统计进行经验分析。

步骤：

- 运行仿真
- 选择仿真|工具|非参数 Bootstrap
- 选择一个预测来进行 Bootstrap，选择 Bootstrap 的统计量，输入需要进行 Bootstrap 的次数点击**确定**（下图 5.16）

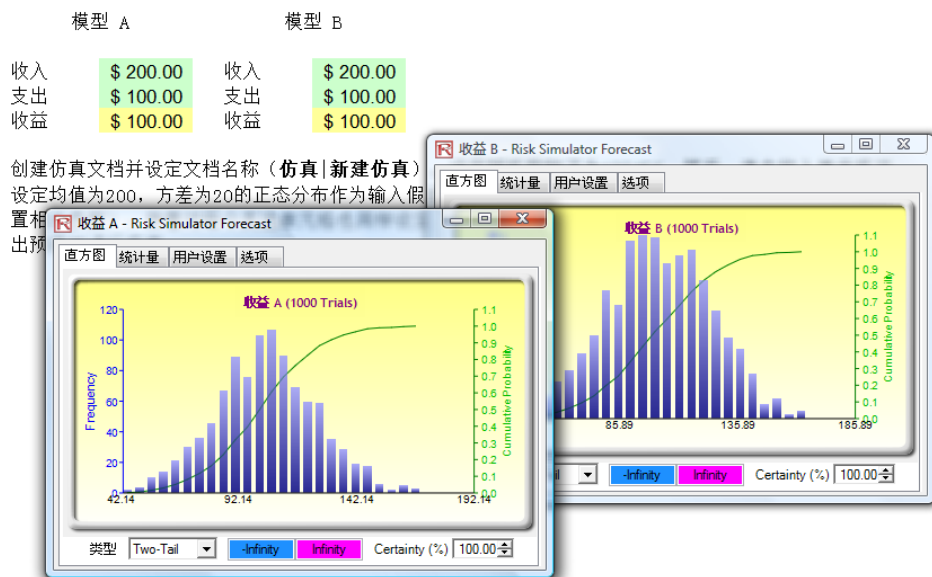


图 5.16——非参数 Bootstrap



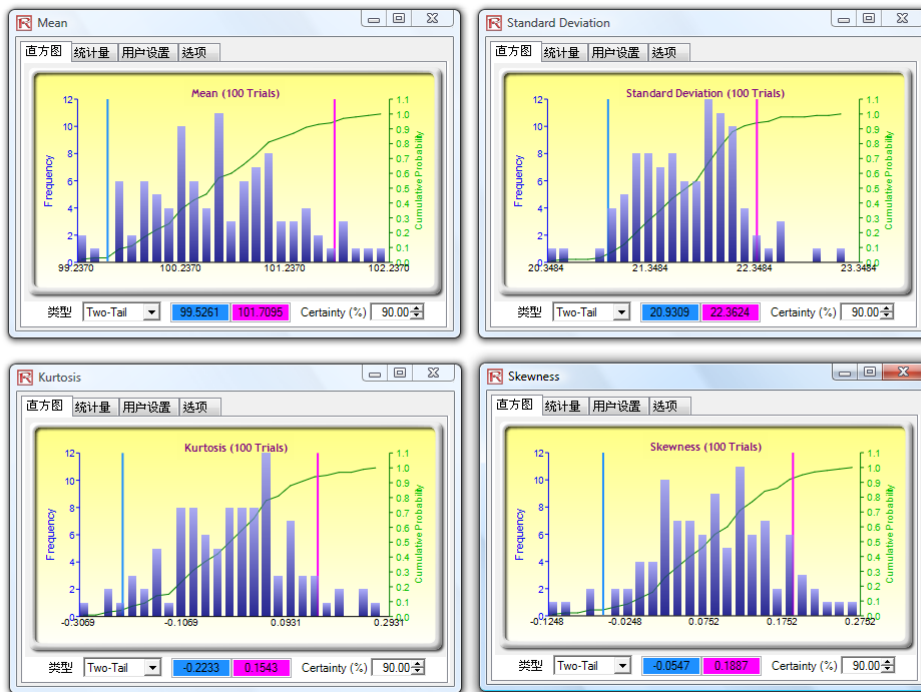


图 5.17——Bootstrap 模拟结果

结果解析:

实际上，非参数 Bootstrap 仿真可以被看作是仿真的仿真。因此，在运行一次仿真之后，结果的统计量会显示出来，但是有时这些统计量的精确度和它们的统计显著性却让人怀疑。例如，如果一次模拟后得到的偏度值为-0.10，那么这个分布真的是负偏的，还是因为随机的影响造成的轻微负值呢？-0.15，-0.20 等等呢？也就是说这个分布被认为是负偏的程度是多少呢？其它的统计量也存在同样的问题。如果计算出的统计量相同，那么可以说一个分布在统计上与另一个分布是相同的，或是它们是显著不同的？图 5.17 是部分 Bootstrap 的结果。例如，置信度为 90%的偏度统计值位于-0.2233 和 0.1543 之间，0 落在这个区间，意味着在 90%的置信水平上，预测的偏度统计上不显著区别于 0，或是这个分布可以被认为是对称的和非偏的。相反的，如果 0 落在这个区间之外，那么反面就成立，分布是有偏的（如果预测值为正，就是正偏，如果预测值为负，就是负偏）。

注意:

Bootstrap 这个词来自于一句俗语，“拎着鞋带把自己提起来”，这种方法利用自身统计量的分布来分析统计量的精确程度。非参数仿真就是简单的从一个大篮子里随机取出高尔夫球然后放回，每个高尔夫球都基于一个历史数据点。假设篮子里一共有 365 个高尔夫球（代表 365 个历史数据点）。想象一下将您每次随机取出的球都记录在一块大黑板上。有放回取出的 365 个球的结果被记录在黑板上的第一列，一共有 365 行。计算出这 365 行的相关统计量（例如均值、中值、标准差等），然后将这个过程重复 5000 次。现在黑板上有 365 行和 5000 列数据。所以我们会得到 5000 个被制成表格的统计量（有 5000 个均值、5000 个中值、5000 个标准差等等）及它们的分布，也计算出统计量的其它相关统计量，从这些结果中我们可以

看出这些仿真统计量的置信度。换句话说，在一个 10000 次试验的仿真中，得出预测结果的均值为 5.00 美元。那么这个结果的可信度是多少呢？Bootstrap 法允许使用者计算均值的置信区间，统计量的分布等等。由于根据统计学中的大数定理和中心极限定理，样本均值的均值是无偏估计量，当样本空间增大的时候，它趋近真实的总体均值，所以 Bootstrap 的结果是重要的

### 5.5 假设检验

理论：

假设检验就是通过检验两个分布的均值和方差来判断这两个分布在统计上是否是一致的。也就是找出对均值和方差的不同预测是由于随机的原因还是由于它们之间的统计显著性差别的原因造成的。

步骤：

- 运行仿真
- 选择仿真|工具|假设检验
- 每次只选择两个预测量来进行检验，选择您想要运行的假设检验类型，点击**确定**（图 5.18）

模型 A		模型 B	
收入	\$ 200.00	收入	\$ 200.00
支出	\$ 100.00	支出	\$ 100.00
收益	\$ 100.00	收益	\$ 100.00

创建仿真文档并设定文档名称（**仿真|新建仿真**）  
 设定均值为200，方差为20的正态分布作为输入假  
 置相关参数）。接着对每个支出单元格也同样设定  
 出预测并进行仿真。

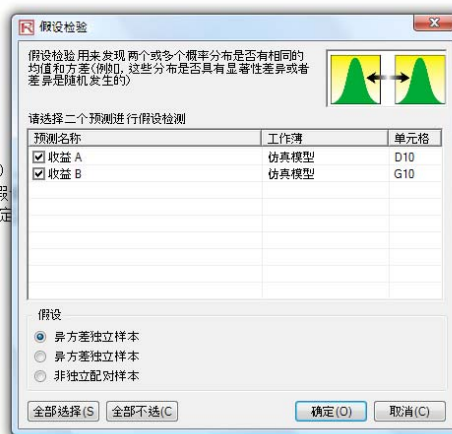


图 5.18——假设检验

结果解析：

双尾假设检验使用零假设 (Ho)：两变量的总体均值是统计一致的。它的备择假设是两者的均值是不一致的。如果计算的 p 值小于或等于 0.01、0.05 或 0.10，这意味着我们要抛弃零假设，也就是说预测平均值在 1%、5%和 10% 的显著性水平是统计显著不一致的。当 p 值较高时，我们就不能抛弃零假设，此时两个预测的分布就是统计一致的。再对两个预测的方差使用 F 检验进行相同的分析。如果得到的 p 值很小，说明方差（标准差）统计不一致的，相反的，对于较大的 p 值，两方差是统计一致的。

## 两预测变量均值和方差的假设检验

## 统计汇总

假设检验就是通过检验两个分布的均值和方差来判断这两个分布是否在统计上是一致的还是不同。也就是说，对均值和方差的不同预测是由于随机的原因还是由于它们之间的统计显著性差别的原因造成的。当预测的分布是来自于不同的样本空间时（例如，在两个不同的地点，在两个不同的商业运作单元搜集的数据等等），使用不等方差的双变量 t 检验（预测1的样本空间方差与预测2的样本空间方差不一致的）。当预测的分布来自两个类似样本空间时（从类似规格的两个不同机械部件上搜集的数据等等），使用等方差的 t 检验（预测1的样本空间方差与预测2的样本空间方差是一致的）。当预测的分布来自同一个样本空间时（在不同情况下从同一组客户那里取得的数据等）可以使用配对双变量 t 检验。

双尾假设检验使用的零假设（ $H_0$ ）是：两个变量的总体均值是统计一致的。备择假设是两者的均值是不一致的。如果计算的 p 值小于或等于 0.01, 0.05 或 0.10, 这意味着我们要抛弃零假设，也就是说预测平均值在 1%, 5% 和 10% 的显著水平上是统计显著不一致的。当 p 值较高时，我们就不能抛弃零假设，此时两个预测的分布就是统计一致的。再对两个预测的方差使用 p 检验进行相同的分析。如果得到的 p 值很小，说明方差（标准差）统计不一致的，相反的，对于较大的 p 值，两方差是统计一致的。

## 结果

假设检验假定	不等方差
t 统计量	1.015722
t 统计量的 P 值	0.309885
F 统计量	1.063476
F 统计量的 P 值	0.330914

图 5.19——假设检验结果

## 注意：

当预测的分布是来自于不同的样本空间时（例如，从两个不同的地点，两个不同的商业运作单元搜集的数据等等），使用不等方差的双变量 t 检验（预测 1 的样本空间方差与预测 2 的样本空间方差不一致的）。当预测的分布来自两个类似的样本空间时（从类似规格的两个不同机械部件上搜集的数据等等），使用等方差的 t 检验（预测 1 的样本空间方差与预测 2 的样本空间方差是一致的）。当预测的分布来自同一个样本空间时（在不同情况下从同一组客户那里取得的数据等）可以使用配对双变量 t 检验。

## 5.6 数据输出和保存仿真结果

使用 Risk Simulator 的数据提取功能可以很容易地提取出仿真的原始数据。假设和预测都可以被提取，但是首先要运行一次仿真。提取的数据可以用于其它的一些分析。

## 步骤：

- 打开或新建一个模型，定义输入和预测，运行仿真
- 选择仿真|工具|数据提取
- 选择想要的假设和预测提取数据，点击**确定**

可以按照不同的形式来提取数据：

- 新工作簿中的原始数据，其中的模拟值（假设和预测）可以被保存并用于进一步分析
- 保存为纯文本文件，这样可以直接导入其它的分析软件
- Risk Simulator 文件，可以在稍后通过选择**仿真|工具|打开数据|导入**来恢复结果数据（假设和预测）

第三个选项是最常用的选项，也就是说，将模拟的结果保存为\*.risksim 文件，这样稍后

可以恢复数据，不必重新运行一次模拟。图 5.21 是提取、导出和保存仿真结果的对话框。

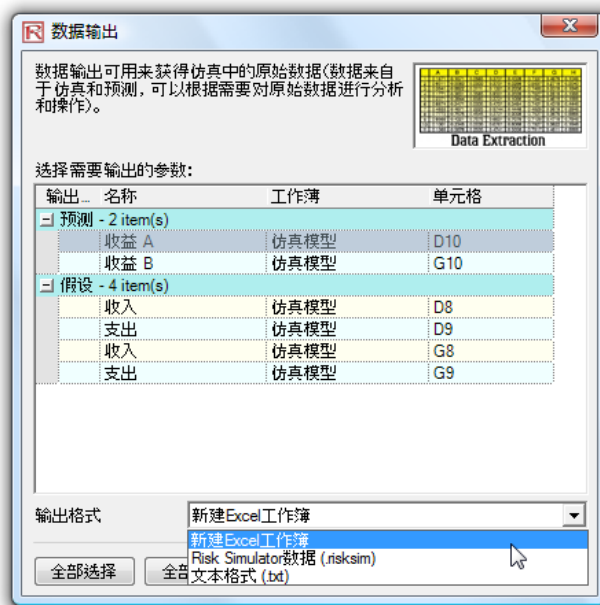


图 5.21——样本模拟报告

## 5.7 创建报告

在运行仿真之后,您可以生成一份关于假设,预测和仿真结果的报告。

步骤:

- 打开或新建一个模型,定义假设和预测,然后运行仿真
- 选择仿真|生成报告

Simulation - Hypothesis Testing

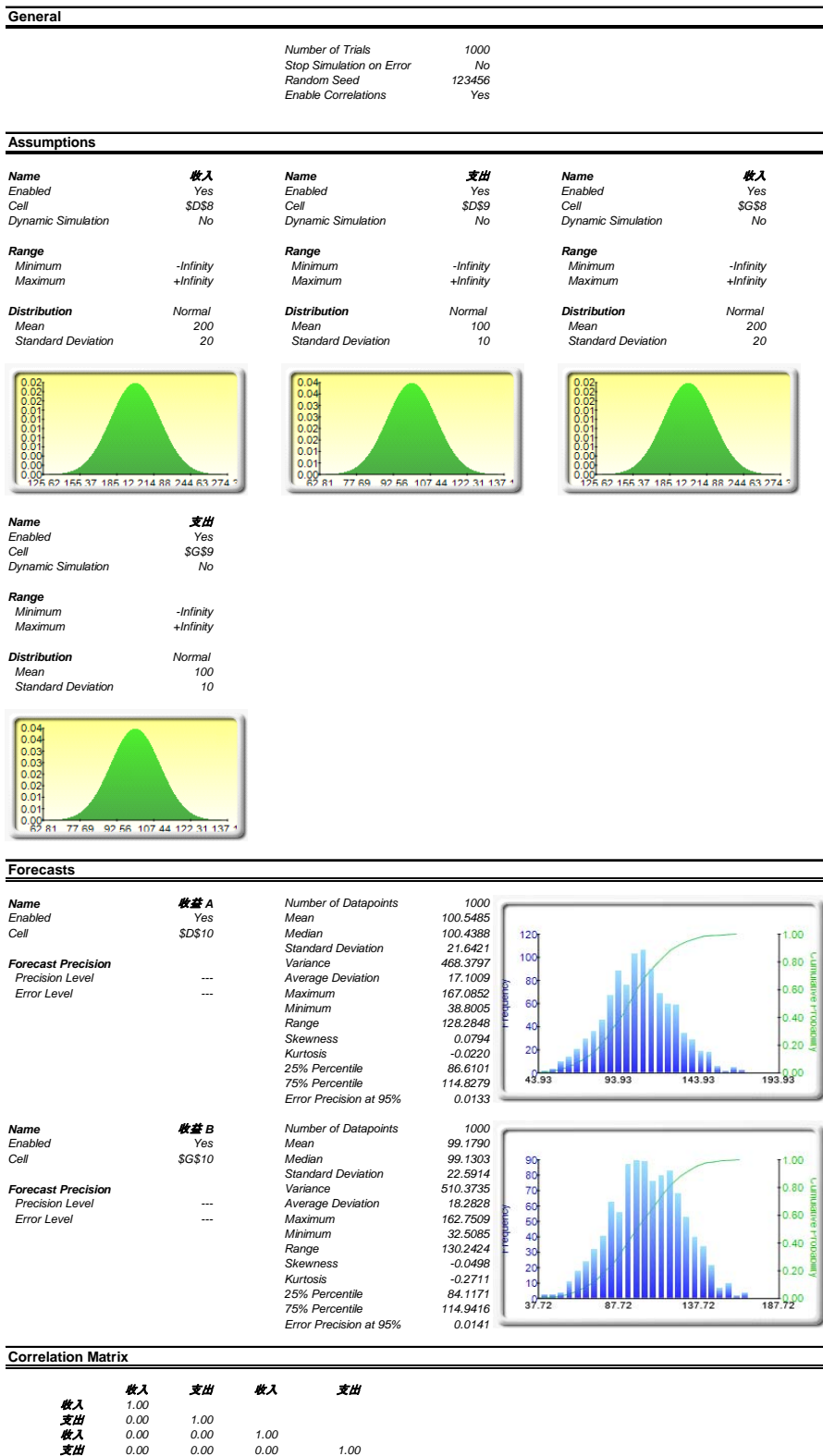


图 5.21——样本仿真报告

## 5.8 回归和预测诊断工具

Risk Simulator 中的高级分析工具可以用来决定数据的计量经济学特性。诊断工具包括测定数据的异方差性，非线性，异常性，规格误差，微数缺失性，平稳和随机性，误差的正态性和球形，以及多重共线性。每个检验在各自模型的报告中都有详细地描述。

操作过程描述:

- 打开示例模型 (Risk Simulator | 示例模型 | 回归诊断) 点击时间序列数据工作表然后选择数据包括数据变量的名称(单元格 C5:H55)。
- 点击 Risk Simulator | 工具 | 诊断工具。
- 点击数据在下拉菜单中选择因变量 Y。点击**确定**完成选择 (图 5.22)。

多元回归分析数据

因变量 Y	变量 X1	变量 X2	变量 X3	变量 X4	变量 X5
521	18308	185	4.041	79.6	7.2
367	1148	600	0.55	1	8.5
443	18068	372	3.665	32.3	5.7
365	7729	142	2.351	45.1	7.3
614	100484	432	29.76	190.8	7.5
385	16728	290	3.294	31.8	5
286	14630	346	3.287	678.4	6.7
397	4008	328	0.666	340.8	6.2
764	38927	354			
427	22322	266			
153	3711	320			
231	3136	197			
524	50508	266			
328	28886	173			
240	16996	190			
286	13035	239			
285	12973	190			
569	16309	241			
96	5227	189			
498	19235	358			
481	44487	315			
468	44213	303			
177	23619	228			
198	9106	134			
458	24917	189			
108	3872	196			
246	8945	183			
291	2373	417			
68	7128	233			
311	23624	349			

Figure 5.22 –运行数据诊断工具

在预测和回归分析中最常见的错误是异方差，也就是说，误差的标准差随着时间的增加不断变大。(参看图 5.23 使用诊断工具作为测试的结果)。视觉上来说，数据在竖直方向上的波动宽度随着时间不断增大或成扇形散开，并且明显地，可决系数 ( $R^2$ ) 当异方差存在时显著下降。如果因变量的标准差不是一个常数，误差的标准差也将不是一个常数。除非因变量的异方差性是显著的，否则它的效果不会非常剧烈：最小二乘法估计仍然是无偏的，当误差是正态分布时，斜率和截距的估计将是正态分布的。当误差不是正态分布时，斜率和截距的估计也将是渐进正态分布的(当数据点的个数很大时)。斜率方差和整体方差的估计量将是不精确的，但是如果自变量的值是关于它们的均值对称，这种不精确性就可能不那么重要了。

如果数据量很少(微数缺失性)，那么就很难判断是否违背使用这些模型的假设。而且在存在违背模型假设的情况下，非正态性或方差的异方差性也是很难察觉的。

尤其对线性回归模型而言，数据点比较少时，很难保证不违背模型假设。同时，这时候往往很难决定是用直线拟合数据点效果好，还是使用非线性函数（曲线）效果好。即使所有假设检验都是符合的，小样本的线性回归也可能没有足够有效性来判断斜率是否为 0。这种有效性与残差项、自变量的方差，假设检验的置信水平和数据点的个数有关。当残差增大时，或置信水平减小时（比如，假设检验更严格），这种有效性减小。当自变量的方差增加或数据点个数增加时，这种有效性增加。

由于异常的存在，数据可能分布上并不一致。异常是数据中那些不正常的值。异常可能会对拟合的斜率和截距产生很强的影响，并且使得对大块数据的拟合较差。异常的存在倾向于增大预测的残差，减小拒绝零假设的概率。比如，产生更高的预测误差。异常的产生可能由于记录误差造成的（可纠正的），也可能由于因变量的值并不是全都从同一个分布中取样的。显然地，异常也可能是由于因变量的值是从一个非正态的总体样本中取得而造成的。但是，自变量和因变量的散列点里面的不寻常的值也可能并不是异常。在回归分析中，拟合的直线对异常是非常敏感的。换句话说就是，最小二乘回归和拟合斜率的估计抗异常的能力较差。一个数据点从另外的一些点中竖直地移动下来会造成拟合的直线更贴近这个点，而不是跟随剩下这些点的线性趋势，尤其当这个点是水平地远离另外一些点的中心时。

但是，当去除这些异常时，必须非常小心。尽管在大多数情况下，当异常被去除后，回归的结果往往看起来更好，但是在此之前必须进行先验论证。举例来说，在对某一特别公司的股票收益表现进行回归时，由于股票市场低迷造成的异常应该被保留下来。这些并不是真正的异常，只是体现了商业周期的必然性。在回归时去掉这些异常来预测如基于公司股票的退休金将有可能产生不正确的结果。但是，如果这些异常是由于不可重现的商业条件（比如，企业的并购）及不可重现的商业结构的变化造成时，往往需要在使用回归分析前去除这些异常点。这里所讲述的内容只涉及如何辨认异常数据，在实际应用中仍需要由使用者根据实际情况来决定保留还是去除这些异常点。

有时候，因变量和自变量之间的关系用非线性关系来描述比线性关系更合适。那么，对于这些情况，使用线性回归将不是最优的选择。如果线性模型不是正确的选择，那么斜率和截距的估计和线性回归的拟合值就是有偏的，同时拟合的斜率和截距估计量将是没有意义的。当规定自变量和变量的范围时，非线性模型可以近似地看成线性模型（事实上，这是线性截距的主要部分），但是要精确预测就要选择对数据解释合理的模型。在使用回归线，必须先对数据进行非线性变换。这方面，一个简单的方法是对自变量进行对数运算（另外的方法包括对因变量进行开根或平方、三次方运算），并且对预测量使用这些非线性变化后的数据进行回归。

诊断结果								
变量	异方差		微数缺测性	异常		潜在的异常数	非线性	
	W检验 p值	假设检验 结果	近似 结果	本性 下届	本性 上界		非线性检验 p值	假设检验 结果
Y			没问题	-7.86	671.70	2		
变量 X1	0.2543	同方差	没问题	-21377.95	64713.03	3	0.2458	线性
变量 X2	0.3371	同方差	没问题	77.47	445.93	2	<b>0.0335</b>	非线性
变量 X3	0.3649	同方差	没问题	-5.77	15.69	3	<b>0.0305</b>	非线性
变量 X4	0.3066	同方差	没问题	-295.96	628.21	4	0.9298	线性
变量 X5	0.2495	同方差	没问题	3.35	9.38	3	0.2727	线性

图 5.23 –异常性，异方差性，微数缺测性，和非线性的检验结果

在预测时间序列数据时一个典型的问题是这些自变量是相互独立的还是存在某种相互关系。因变量的时间序列数据可能是自相关的。对那存在序列相关的因变量数据，斜率和截距的估计是无偏的，但是得到的预测值和方差是不可靠的，因此拟合的统计检验有效性是有缺陷的。譬如，利率、通货膨胀、销售量、收入、和其它诸如此类的时间序列数据明显是自相关的，现阶段的值和前一阶段的价值有关（显然，三月的通货膨胀数据和二月份的通货膨胀数据有关，二月份的和一月份的有关，这种相互关系可一直找寻下去）。

如果忽略这种关系，那得到的预测是有偏的，也是不精确的。对这样的情况，使用自回归模型或自回归求和滑动平均效果将会更好（**Risk Simulator** | 预测 | ARIMA）。最后，需要指出，对那些非稳态数据的自相关函数趋向于较慢的衰减（参见非稳态报告）。

例如，如果自相关 AC (1) 为非零值，意味着序列是一阶序列相关的。如果 AC 随着滞后的增加呈几何下降趋势，这意味着序列遵循一个低阶自回归过程。如果经过几次滞后之后 AC 值趋于 0，这意味着序列遵循一个低阶移动平均过程。相反的，PAC 衡量了在移除了滞后干扰后的 k 阶相关值。如果自相关模式可以通过小于 k 阶的自回归解决，那么 k 阶滞后的部分自相关值趋近于 0。报告里同时还提供了 Ljung-Box 的 k 阶滞后 Q 统计值和 p 值，此时被检验的原假设是 k 阶时不存在自相关。自相关的虚线图近似在两个标准差的范围。如果自相关值在此范围之内，那么在 5% 的显著性水平内它不显著区别于 0。寻找到合适的 ARIMA 模型需要尝试和经验。AC, PAC, SC 和 AIC 都是识别正确模型的有效诊断工具。

自相关用来测量自变量 Y 的现在的数值和过去的数值的相互关系。与这种分布上滞后相对应的是自变量和不同因变量 X 之间的时间滞后关系。举例来说，抵押率的走势往往会跟从联邦准备金率的走势，但是会有时间上的滞后性（典型的是 1 到 3 个月）。有时，时间滞后带有周期性和季节性（例如，冰激凌的销售量会在夏季的月份里达到最大，因此会和 12 月前夏季的销售量有关）。下面的分布滞后分析显示了在各种时间滞后下（这里的滞后是同时发生的），因变量和每个自变量之间的关系，并判断那些时间滞后是统计上显著的，应该被考虑。



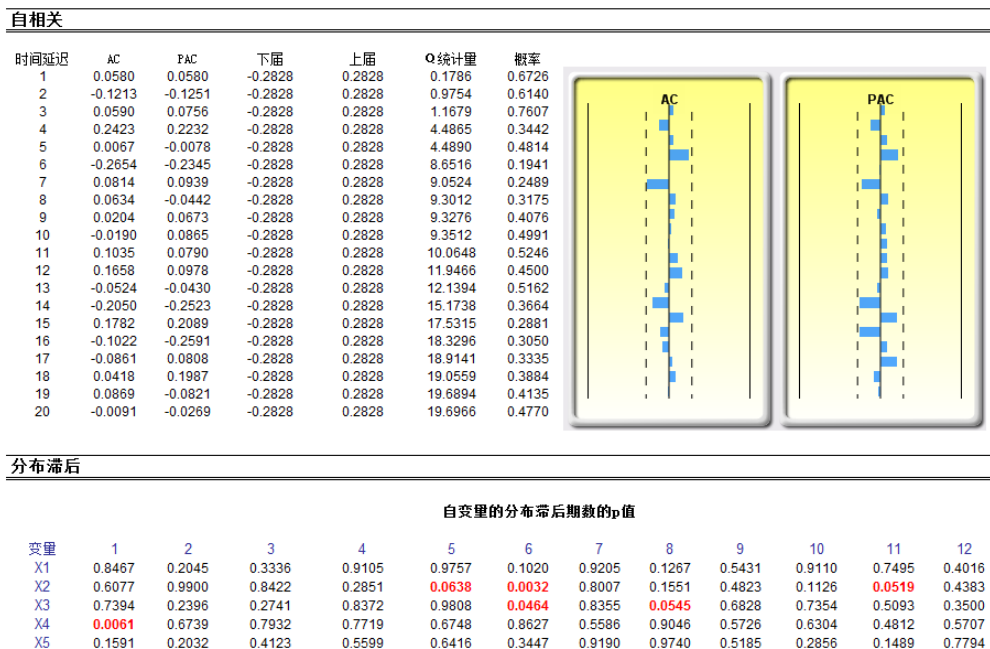


图 5.24 – 自回归分布滞后结果

在运行回归模型时的另外一个假设是误差项的正态性和球形假设。如果正态性假设违背或出现异常，那么线性回归模型的拟合度检验（用来表明是否采取线性拟合）就可能不是判别模型好坏的最有效的检验方法。如果误差不是独立的和正态分布的，这就可能表明数据可能是自相关的，或含有非线性或者另外的更具破坏性的误差。误差的独立性也可通过异方差性检验（图 5.25）来探测。

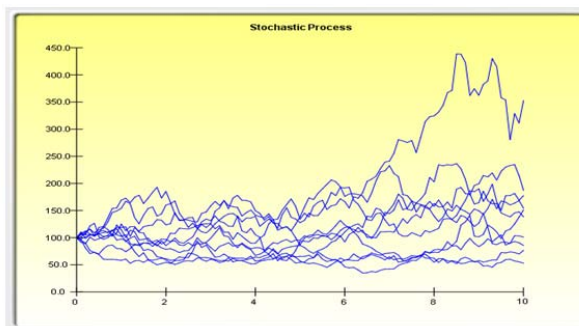
误差的正态性检验是一个非参数检验，该方法并不需要样本总体形状的假设，对小样本数据的分析也可使用。该检验使用的零假设是样本误差服从正态分布的，备择假设是样本误差不是正态分布的。如果在各种有效性水平下，计算得到的 D 统计量都大于或等于 D 临界值，那就拒绝零假设并接受备择假设（误差不是正态分布的）。相反地，如果 D 统计量小于 D 临界值，那么就不拒绝零假设（误差是正态分布的）。该检验依靠两类累积频率：一种来自于样本数据集，另一种来自于基于样本均值和方差的理论分布。

检验结果					
	误差	相关频率	观察值	预期值	O-E
回归平均误差	0.00				
误差的标准差	141.83	-219.04	0.02	0.0612	-0.0412
D 统计量	0.1036	-202.53	0.02	0.0766	-0.0366
1%下的D临界值	0.1138	-186.04	0.02	0.0948	-0.0348
5%下的D临界值	0.1225	-174.17	0.02	0.1097	-0.0297
10%下的D临界值	0.1458	-162.13	0.02	0.1265	-0.0265
零假设：误差是正态分布的。		-161.62	0.02	0.1272	-0.0072
		-160.39	0.02	0.1291	0.0109
		-145.40	0.02	0.1526	0.0074
<b>结论：误差是正态分布的在 1% 显著性水平下。</b>		-138.92	0.02	0.1637	0.0163
		-133.81	0.02	0.1727	0.0273
		-120.76	0.02	0.1973	0.0227
		-120.12	0.02	0.1985	0.0415

图 5.25 – 误差的正态性检验

有时，某种类型的时间序列数据由于其代表的事件在本质上是随机的，从而使得除了随机过程外不能使用其它的方法来建模。举例来说，由于股票价格，利率，石油价格和其它商品的价格是高度不确定和波动的，你并不能通过使用单变量回归模型来充分地模拟和预测这些值。换句话说，这些过程不是平稳的。平稳性这里通过使用 Runs 检验来测试，并且在自相关报告中也可得到其它的可视化效果（自相关系数趋向于缓慢地减少）。随机过程是指服从某一概率分布的一系列事件和轨迹的集合。就是说，随机事件虽然随着时间发生但这些事件都符合某一特殊的统计和概率法则。我们主要碰到的随机过程包括随机游走（布朗运动），均值回复，跳跃-扩散这些。这些过程能用来预测大多数体现随机倾向但又服从概率分布的变量的变化。产生该过程的方程是实现已知的，但是产生的结果确实未知的。（图 5.26）

随机游走（布朗运动）能用来预测股票的价格，商品的价格和任何沿着漂移路径有着漂移或增长率和波动率的随机时间序列数据。均值回复可以通过远期目标水平来减小随机游走的波动。该过程可以用来预测如利率，通胀这些有长期目标水平（这些长期目标水平由权威机构和市场提供）的时间序列变量。跳跃-扩散过程可以用来预测如石油价格，电力价格（个别外部事件的发生能使价格往上跳跃或下降）这些偶尔伴有随机跳跃的时间序列数据。最后，这三类随机过程可根据需要相互配合使用。



统计汇总

下面给出的是对上述数据通过随机过程得到的估计参数。这可由您来决定是否拟合的概率（和拟合度计算相似）能充分保证使用随机过程来预测。如果可以的话，那么这个随机过程是随机游走，均值回复，跳跃扩散，还是它们的联合模型。您必须依靠数据过去的表现，经济学上的预先判断和金融预期来选择正确的随机过程模型。这些参数可设置到随机过程预测里（仿真|预测|随机过程）。

周期		回复率	283.89%	跳跃率	20.41%
漂移率	-1.48%	远期值	327.72	跳跃大小	237.89
波动率	88.84%				

随机模型拟合的概率： 46.48%  
高的拟合值意味着随机模型比传统模型要好

Runs	20	标准正态	-1.7321
正的	25	P值 (1-tail)	0.0416
负的	25	P值 (2-tail)	0.0833
期望的 Run	26		

P值较低（小于0.10, 0.05, 0.01）意味着序列不是随机的，因此涉及到平稳性问题，那么ARIMA模型可能更合适。相反地，P值较大意味着随机模型更合适。

图 5.26 – 随机过程参数检验

当自变量之间存在线性关系时，就称之为存在多重共线性。当处于这样的情况时，就不能使用回归模型来估计了。在共线性情况下，回归估计是有偏的，结果是不精确的。当使用逐步回归时，得到的结果就会像上述描述的情况，统计上有效的自变量会很早地从回归模型中剔除，使得得到的结果既不有效，也不精确。一种在多元回归模型中判断多重共线性的快速有效的方法是查看当t统计量比较小时，R方的值是否比较大。

另一种快速的检验方式是创建自变量的相关性矩阵。相互的相关性比较高意味着潜在的自相关性。当相关系数的绝对值大于 0.75 时，就认为自变量间存在剧烈的多重共线性。多重共线性的另一种检验方法是通过计算方差膨胀因子（VIF）。可通过对每个自变量用其它自变量来回归后，得到 R 方来计算 VIF。如果 VIF 大于 2.0，表明存在剧烈的多重共线性。如果 VIF 大于 10.0，表明存在破坏性的多重共线性。（图 5.27）

相关性矩阵				
相关性	X2	X3	X4	X5
X1	0.333	0.959	0.242	0.237
X2	1.000	0.349	0.319	0.120
X3		1.000	0.196	0.227
X4			1.000	0.290

方差膨胀因子				
VIF	X2	X3	X4	X5
X1	1.12	12.46	1.06	1.06
X2	N/A	1.14	1.11	1.01
X3		N/A	1.04	1.05
X4			N/A	1.09

图 5.27 – 多重共线性误差

相关性矩阵通过变量之间的 Pearson 乘积项（一般被称为 Pearson 系数）来表示变量之间的相互关系。这些相关系数取值在-1 到 1 之间，包括两个端点。它们的符号反映了变量之间联系的方向性，大小则反映了这种联系的强弱。Pearson 的相关系数仅仅测量了线性相关性，对非线性相关性不是很有效。

双尾假设检验用来判断变量之间的相互关系是否是显著的，检验结果的 P 值被显示出来。P 值小于 0.1, 0.05 和 0.01 的用蓝色显示，表明是统计上有效的。换句话说，某个相关变量的 p 值小于给定的有效性水平时，表明变量之间的相互关系是统计上显著不同于 0 的，表明两个变量之间存在显著的线性关系。

两个变量（x 和 y）的 Pearson 乘积项系数（R）是和协方差（cov）有关，表达式

为：
$$R_{x,y} = \frac{COV_{x,y}}{s_x s_y}$$
。这里协方差除以两个变量的标准差（s）的优势是可以把相

关系数控制在-1 到 1 的范围内。这使得这种测量方法能较好地反映不同变量之间

的相互关系（尤其对那些不同单位和量级的变量）。Spearman 基于排列的非参数相关性也在下面给出。Spearman 的相关系数是通过将数据先排列，然后在求排列的这种相关性得到的。当变量之间存在非线性关系时，排列的相关性能提供一种更好的估计。

需要指出的是，存在显著的相关性并不意味着变量之间含有因果关系。变量之间的这种联系并不表示改变一个变量的值，另一个变量也会改变。当两个变量以相关的路径各自独立地变动时，它们可能是相关的，但这种相关关系可能是没有道理的（例如，太阳黑子的数目和股票的市场价格存在较强的相关性，但是这里面并不存在因果关系，这种相关性完全是伪造的）。

### 5.9 统计分析工具

Risk Simulator 软件中另外一个非常有用的工具就是统计分析工具，它可以发现数据的统计特征。诊断工具运行包括对数据的统计特性进行描述，对随机数据进行基本的描述性统计检验和校正。

#### 操作过程描述：

- 打开示例模型 (**Risk Simulator | 示例模型 | 统计分析**) 回到数据工作簿，选择数据包括变量名称(单元格 **C5:E55**)。
- 点击 **Risk Simulator | 工具 | 统计分析** (图 5.28)。
- 点击 **数据类型**，选择的数据来自于一列或者多列。本例中，数据来自于多列。点击**确定**完成选择。
- 选择想要进行的统计检验。建议（软件默认）是选择所有的检验。点击**确定**完成选择（图 5.29）。

请参阅生成的报告更好的理解统计检验结果的意义。（示例报告显示如下图 5.30-5.33）

数据集

变量 X1	变量 X2	变量 X3
521	18308	185
367	1148	600
443	18068	372
365	7729	142
614		
385		
286		
397		
764		
427		
153		
231		
524		
328		
240		
286		
285		
569		
96		
498		
481		
468		
177		
198		
458		
108		
246		
204		
	8945	183
	2272	417

**统计分析**

此工具用来预测和发现一系列原始数据之间的统计关系。

选取数据

变量 X1	变量 X2	变量 X3
521	18308	185
367	1148	600
443	18068	372
365	7729	142
614	100484	432
385	16728	290
286	14630	346
397	4008	328
764	38927	354
427	22322	266
153	3711	320
231	3136	197

数据来自于单一变量  
 数据包含多栏中的多个变量

图 5.28 – 运行统计分析工具

**统计分析**

选择需要运行的分析

<input checked="" type="checkbox"/> 描述性统计	<input checked="" type="checkbox"/> 随机过程参数估计
<input checked="" type="checkbox"/> 分布拟合	周期性 <input type="text" value="Annual"/>
<input checked="" type="radio"/> 连续的 <input type="radio"/> 离散的	<input checked="" type="checkbox"/> 时间序列自回归
<input checked="" type="checkbox"/> 直方图和其他图表	<input checked="" type="checkbox"/> 时间序列预测
<input checked="" type="checkbox"/> 假设检验	季节性(周期/循环) <input type="text" value="4"/>
假设的均值 <input type="text" value="0"/>	预测(期) <input type="text" value="4"/>
<input checked="" type="checkbox"/> 非线性外推	<input checked="" type="checkbox"/> 趋势线预测
预测(期) <input type="text" value="4"/>	预测(期) <input type="text" value="4"/>
<input checked="" type="checkbox"/> 正态性	

图 5.29 – 统计检验

描述性统计

统计分析

几乎所有的分布都可以用4个矩来描述（有的需要一个矩，有的需要两个矩，等等）。描述性统计从量上描述了这是个部分。第一矩描述了分布的位置（例如，均值，中位数，和众数）常常用来描述期望值，期望回报，或者事件发生的平均值。

算术平均数通过加总所有的数据值再除以数据的个数得到发生的平均值。几何平均数就是数据乘积的算术根，并且数据都必须是正数。在计算百分比或者比例的时候，使用几何平均数则更为精确。例如，使用几何平均数根据变化率计算平均增长率。截尾平均数是截取两边极大值之后的代数平均值。当两边的极值存在的时候，均值会向重要差异的方向倾斜，而截尾平均数在有偏的分布中避免了这个问题。

均值的标准误差计算了样本均值的误差大小。样本越大，误差越小，对于无限大的样本总体，误差就越接近于零，说明已经估计到了样本参数。由于抽样的误差，需要提供95%的置信区间用于均值的估计。基于样本数据的分析，实际的样本均值落在上下区间之内。

中位数处于数据的中间，有50%的大于它，有50%的小于它。在三个第一矩统计量中，中位数受分布两边的影响最小。对称的分布的中位数和分布的代数平均值相等。当中位数远离均值的时候就在偏斜的分布。众数测量了最为经常发生的数据点。极小值是数据中的最小值而极大值是数据中的最大值。极差就是极小值和极大值之差。

第二矩用来测量分布的幅度或者宽度，通常使用的统计量包括标准差，方差，全距，四分位间距。标准差说明所有数据与均值的高中程度。它通常作为与风险相关的一个测定值（越高的标准差意味着越广的分布，越高的风险，或者数据在均值附近分散的更广）它的单位等于原始数据的单位。样本标准差不等于总体标准差，前者使用了对于小样本的自由度，总体样本的标准差会落入这个区间，如果数据包含整个样本总体，使用总体标准差。因此，两个方差也就是各自标准差的平方。

变异系数是样本的标准差除以样本的均值，提供了一个无单位的计算结果，可以在多个分布之间比较（可以将以百万作为计数单位的分布和以十亿作为计数单位的分布做比较，或者以米和千米做单位的分布进行比较，等等）。第一分位数用来度量按照数据从小到大排列的第二十五百分数。第叁分位数是七十五百分数。有时候，分位数可以作为分布的上下限，因为它截取了数据不去考虑分布的两端。四分位间距不同于第一和第三分位数，常用来度量分布中心范围的宽度。

偏度是分布的第三矩。偏度体现了分布不对称的特性。正偏斜说明分布的尾部不对称向正值偏斜。负偏斜说明尾部不对称向负值偏斜。

峰度体现了分布与正态分布相比较平坦或者突兀的特性。是分布的第四矩。正的峰度说明分布相当的突兀。负的峰度说明分布相当的平。这里测量的峰度以零为中心（峰度也可以以3.0为中心）。两者都是有效的测定标准，以零为标准的便于解释。较高的正峰度说明分布的中心部分较尖，尾部较平坦。说明极端事件的出现具有较高的可能性（例如，灾难事件，恐怖袭击，股票市场的大幅下跌）而非正态分布表现的情形。

统计汇总

统计量	变量 X1		
观测值	50 0000	标准差 (样本)	172.9140
算术平均数	331.9200	标准差 (总体)	171.1761
几何平均数	281.3247	标准差的下置信区间	148.6090
截尾平均	325.1739	标准差的上置信区间	207.7947
算术平均数的标准差	24.4537	方差 (样本)	29899.2588
均值的下置信区间	283.0125	方差 (总体)	29301.2736
均值的上置信区间	380.8275	变异系数	0.5210
中位数	307.0000	第一四分位数 (Q1)	204.0000
众数	47.0000	第三四分位数 (Q3)	441.0000
极大值	764.0000	四分位间距	237.0000
极小值	717.0000	偏度	0.4838
极差		峰度	-0.0952

图 5.30 – 示例统计分析报告

假设检验（对于单变量样本均值的t检验）			
<b>统计汇总</b>			
来自数据样本的统计量:		计算得到的统计量	
观测值	50	t 统计量	13.5734
样本均值	331.92	P 值（右尾）	0.0000
样本标准差	172.91	P 值（左尾）	1.0000
<b>使用者使用的统计量</b>		P 值（双尾）	0.0000
假设均值	0.00	零假设 (H <sub>0</sub> ):	$\mu =$ 假设的均值
		备择假设 (H <sub>a</sub> ):	$\mu <>$ 假设的均值
		注意：“<>”代表“右尾检验中”大于的意思”，代表作为检验“小于的意思”，或者“不等于”用于双尾检验。	
<b>假设检验汇总</b>			
单变量检验适用于在不知道总体标准差，假设样本分布大致服从正态的条件下使用（t 检验用于样本的数据小于 30 的情形，事实上对于检验提供了一个保守的结果）			
<b>双尾假设检验</b>			
双尾假设检验的零假设：总体均值在统计意义上等于假设的均值。备择假设是真实的总体均值在统计上是否不等于样本的均值。使用t 检验，如果p值小于指定的显著性水平（一般为0.1，0.05，或者0.01），这就意味着在指定显著性水平10%，5%或者1%（或者90%，95%和99%的置信度），样本均值和假设的均值在统计上具有显著性差异。相反地，如果p 值大于0.1，0.05，或者0.01，总体的均值在统计上就等于假设的均值，任何差异都是随机造成的。			
<b>左尾假设检验</b>			
左尾假设使用零假设：总体均值在统计意义上小于或者等于假设的均值。备择假设是总体均值在统计上大于样本的均值。使用t 检验，如果p值小于指定的显著性水平（一般为0.1，0.05，或者0.01），这就意味着在指定显著性水平10%，5%或者1%（或者90%，95%和99%的置信度），样本均值和假设的均值在统计上具有显著性差异。相反地，如果p 值大于0.1，0.05，或者0.01总体的均值在统计上就小于或者等于假设的均值，任何差异都是随机的。			
<b>右尾假设检验</b>			
右尾假设检验使用零假设：总体均值在统计意义上大于或者等于假设的均值。备择假设是总体均值在统计上小于样本的均值。使用t 检验，如果p值小于指定的显著性水平（一般为0.1，0.05，或者0.01），这就意味着在指定显著性水平10%，5%或者1%（或者90%，95%和99%的置信度），样本均值和假设的均值在统计上具有显著性差异。相反地，如果p 值大于0.1，0.05，或者0.01总体的均值在统计上就大于或者等于假设的均值，任何差异都是随机的。			
由于t 检验更为保守并不像Z 检验那样要求已知样本的标准差，这里只用t 检验。			

图 5.31 – 示例统计分析报告 (单变量假设检验)

非线性外推法

正态性检验是非参数检验的一种，它对样本的形状不做任何假设，可以对较小的样本进行分析。检验的零假设是样本遵循正态分布，备择假设为数据的分布不遵循正态分布。如果计算得到的p值小于或者等于显著性水平的大小，就拒绝原假设，接受备择假设。否则的话，如果p值大于显著性水平的大小，就不拒绝零假设。检验取决于两个累计的频数：一种来自于样本数据集，另一种来自于基于样本均值和方差的理论分布。另外一种正态性检验的方法叫做卡方检验。卡方检验需要更多的数据进行正态性检验。

检验结果						
数据的均值	331.92	数据	相关频率	观察值	估计值	O-E
标准差	172.91	47.00	0.02	0.02	0.0497	-0.0297
D 统计量	0.0859	68.00	0.02	0.04	0.0635	-0.0235
1%显著性水平的D值	0.1150	87.00	0.02	0.06	0.0783	-0.0183
5%显著性水平的D值	0.1237	96.00	0.02	0.08	0.0862	-0.0062
10%显著性水平的D值	0.1473	102.00	0.02	0.10	0.0918	0.0082
零假设：数据的分布服从正态分布		108.00	0.02	0.12	0.0977	0.0223
		114.00	0.02	0.14	0.1038	0.0362
		127.00	0.02	0.16	0.1180	0.0420
		153.00	0.02	0.18	0.1504	0.0296
		177.00	0.02	0.20	0.1851	0.0149
		186.00	0.02	0.22	0.1994	0.0206
		188.00	0.02	0.24	0.2026	0.0374
		198.00	0.02	0.26	0.2193	0.0407
		222.00	0.02	0.28	0.2625	0.0175
		231.00	0.02	0.30	0.2797	0.0203
		240.00	0.02	0.32	0.2975	0.0225
		246.00	0.02	0.34	0.3096	0.0304
		251.00	0.02	0.36	0.3199	0.0401
		265.00	0.02	0.38	0.3494	0.0306
		280.00	0.02	0.40	0.3820	0.0180
<b>结论：样本数据是 正态 分布的 1% 显著性水平。</b>						

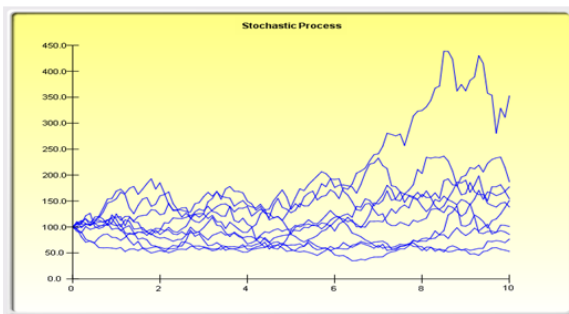
图 5.32 – 示例统计分析报告(正态检验)

随机过程

统计汇总

随机过程是由概率论的原则生成的一系列的事件或者路径。即，随机事件随着时间的延续而发生，但是出现的事件满足指定的统计理论和概率原则。主要的随机过程包括随机行走或者布朗运动，均值回复，以及条约扩散。这些过程可以用来预测遵循随机趋势但是却严格受概率论限制的变量的多个层面。生成随机事件的公式事先已知但是最后生成的结果却无法预知。

随机行走的布朗运动过程可以用来预测股票价格，商品价格，以及已知漂移或者增长，围绕漂移路径的波动性的其他随机时间序列数据的行为。均值回复行为通过锁定路径的远期值用来减少随机行走的波动性，使之可以用于预测时间序列变量的长期比率例如利率和通胀率（这些都是远期目标值都被权威或者市场调整）。跳跃扩散过程可以用来预测时间序列数据，当变量表现出随机跳跃，例如石油价格或者电力价格（离散的外生事件冲击可能使价格高升或者下降）。最终，这三种随机过程可以混合出现。



统计汇总

下面是在给定数据的条件下估计随机过程的参数。这可由个人决定该是否应该使用随机过程来拟合（类似于拟合优度的计算），如果是的，这个随机过程是一个随机游走，均值回复，还是跳跃扩散的模型，或者是一个三者混合在一起的模型。需要根据以往的经验和经济和金融上的先验估计，选择正确的随机模型来表现数据的特征。这些估计参数可以用于随机过程的预测（[仿真/预测随机过程](#)）。

按年计算		回复比率	263.89%	跳跃率	20.41%
漂移率	-1.46%	长期值	327.72	跳跃大小	237.89
波动性	88.84%				
随机模型的概率拟合			46.48%		

图 5.33 –示例统计分析报告(随机参数估计)

### 5.10 分布分析工具

这是 Risk Simulator 软件中的统计概率密度工具，在进行一系列的设定之后会十分的有用，可以用来计算概率密度函数 (PDF)，和离散数据的概率密度函数 (PMF)，两者可以交替使用。再给出了某个分布的参数，我们就可以决定某些事件  $x$  发生的概率水平。此外，累积的概率密度函数 (CDF) 也可以计算出来，它就是对发生事件  $x$  的 PDF 的加总。最后，逆累积概率密度函数 (ICDF) 被用来计算在给定概率发生水平时的  $x$  值。

可以通过点击 **Risk Simulator | 工具 | 分布分析**。作为示例，图 5.34 显示了二项分布（例如，一个事件只有两种可能性水平，例如掷钱币，只有头像和背面两种可能性，这里的事件可以是头像也可以是背面，事先定义好头像出现的可能性水平）。假设投掷钱币两次，设定头像出现为成功，使用二项分布，试验次数为 2（投掷两次），概率水平=0.50（成功的概率或者头像出现的概率）。选择 PDF，设定  $x$  的范围，从 0 到 2 步长为 1，（这意味着将 0, 1, 2 作为  $x$  值），结果以图像和表格的形式输出，也包括理论分布四矩。出现的结果可能是头像-头像，背面-背面，头像-背面，背面-头像，因此头像不出现的概率为 25%，一个头像出现的概率为 50%，两个头像出现的概率为 25%。



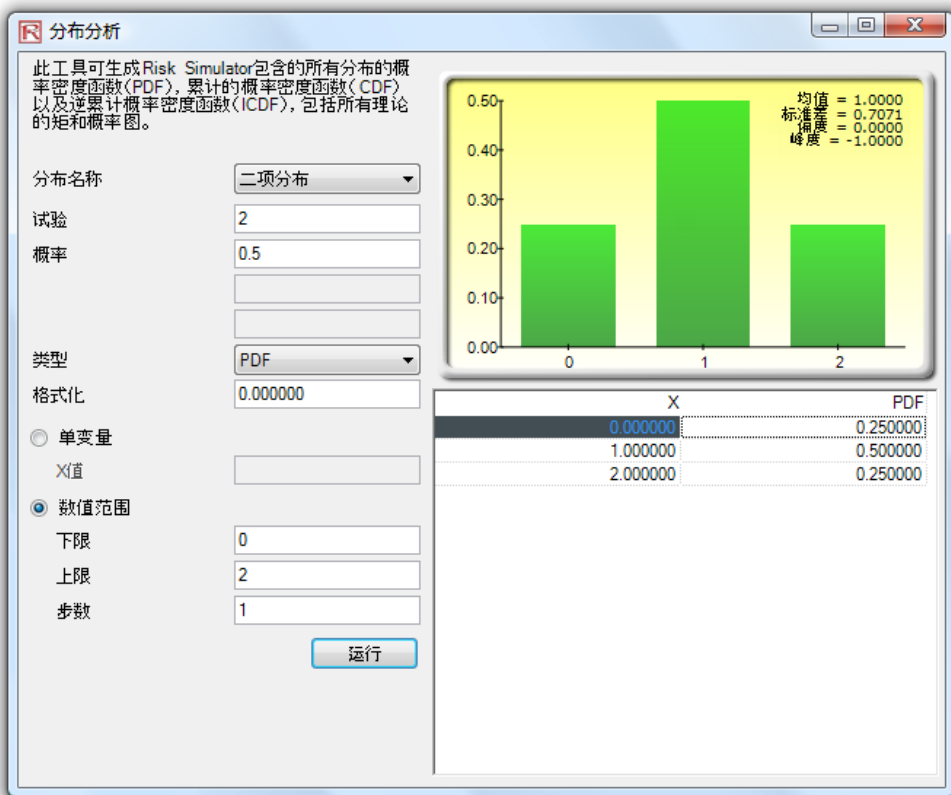


图 5.34 – 示例统计分析报告(2 次试验的二项分布)

同样地，我们可以得到投掷钱币的概率，以 20 次为例，如图 5.35。结果同样以表格和图形的形式输出。

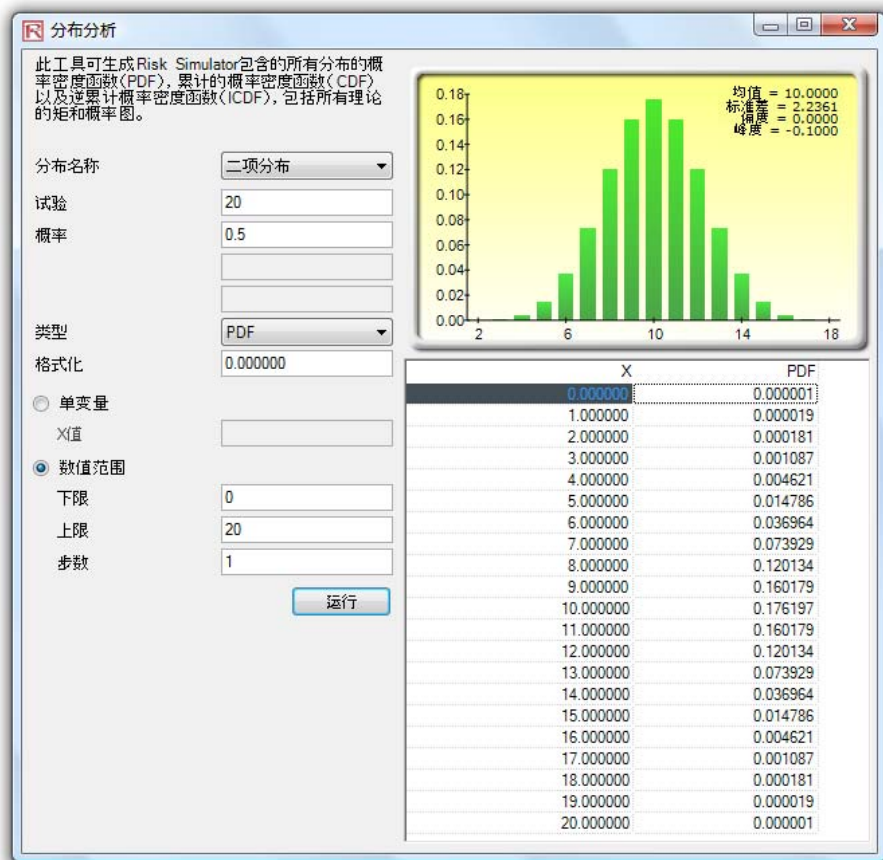


图 5.35 -示例统计分析报告(20 次试验的二项分布)

图 5.36 显示了二项分布以及如何计算 CDF 的。CDF 就是对每个  $x$  点对应的 PDF 进行加总。例如，在图 5.35，我们看见对于试验 0, 1, 2 对应的概率为 0.000001, 0.000019, 和 0.000181, 总和为 0.000201, 也就是  $x = 2$  时的 CDF 值，如图 5.36。PDF 计算了出现两次头像的概率，CDF 计算出不超过两次头像的概率（或者出现 0, 1, 2 次头像的概率）。相减之后（例如， $1 - 0.00021$  得到 0.999799 或者 99.9799%）提供了出现 3 次或者超过 3 次头像的概率。

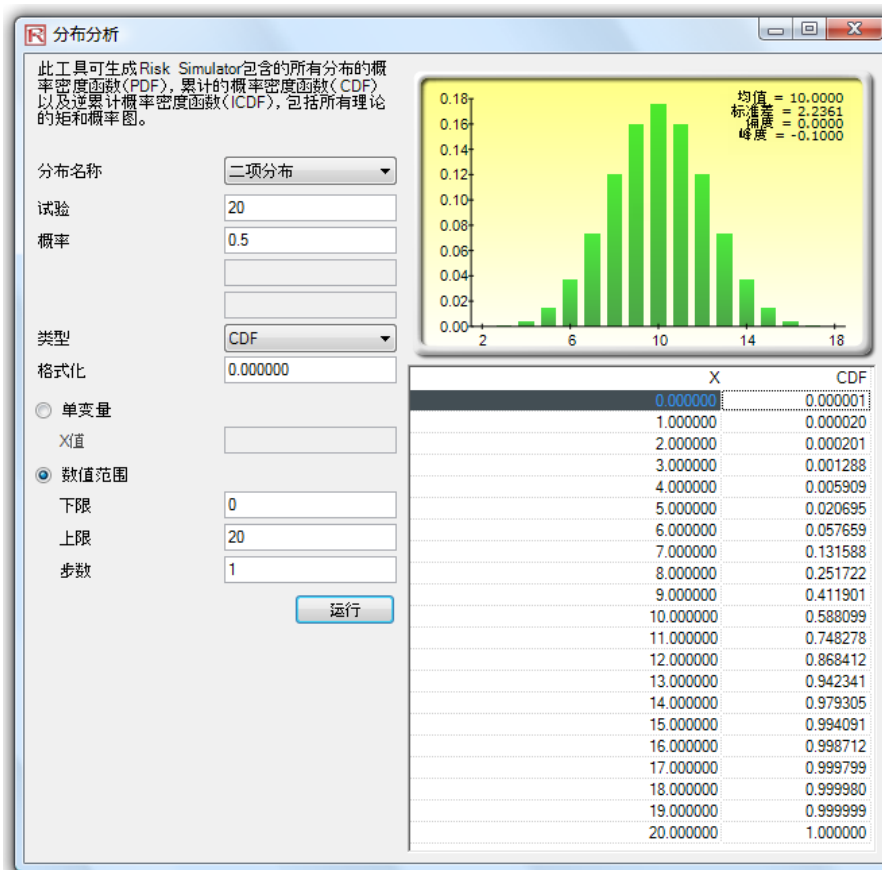


图 5.36 – 示例统计分析报告(20 次试验二项分布的 CDF)

使用分布分析工具，可以对一些高级的分布进行分析，例如 gamma 分布，beta 分布，逆二项分布，等其他一些包含在 Risk Simulator 中的分布。下面是对于连续型概率分布在该工具应用的示例，图 5.3.7 显示了标准正态分布（均值为 0，标准差为 1 的正态分布），这里使用 ICDF 发现对于累积的概率 97.5%（CDF）对应的  $x$  值。即，单尾 97.5% 的 CDF 等于双尾的 95% 的置信区间（即 2.5% 概率水平在左尾，2.5% 的概率水平在右尾，中间或者置信域的水平为 95%，就等于单尾 97.5% 的区域）。结果类似于  $Z$  值的 1.96。因此，使用分布分析工具，可以得到标准的分布值，具体的和累积的概率值，十分地方便和简单。

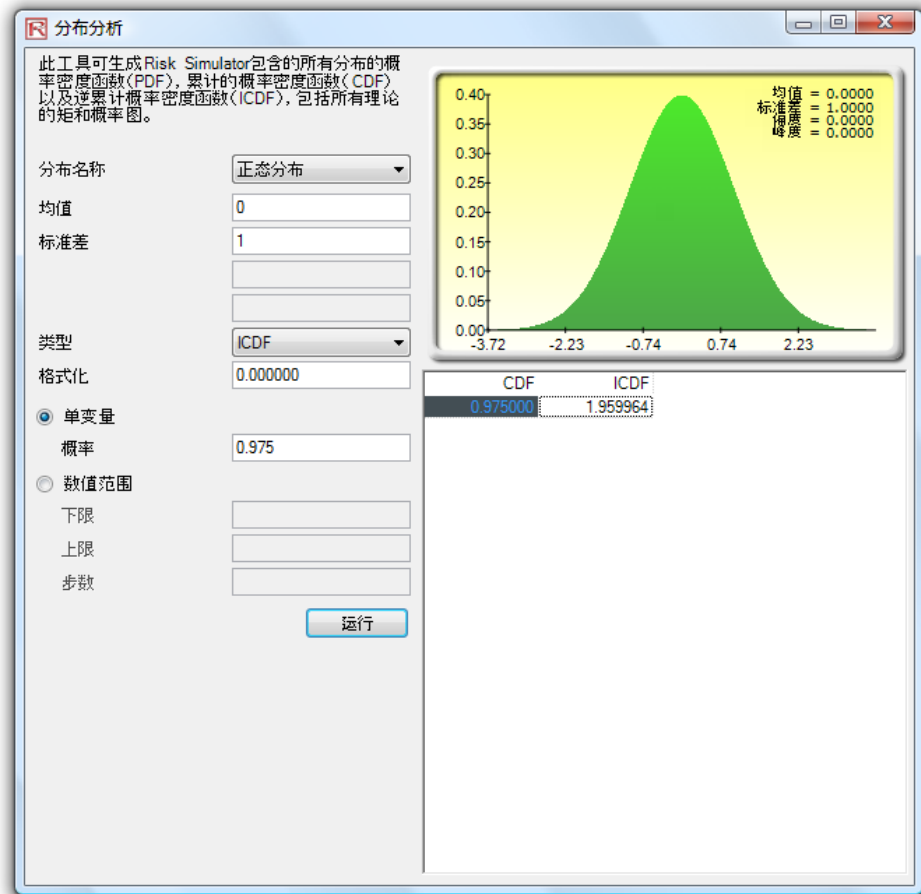


图 5.37 – 示例统计分析报告(正态分布的 ICDF 和 Z 值)

## 5.13 Risk Simulator 2011/2012 新工具

### 5.14 随机数生成，蒙特卡罗与拉丁超立方，以

#### 及相关性 Copula 方法

启动 Risk Simulator 2011/2012 版,有 6 种随机数生成器,3 种相关性 Copula 方法,及 2 种仿真抽样方法供以选择(图 5.41)。这些选项可以通过打开位置 [Risk Simulator/选项](#) 进行设置。

随机数生成器 (RNG) 是任何仿真软件的核心。基于产生的随机数,可以构造不同的数学分布。本软件默认使用方法是专有的 ROV 风险仿真方法,它提供了最好的和最可靠的随机数。共有 6 种随机数发生器提供支持,在一般情况下,ROV Risk Simulator 的默认方法和高级减法随机打乱法是推荐使用的两种方法。除非你的模型或分析方法有特定的要求,不需要应用其它的方法,即使如此,我们仍然建议针对这两种方法对结果进行测试。越向 RNG 方法清单的下面,算法越简单,运行的速度越快,相反,越向 RNG 方法清单的上面,结果越可靠。

在相关性部分,支持三种方法:正态 Copula, T-Copula, 和准伪正态 Copula。这些方法依赖数学积分技术,当你有疑问的时候,正态 copula 方法提供最安全和最保守的结果, T - Copula 方法提供仿真分布尾部的极值,而伪正态 copula 方法返回这些极值之间的结果。

在仿真方法部分,支持蒙特卡罗仿真 (MCS) 和拉丁超立方抽样 (LHS) 方法。请注意, Copula 方法和其它多元函数与 LHS 是不兼容的。这是因为 LHS 可以应用到一个单变量而不是一个联合分布。实际上,一个模型的分布越多, LHS 对模型输出结果的精度只有非常有限的影响,因为只是分别应用到每个分布。如果用户在开始的时候没有确定样本的数量, LHS 的优势会受到削弱,例如,当用户在仿真的过程中停止运行仿真的时候。LHS 也会对一个有着大量输入的模型有沉重的负担,因为在一个分布中运行第一个样本之前,它需要在每个分布生成和组织样本。这在运行一个较大的模型的时候产生一个较长时间的滞后,但只提供非常有限的额外准确性。最后, LHS 最好应用在当分布表现良好,对称并且没有任何相关性的时候。然而, LHS 是一个功能强大的方法,可以产生一个抽样均匀的分佈,当应用 LHS 的时候,与一个抽样更加均匀的分佈相比(分佈的每一个部分都进行抽样), MCS 有时候可以产生块状分佈(抽样数据有时可能产生在一个集中的区域)。

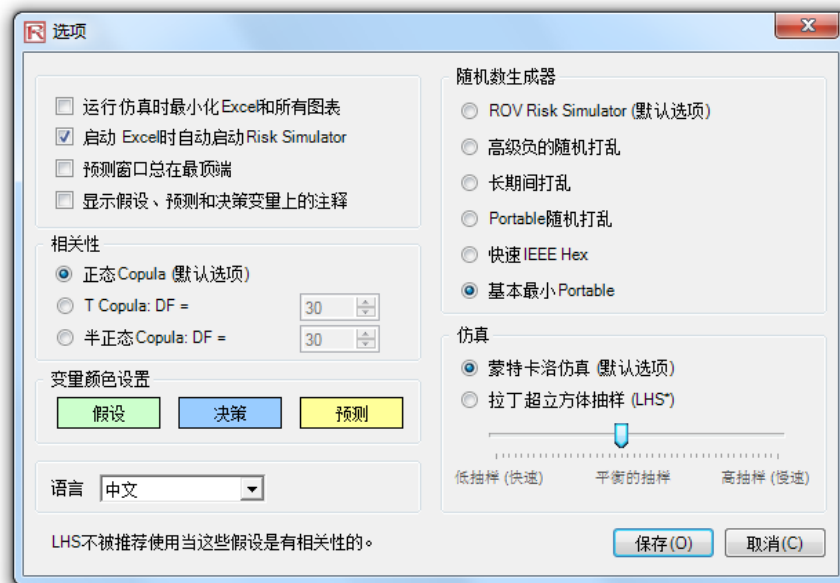


图 5.41 – Risk Simulator 选项

## 5.15 数据季节性因素和趋势性因素消除

此工具能够对你的原始数据进行季节性因素和趋势性因素消除处理，意消除季节性和趋势性的部分（图 5.42）。在预测模型中，该过程通常用于消除包括来自于季节性和趋势性的累积影响，仅显示数值的绝对变化，在消除一系列时间序列数据一般的漂移，趋势，扭曲，弯曲和季节性周期的影响之后确定潜在的循环模式。例如，从一个消除趋势性之后的数据集可以更加准确地了解到一个公司的销售情况，通过将整个数据集从一个斜坡转移到一个平面上，可以更好地理解潜在的循环和波动性。

很多时间序列数据呈现出季节性特点，其中某些事件在一段时间或者季节性周期之后会重现（滑雪胜地在冬天的收入比夏天的收入高，并且这个可预见的事件在每年冬季会重现）。季节性周期代表在循环自身重复之前需要经过多少个期间（例如，一天 24 小时，1 年 12 个月，一年 4 个季度，1 小时 60 分钟，等等）。此工具可以将你原来的数据消除周期性影响。一个季节性指数大于 1 表明是季节循环的一个高峰值，同样一个季节性指数小于 1 表明是季节循环的一个低峰值。

操作流程(节性因素和趋势性因素消除):

- 选择您要分析的数据（例如，B9: B28）然后点击 Risk Simulator|工具|消除数据季节因素和消除趋势因素
- 选择消除数据季节因素和/或消除数据趋势因素，选择你希望运行的任何消除趋势模型，并输入相关的处理选项（例如，多项式，阶移动平均线，差分，比率），然后点击确定
- 检查这两份详细报告，了解提供的方法，应用程序，以及由此产生图表和消除季节性因素/消除趋势性因素的数据

操作流程(节性因素和趋势性因素消除):

- 选择您要分析的数据（例如，B9：B28）然后单击 Risk Simulator|工具|数据季节性因素检验
- 输入进行测试的最大季节性期间。也就是说，如果你输入 6，该工具将测试以下的季节性时期：1，2，3，4，5，6。时期 1 当然意味着数据没有季节性
- 检查这两份详细报告，了解提供的方法，应用程序，以及由此产生图表和季节性测试结果。最好的季节性周期排列在第 1 位（通过最低的 RMSE 错误测量结果来进行排序）和包含所有的相关错误测量以进行比较：均方根误差（RMSE），均方误差（MSE），平均绝对差（MAD）的，及平均绝对百分比误差（MAPE）

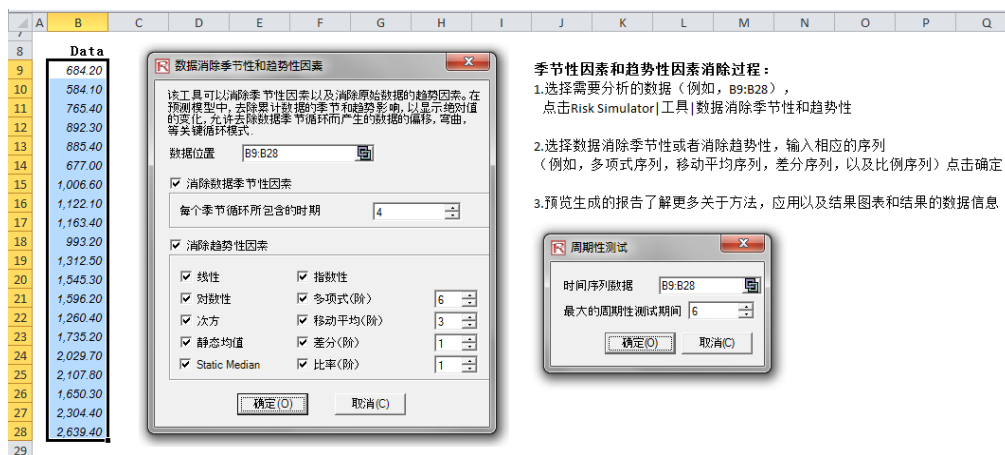


图 5.42 – 消除数据季节因素和消除趋势因素

### 5.16 主成分分析

主成分分析是一种识别数据模式的方式，并且重新塑造数据以突出其共性和差异（图 5.43）。当多个变量存在的时候，数据模式不容易被辨认。一旦发现了数据的模式以后，可以压缩数据使其维度减少。数据维度的减少并不意味着信息量的减少。相反，通过较少的变量，也可以获取相同水平的信息量。

操作程序:

- 选择数据进行分析（例如，B11：K30），单击 Risk Simulator|工具|主成分分析和单击确定
- 检查生成的报告获取计算结果

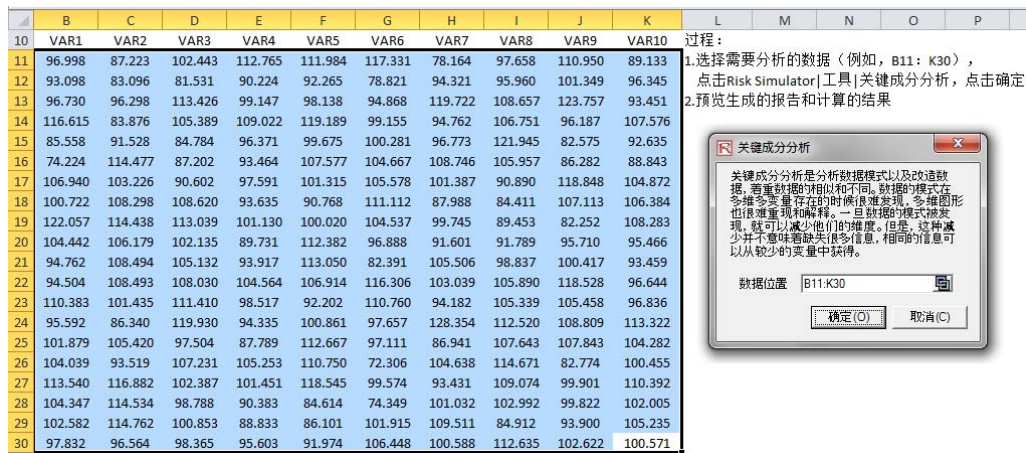


图 5.43 –主成分分析

### 5.17 结构性中断分析

结构性中断分析可以检验不同数据组的系数是否相同，在时间序列分析中常用于检验数据是否存在结构性突变（图 5.44）。时间序列数据集可以分为两个子集，然后可以对每个子集分别进行检验，统计学上确定数据在某个特定的时期是否开始突变。结构性突变检验经常被用来确定独立变量是否对数据总体的子集有不同的影响，进行的测试包括例如一个新的营销计划，活动，重大事件，收购，剥离等等对时间序列数据是否有影响。假设数据集有 100 个时间序列数据点，您可以设置各种断点进行检验，例如，数据点 10, 30 和 51（这意味着有三个结构性中断检验将在以下数据集进行：数据点 1-9 和 10-100；数据点与 1-29 和 30-100；以及 1-50 和 51-100，确实在第 10 个点，第 30 个点和第 51 个点开始的数据是否有结构性中断）。一个单尾假设检验对零假设 ( $H_0$ ) 进行检验，检验两个数据集之间在统计上是否显著相同。可替换假设 ( $H_a$ ) 就是两个数据集之间在统计上显著不同，这表明一个可能的结构性中断。如果计算的 P 值小于或者等于 0.01, 0.05 或者 0.1, 这意味着假设被拒绝，这意味着在 1%, 5%和 10%的显著性水平，这两个数据子集统计上显著差异，较高 P 值表示没有显著的结构性突变。

操作流程：

- 选择您要分析的数据（例如，B15:D34）和点击 **Risk Simulator|工具|结构性中断检验**，然后输入你想要进行检验的点（例如，6, 10, 12），然后点击 **确定**。
- 检查报告以确定你的数据上哪些测试点在统计上存在显著突变和哪些点不存在显著突变



	A	B	C	D	E	F	G	H	I	J
13										
14		Y	X1	X2						
15		521	18308	185						
16		367	1148	600						
17		443	18068	372						
18		365	7729	142						
19		614	100484	432						
20		385	16728	290						
21		286	14630	346						
22		397	4008	328						
23		764	38927	354						
24		427	22322	266						
25		153	3711	320						
26		231	3136	197						
27		524	50508	266						
28		328	28886	173						
29		240	16996	190						
30		286	13035	239						
31		285	12973	190						
32		569	16309	241						
33		96	5227	189						
34		498	19235	358						

**过程：**

- 1.选择需要分析的数据（例如，B15:D34）点击 Risk Simulator|工具|结构性终端检验，然后输入希望应用于数据的检验点（例如，6,10,12）点击确定
- 2.预览报告决定这些检验中断点是否统计显著

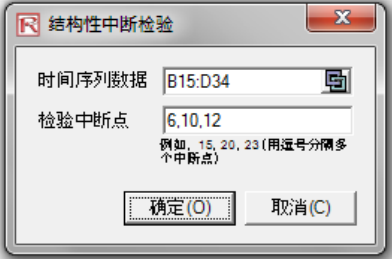


图 5.44 – 结构性中断分析

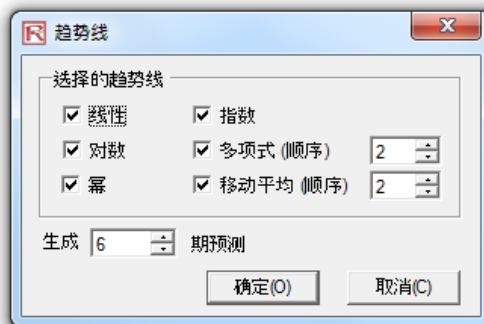
## 5.18 趋势线预测

趋势线可以用来确定一个时间序列数据集是否遵循任何明显的趋势（图 5.45）。趋势可以是线性或非线性（例如指数，对数，移动平均，电力，多项式，或幂）。

操作程序：

- 选择您要分析的数据然后点击 **Risk Simulator|预测|趋势线**，选择你想应用于数据的相关趋势线，输入预测的期数（例如，6 期）然后点击确定。
- 检查报告确定哪些检验趋势线针对你的数据提供了最优的拟合和最优的预测

年份	季度	时间点	销售额
2006	1	1	\$684.20
2006	2	2	\$584.10
2006	3	3	\$765.40
2006	4	4	\$892.30
2007	1	5	\$885.40
2007	2	6	\$677.00
2007	3	7	\$1,006.60
2007	4	8	\$1,122.10
2008	1	9	\$1,163.40
2008	2	10	\$993.20
2008	3	11	\$1,312.50
2008	4	12	\$1,545.30
2009	1	13	\$1,596.20
2009	2	14	\$1,260.40
2009	3	15	\$1,735.20
2009	4	16	\$2,029.70
2010	1	17	\$2,107.80
2010	2	18	\$1,650.30
2010	3	19	\$2,304.40
2010	4	20	\$2,639.40



图标 5.45 – 趋势线预测

## 5.19 模型检验工具

在模型创建好并且假设和预测设置好以后，你可以运行仿真或者运行模型检验工具（图 5.46）来检验该模型是否被正确设置。另外，如果模型不能运行或者你怀疑模型的一些设置可能不正确，可以从 [Risk Simulator/工具/模型检验](#) 来确定你的模型可能存在的问题。请注意，这个工具检验的是最常见的模型问题以及在 Risk Simulator 假设和预测中的问题，当然不能全面测试所有类型的问题，这仍然取决于模型的开发者来确保模型是否能够正常运行。

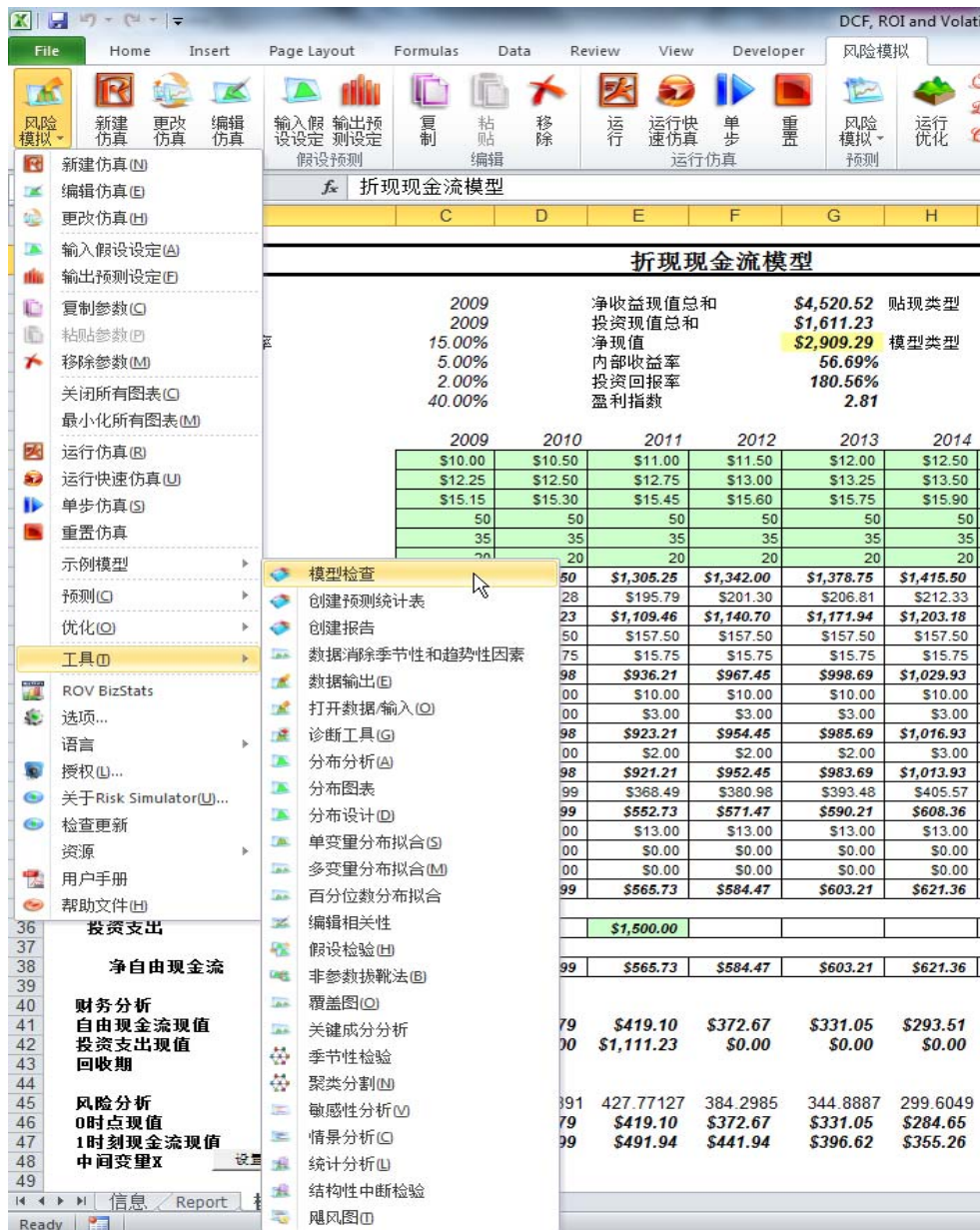


图 5.46 – 模型检验工具

## 5.20 百分位分布拟合工具

百分位分布拟合工具（图 5.47）是另外一种可替代的拟合概率分布的方法。有几个与此相关的分析工具，每一种工具都有自己的用途和优势：

- 分布拟合(百分位) - 用另一种输入的方法(百分位和第一/第二矩组合)，你可以找到一个特定分布的最优拟合参数而不需要原始数据。当没有足够数据的时候，这种方法也适合使用，只有当百分位和矩可以获得的时候，或者当只有两个或者三个数据点但是分布类型是已知或者被假定的时候，这种方法可以作为一种方式来恢复整个分布

- 分布拟合（单变量） - 利用统计方法来拟合你的原始数据到所有 42 种分布，然后进行比较以找出最佳拟合及其输入参数。好的拟合需要多个数据点，分布的类型可以不需要提前知道
- 分布拟合（多变量） - 使用与单变量拟合相同的算法，使用统计方法对多个变量同时拟合你的原始数据，但是可以在变量之间整合一个相关性矩阵。好的拟合需要多个数据点，分布的类型可以不需要提前知道
- 定制分布（设置假设） - 使用现存的原始数据，利用非参数重采样技术，以生成一个定制分布，及基于这个经验分布来仿真分布。这种方法需要较少的数据，分布类型不需要提前知道

操作程序：

- 点击 **Risk Simulator|工具|分布拟合（百分位）**，选择概率分布和你希望使用的输入类型，输入参数和点击**运行**来获取结果。检查拟合的 R 方结果及比较实证的和理论的拟合结果来确定你的拟合是否拟合得较好。

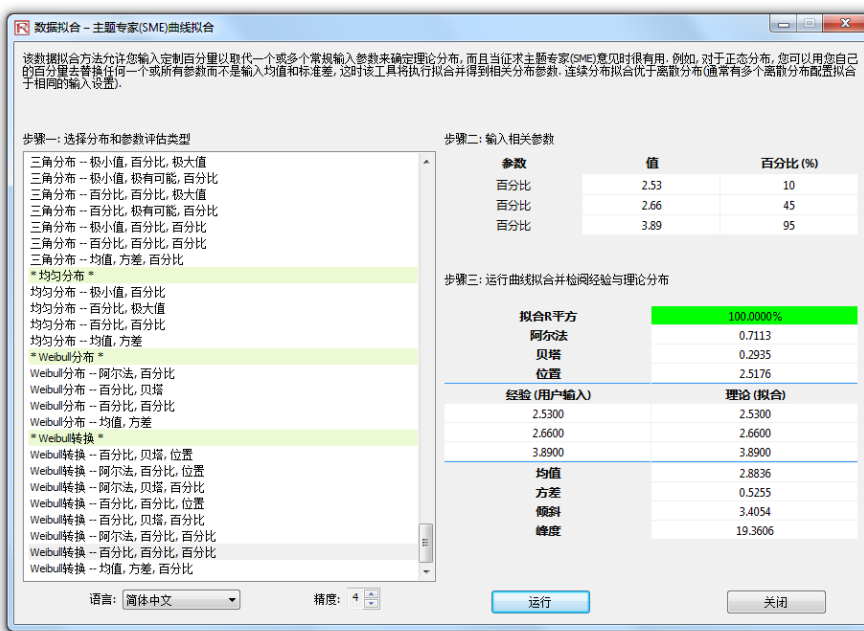


图 5.47 – 百分位分布拟合工具

## 5.21 分布图表：概率分布工具

新的概率分布工具非常地强大和快捷，可用于生成分布图表（图 5.48-5.51）。注意 Risk Simulator 软件中包含三种类似的工具，但是彼此的功能各不相同：

- 分布分析—用于快速的计算 Risk Simulator 中 42 种分布的 PDF, CDF, 和 ICDF，并且返回这些值的概率表。
- 分布图表—概率分布工具在这里用来比较相同分布的不同参数（例如，具有不同 Alpha, Beta 参数[2,2], [3,5], [3.5,8]的 Weibull 分布的 PDF, CDF, ICDF 的形状，并且将彼此重叠到一起）。
- 重叠图—用于比较不同的分布（理论输入假设和试验仿真输出预测）并且将彼此重叠比较。

过程:

在 *Risk Simulator*|分布图表中运行 ROV BizStats，点击 *应用全局输入* 按钮加载一系列输入参数或者输入相应的输入，点击 *运行* 计算结果。相应的 45 个分布的四个矩和 CDF，ICDF，PDF 结果都进行了计算（图 5.48）。



图 5.48—概率分布工具（45 种概率分布）

- 点击 *图表栏* (图 5.49)，选择分布[A](例如，正弦函数)，选择想要返回的 CDF，ICDF，或者 PDF[B]，输入相应的参数，*点击运行图或者表* [C]。就可以在图和表栏之间切换，查看运行的结果，或者点击图的图标[E]查看图的效果。
- 可以更改两个参数[H]输入到 *从到步* 生成多个图表或者使用 *自定义* 输入点击 *运行*。例如，图 5.50 展示的，运行 Beta 分布，选择 PDF[G]，选择 Alpha 和 Beta 改变[H]使用自定义输入[I]，输入相应的输入参数：2;5;5 用于 Alpha; 5;3;5 用于 Beta 分布[J]点击 *运行图表*。就可以生成三个 Beta 分布[K]：Beta (2,5)，Beta (5,3)，Beta (5,5) [L]。查看不同的图形类型，网格，语言和小数设置[M]，以及再次运行分布使用理论 vs 试验仿真的值 [N]。
- 图 5.51 显示了伯努利分布生成的概率表，使用 *从到步* 选项[O]，选择不同的试验成功概率和试验成功次数（随机变量 X）。重复计算，点击表栏 [P]查看创建的概率密度函数结果。在案例中，伯努利分布的输入起始点，试验=20，成功概率=0.5，成功的试验次数 X=10,显示的成功概率允许从 0,0.25,...,0.50 变动，如列变量所示，成功试验次数允许从 0,1,2,...,8，如行变量所示。选择 PDF，因此结果在表中展示。例如，进行 20 次试验，每次试验都 25%机会成功，因此 2 次成功的概率为 0.0669 或者 6.69%的概率值。

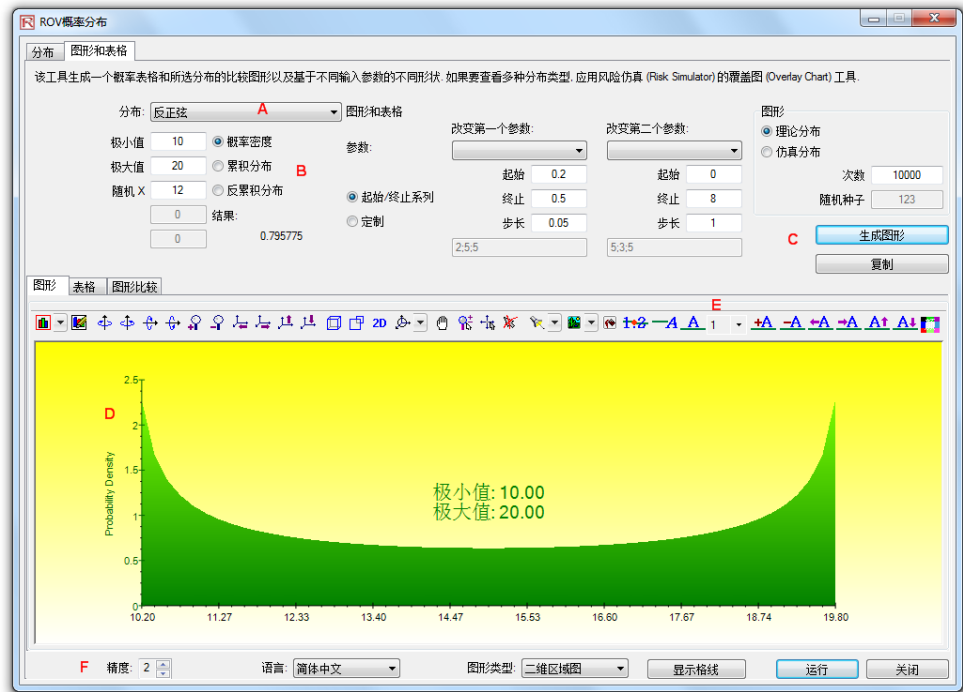


图 5.49—ROV 概率分布 (PDF 和 CDF 图)

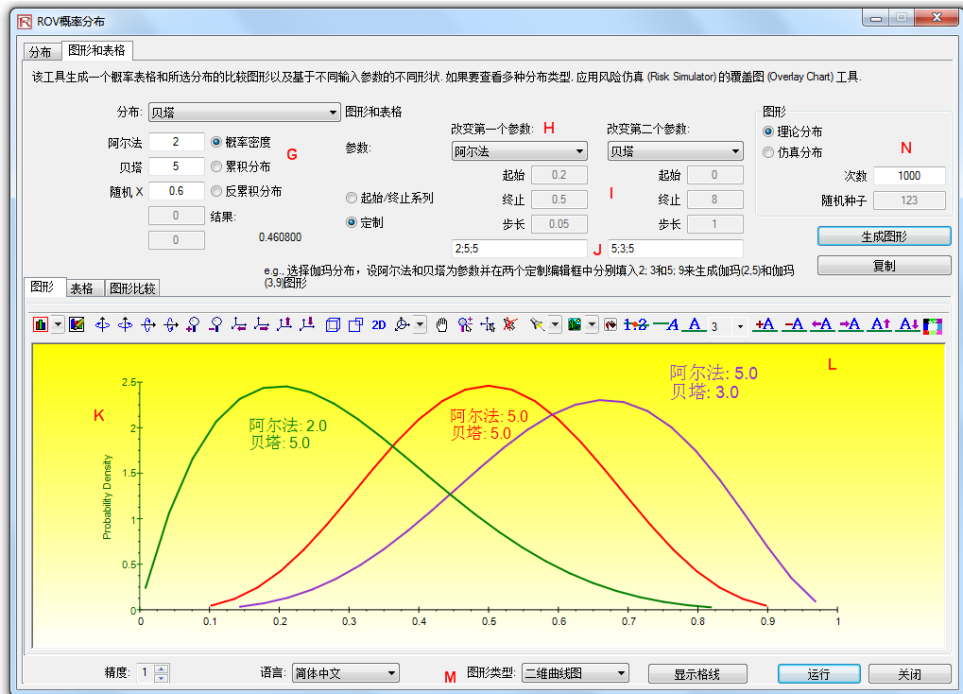


图 5.50—ROV 概率分布 (多重重叠图)

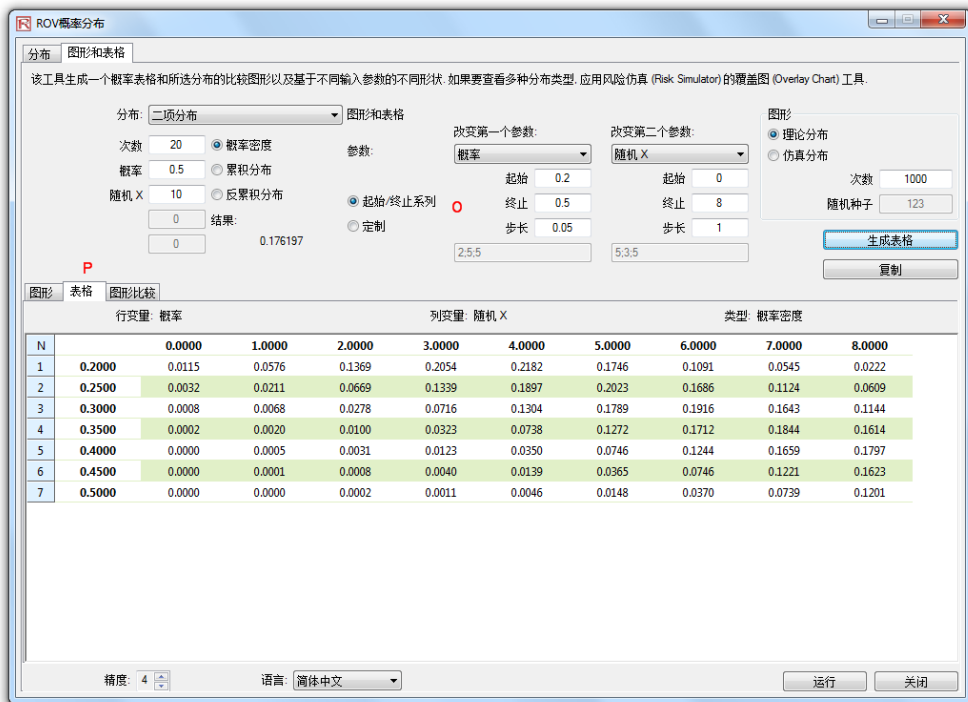


图 5.51—ROV 概率分布 (分布表)

## 5.22 ROV BizStats

最新的 ROV BizStats 工具是 Risk Simulator 中非常强大和便捷的模块，用于进行商业统计和数据分析模型，它包含 130 多种商业统计和分析模型（图 5.52-5.55）。下面提供了快速应用步骤和详细的软件信息。

### 过程:

- 在运行 **Risk Simulator / ROV BizStats** ROV BizStats，点击案例加载数据和模型文档[A]或者手动输入或者复制/粘贴数据到表格中 [D]（图 5.52）。也可以在第一个注释行中添加注释或者变量名称[C]。
- 选择相应的模型[F]在第 2 步运行，使用示例数据输入设置[G]，输入相应的变量[H]。相同参数的不同变量用分号隔开，不同参数的另起一行。
- 点击运行[I]计算结果[J]。可以在不同的栏目中查看任何相关的分析结果，图表或者统计结果。
- 如果需要，可以提供模型的名称在步骤 4[L]中保存为文档。多个模型可以使用同一个文档。现有的模型可以通过编辑或者删除[M]，重新排列配置[N]，所有的更改都保存在[O]一个单一文档中以后缀名.bizstats 结尾。

### 注意:

- 可以在菜单栏中设置数据表的大小，数据表支持高达 1000 个变量行，1 百万列的数据格。菜单也允许更改语言设置和数据的小数位。
- 开始，可以加载案例文件[A]包含完整的数据和预先创建好的模型[S]。可以双击任何一个模型在报告区得到结果[J],有的时候会是图表或者统计结果[T/U]。使用案例文件，可以了解输入参数[H]是根据模型[描述[G]]，可以创建任何自定义的模型。
- 点击变量的标头[D]一次选择单个或者多个变量,右键单击添加，删除，复制，粘贴或者可视化[P]选择的变量。
- 模型也是可以通过使用命令行控制台程序输入[V/W/X]。查看运行过程，双击运行模型[S]进入命令行控制台[V]。也可以重建模型或者新建模型，点击运行命令[X]完成模型。控制台下的每行反应了模型和相关的参数。
- 整个.bizstats 文档（当数据和多个模型已经创建和保存）可以直接通过文件菜单中的 XML 编辑器对 XML[Z]文件编辑。文档中的更改可以程序化地应用，当文件被保存的时候发生作用。

### 贴士

- 点击表格列的头选择整个表格或者变量，一旦选择，可以鼠标右键自动调整适合列的大小，剪切，复制，删除，或者粘贴数据。也可以点击然后选择列的头部选择多个变量，右键选择可视化数据。
- 如果单元格包含很大的数据而无法完全显示，点击选择，在单元格附近挪动鼠标，可以看到跳出的提示显示整个值，或者简单的调整变量列的大小（拖动数据列使之显示更宽，双击数据列的边缘自动调整大小，或者右键点击数据头部然后选择自动拟合）。
- 使用上，下，左，右键在表格内移动，或者使用 *Home*，或者 *End* 键跳到行的最左端和最右端。也可以使用组合键例如：*Ctrl+Home* 跳到左侧顶端，*Ctrl+End* 右侧底端，*Shift+Up/Down* 选择整个区域，等等。
- 可以在提示区域输入简短的提示。尽量使提示简短。



- 在可视化栏尝试点击多个按钮，改变试图的外观感（例如，翻转，移动，放大，等等）。
- 在第3步模型运行完之后，**复制**按钮可以用来复制**结果**，**图表**，以及**统计**栏，否则的话复制的内容都会是空。
- 报告按钮只会在第4步存在保存的模型或者表格中包含数据的时候生成报告，否则的话生成的报告也会是空的。运行数据输出和生成报告的时候需要预先安装 Microsoft Excel 和 PowerPoint。
- 在如何运行特定统计模型时可以查看示例文档查看如何在第1步设置数据或者如何在第2步输入参数。可以使用开始指南和模板存放数据和模型。
- 可以在语言菜单更改语言。目前软件包含10种语言，更多地都会在未来加入。可是，有时候某些结果还是以英语显示。
- 在第2步可以通过显示的下拉列表，更改模型列表的显示。可以按照首字母，内容，或者输入的要求—注意在某些字符语言（例如，中文，日文，韩语），没有首字母排列，所以这一项不可用。
- 软件可以处理小数点和数字的设定（例如，一千零五十美分，可以写作1,000.50 1.000,50 或者 1'000,50）。可以在 ROV BizStats 菜单 **数据/小数**设置选择小数设置。可是，请更改计算机的区域设置到 English USA，然后在 ROV BizStats 保留默认的北美 1,000.50（这个设置用来保证能够与 ROV BizStats 和默认示例兼容）

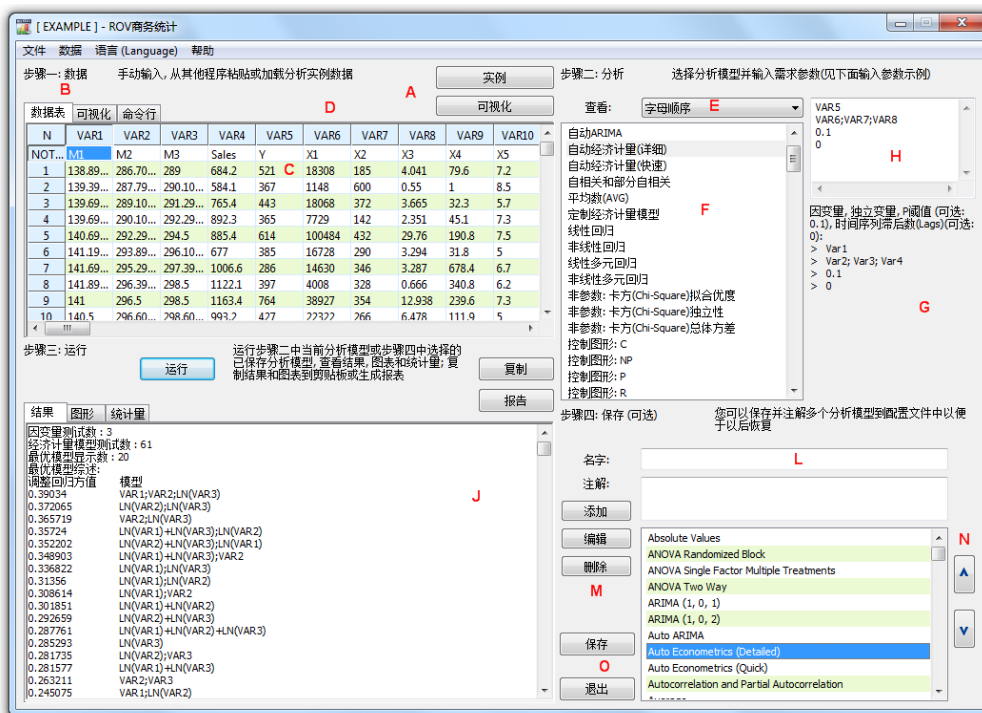


图 5.52—ROV BizStats（统计分析）

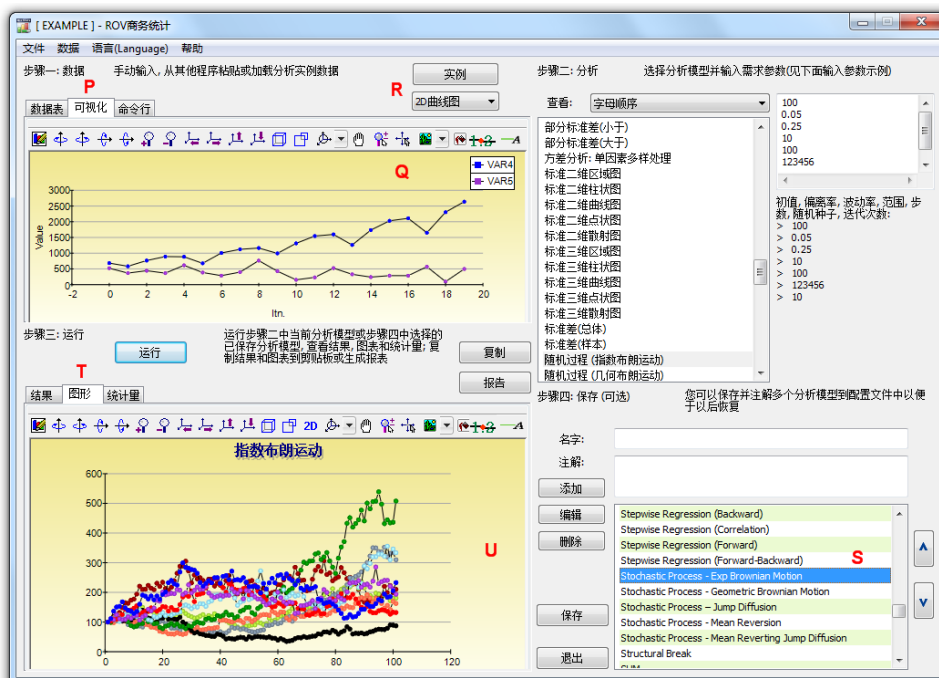


图 5.53—ROV Bizstats（数据可视化和结果图）

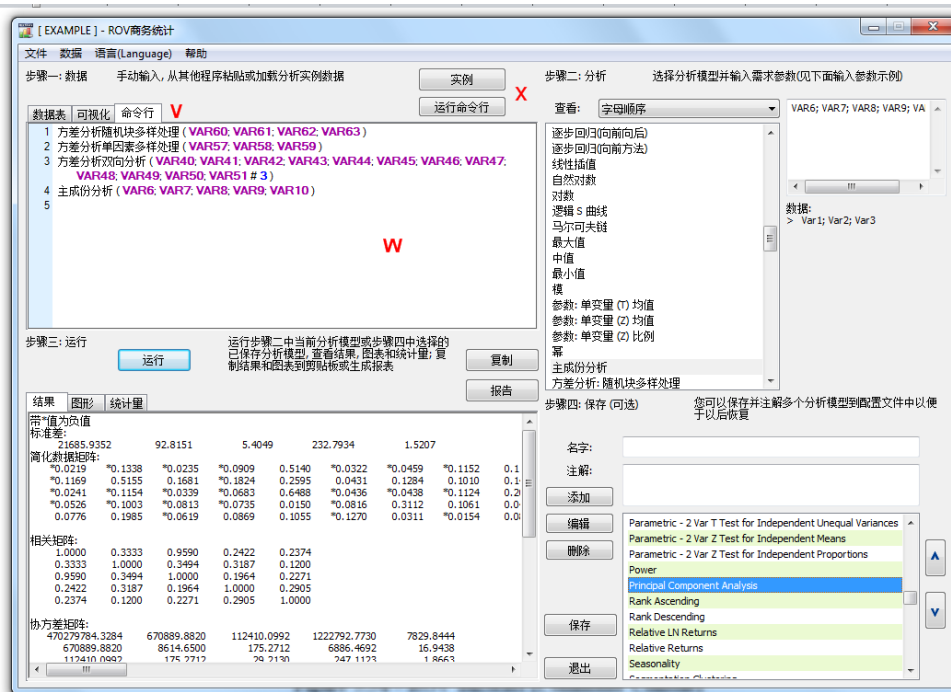


图 5.54—ROV BizStats（命令行控制台）

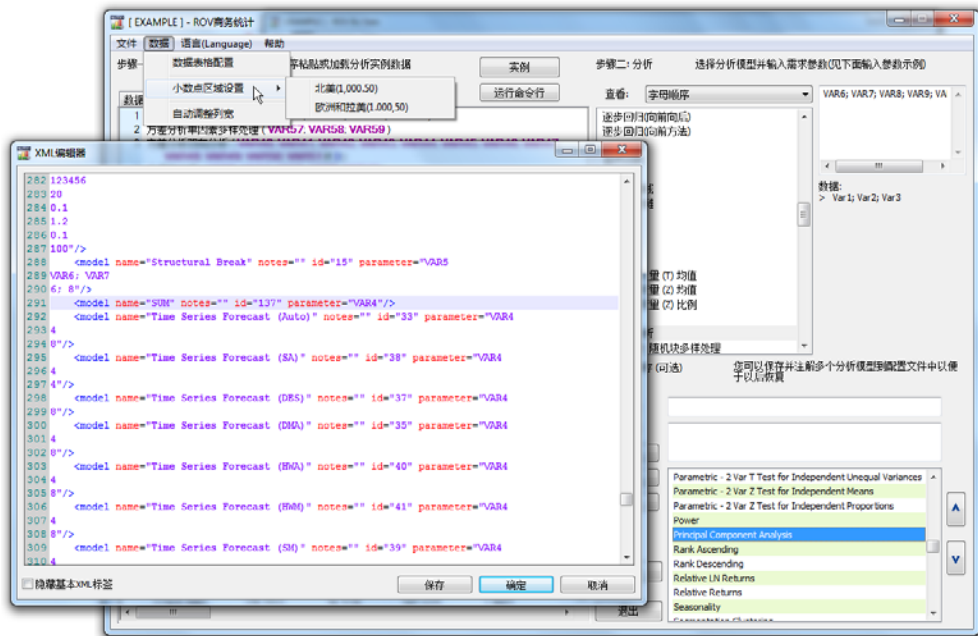


图 5.55—ROV BizStats (XML 编辑器)

### 5.23 神经网络和组合模糊逻辑预测方法

神经网络一词通常用来指一个网络或生物神经元电路，而这个名词的现代用法通常是指在一个软件环境中重建的由人工神经元或网络节点组成的人工神经网络。这种方法通过一种思考和辨认模式的方法试图模仿人类大脑或神经元。在这里，通过辨认模式来实现预测时间序列数据的目的。你可以在 Risk Simulator 里面的 **ROV BizStats** 模块找到该方法。位于 **ROV 的 BizStats |神经网络**，以及在 **Risk Simulator|预测|神经网络**。图 5.56 显示了神经网络预测方法。

#### 操作流程

- 点击 Risk Simulator | |预测|神经网络
- 通过手动输入数据或从剪贴板粘贴进行启动（例如，选择并复制 Excel 中的一些数据，启动这项功能，并通过点击 **粘贴** 按钮来粘贴数据）
- 选择你希望运行的 **线性** 或 **非线性** 神经网络模型，输入你期望的预测期间（如 5），隐藏层的神经网络数（如 3），和 **测试周期数**（如 5）。
- 点击 **运行** 以执行分析和审查计算的结果和图表。您还可以将结果 **复制** 到剪贴板和图表，并粘贴到另一个软件应用程序中。

请注意，在网络中隐藏层的数目是一个输入参数，并需要使用数据来校准。通常情况下，数据模式越复杂，你将需要更多隐藏层数，并且需要更多的时间来计算。建议你在 3 层开始。测试期数仅仅是用于神经网络模型所最后的校准所使用的数据点的数量，我们建议使用至少是你希望预测的期数。

与此相反，模糊逻辑这个词是从模糊集合理论而来，用于近似而不是精确的推理，与“清脆的逻辑”相反。其中二元集合有二元逻辑，模糊逻辑变量可能有一个其

范围介于 0 和 1 之间的真实数值，而不同于经典命题逻辑的两个真实数值。这种模糊加权特性与一个组合方法一起使用产生时间序列数据预测结果，如图 5.57 所示的结果，这种方法应用到具有季节性和趋势性的时间序列数据时是最适用的。你可以在 Risk Simulator 里面的 **ROV BizStats** 模块来找到这种方法。位于 **ROV BizStats | 组合模糊逻辑** 和 **Risk Simulator | 预测 | 组合模糊逻辑**。图 5.57 显示了模糊逻辑预测方法。

### 操作流程

- 点击 Risk Simulator | 预测 | 组合模糊逻辑。
- 通过手动输入数据或从剪贴板粘贴进行启动（例如，选择并复制 Excel 中的一些数据，启动这项功能，并通过点击 *粘贴* 按钮来粘贴数据）
- 从下拉列表中选择你希望运行分析的变量，输入季节性周期（例如，季度数据是 4，月度数据是 12，等等），以及 *预测期数*（如 5）。
- 点击 *运行* 以执行分析和审查计算的结果和图表。您还可以将结果 *复制* 到剪贴板和图表，并粘贴在另一个软件应用程序中。

请注意，神经网络和模糊逻辑技术都还没有被确定为商业预测领域有效和可靠的方法，无论是在战略，战术或运营层面。在这些高级预测领域，还需要进行大量持续的研究。然而，Risk Simulator 提供了这两种技术的基本方法来进行时间序列预测。我们建议你不要孤立地使用这些技术，而与其它的风险模拟预测方法相结合，建立更强大可靠的模型。

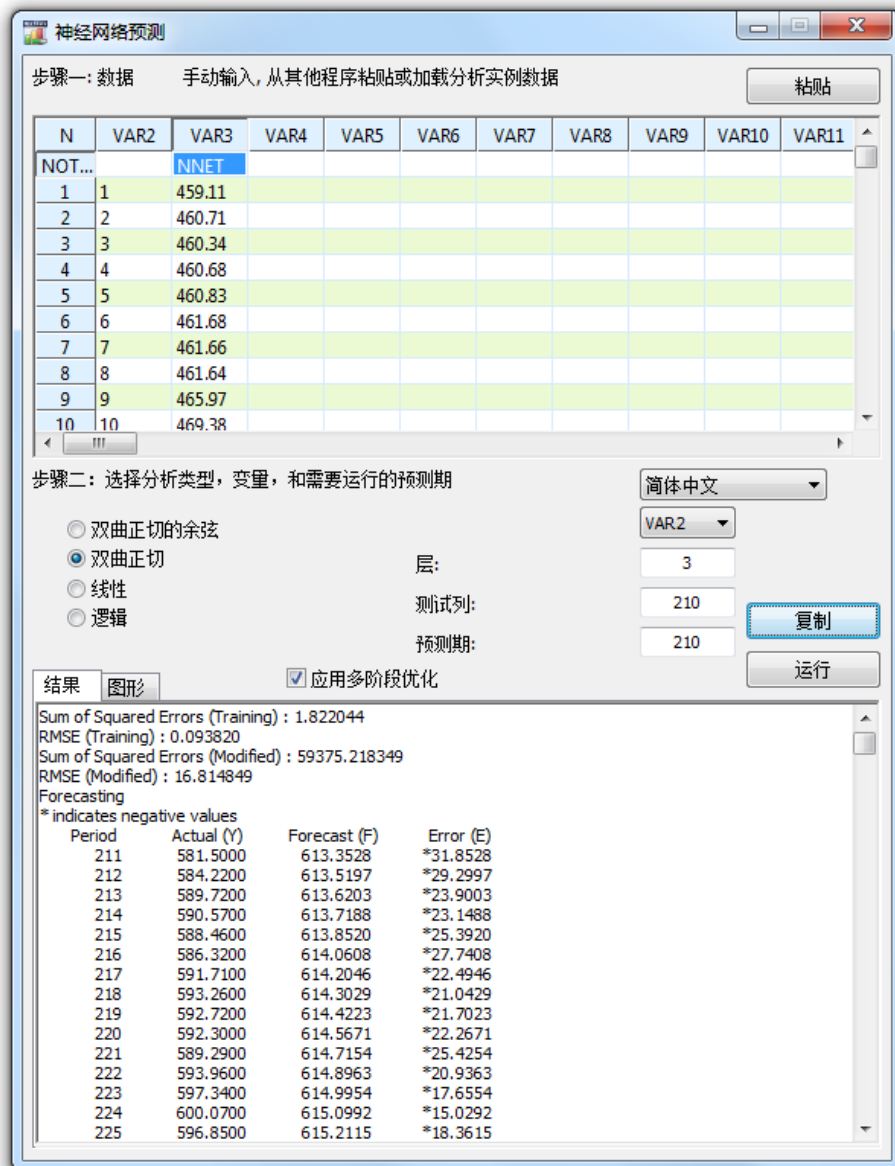


图 5.56 – 神经网络预测

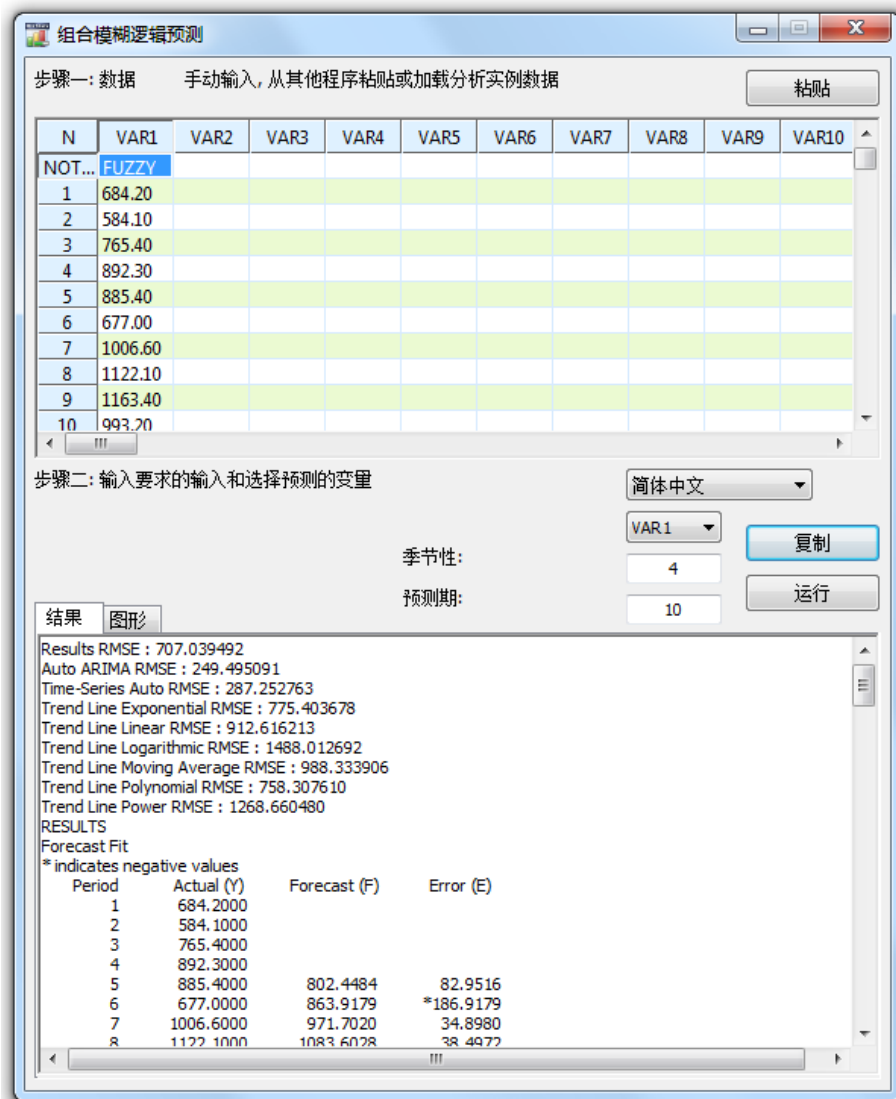


图 5.57 – 模糊逻辑时间序列预测

## 5.24 优化器目标搜寻

目标搜寻工具是一个搜索算法，用于模型来寻找了一个单变量的解。如果你知道一个公式或模型的结果，但不知道什么样的输入值才能得到这一结果，可以使用 **Risk Simulator|工具|目标搜寻** 功能。请注意，目标搜寻功能只适用于只有一个变量输入数值。如果你想接受多个输入值，使用 Risk Simulator 高级的优化程序。图 5.58 显示了一个简单的模型和如何运行目标搜寻功能。

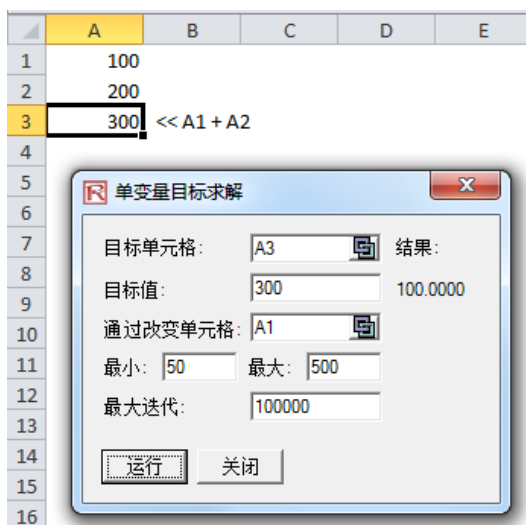


图 5.58 – 目标搜寻

## 5.25 单变量优化

单变量优化工具是一个搜索算法，应用于模型来寻找了一个单变量的解，就像前面讨论的目标搜寻方法。如果你想最大化或者最小化一个模型中可能的结果，但不知道公式需要什么样的输入值来得到这一结果，可以使用 **Risk Simulator|工具|单变量优化** 功能（图 5.59）。请注意，此单变量优化程序运行速度很快，但只适用于寻找只有一个变量的输入。如果你想接受多个输入值，使用 Risk Simulator 高级优化方法。请注意，此工具包括在 Risk Simulator 内，因为有时你可能需要一个快速的优化计算来解决单个决策变量问题。此工具提供了此功能而不需要设置一个仿真文档，仿真假设，决策变量，目标和约束的优化模型。

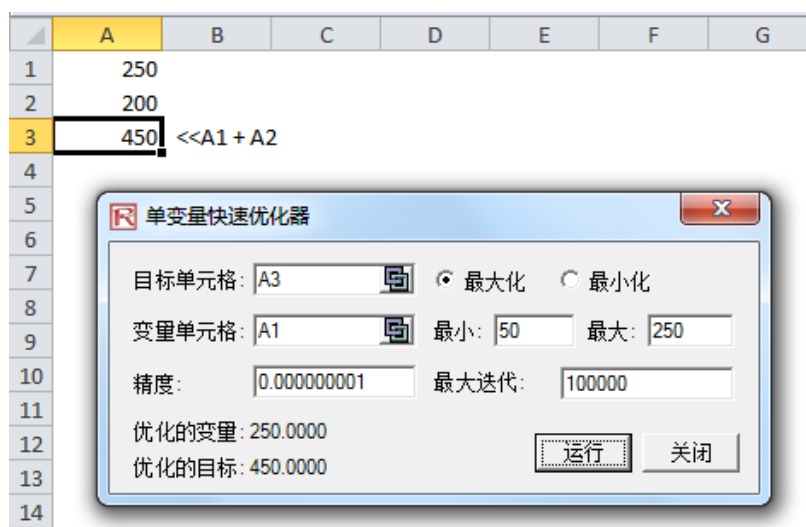


图 5.59 – 单变量优化

## 5.26 遗传算法优化

遗传算法是一种启发式模仿自然演化过程的方法。此方法一般用于获得优化和搜索问题的有用的解。遗传算法属于进化算法的较大级别类，进化算法为优化问题提供了解决方法，使用的技术包括自然进化，如继承，变异，选择和交叉。

遗传算法可在 **Risk Simulator|工具|遗传算法** 中找到（图 5.60）。应注意校正模型的输入因为其结果对输入是相当敏感（对于最常见的输入水平，默认输入可作为一般指导），并建议将梯度搜索测试作为一个更准确的结果集（可以取消选择该选项就可以开始，然后选择该选项，重新运行分析和比较结果）。

注：对于许多问题，遗传算法可能有一种向局部最优值收敛的倾向，甚至是任意点而不是要解决的问题的全局最优值。这意味着此方法不知道如何牺牲短期的拟合优度来获取长期的拟合优度。对于特定的优化问题，其它的优化算法可能提供比遗传算法更好的解（考虑到同样的计算时间）。因此，建议您首先运行遗传算法，然后重新运行，通过检查选择 **应用梯度搜索测试**（图 5.60），以检查模型的稳健性。这种梯度搜索测试将尝试运行传统遗传算法优化技术与遗传优化方法的组合，并返回最佳的解。最后，除非有特定的理论需要使用遗传算法，我们建议您使用 **Risk Simulator** 的优化模块以获取更稳健的结果，它使您可以运行更高级的基于风险的动态和随机优化方法。





图 5.60 – 遗传算法

## 5.27 ROV 决策树模组

### 5.27.1 决策树

ROV 决策树软件用于创建和评估决策树模型（图 5.61）。高级方法和分析包括：

- 决策树模型
- 蒙特拉罗仿真
- 敏感性分析
- 情景分析
- 贝叶斯定理(连接及后可能性更新)
- 期望数值信息
- 极大极小
- 极大极小
- 风险文件

以下是快速开始使用流程：

- 模组共有十一种语言，可以通过语言菜单改变。
- 选择任意节点，*插入选项点*或者*插入终止点*，点击选项节点图标(方形)或者终止节点图标(三角形)或使用功能*插入菜单*。
- 双击节点，修改单个*选项节点*或者*终止节点*。有时候当您点击一个节点，所有下属于节点也会被选择(这样可以移动从所选节点下的所有树)。如果您希望仅仅选择那个节点，您可以点击空白背景，单独选择节点。此外，您可以移动单个节点或者整个树(右键，或者在*编辑菜单*中，选择*单个移动节点*或者*一起移动节点*)。
- 以下是描述可以并配置到节点属性中。可以尝试属性的不同设置来判断战略树中的效果：
  - *名称*. 节点上方显示名称
  - *数值*. 节点下方显示数值
  - *Excel 连接*. 从 Excel 单元格中连接数据
  - *注释*. 可以在节点上方或下方插入节点
  - *模型展示*. 名称，数值，注释的组合
  - *本地颜色与全部颜色*. 节点颜色可以改变
  - *标签内形状*. 文本可以放置于节点中(需要将节点扩大些)
  - *子事件名*. 文本可以放置在引导节点提示事件。
  - *选择实物期权*. 实物期权模型可以指派到当前节点。指派节点可以生成一系列输入变量。
  - *全部元素*是可定制化的，包括战略树的背景，连接线，选择节点，终止节点，文本盒。具有以下设置项：
    - *字体* 名字，数值，注释，标签，项目名
    - *节点大小*(最小和最大的高度和宽度)
    - *边界*(线类型，宽度，颜色)
    - *阴影* (颜色和是否应用阴影)
    - *全部颜色*

- 全部形状
- *编辑* 菜单中的 *查找数据要求窗口*能够在战略树右侧打开一个窗口,当一个选择节点或者终止节点被选择时候,能够展示节点的属性,并直接更新。这个功能能够每次在节点上双击。
- *案例文件*能够通过 *文件*菜单开始构建决策树。
- *文件 保护文件*可以将战略树加密到 256 个字节密码。在加密的时候,请牢记密码。
- *文件 截屏*可以截屏现有模型。图片可以在其他软件中使用。
- 添加, 复制, 重命名, 删除决策树 可以通过右键决策树或者 *编辑* 菜单。
- 您也可以任意选项或者终止节点 *插入文件链接* 或者 *插入评论*, 在背景处 *插入文本* 或者 *插入图片*。
- 您也可以改变现有类型, 管理和创建自定义的战略树(包括尺寸, 形状, 颜色, 字体大小, 整个决策树的颜色)。
- *插入决定因素*, *插入不确定性*, *插入终止*, 选择已有节点, 点击决策节点图标(方形), 不确定性节点图标(圆形), 终止节点图标(三角形), 在 *插入* 菜单中使用这些功能。
- 双击节点来修改单个决策, 不确定性, 终止节点属性。以下是决策树模组中的一些独特项目, 可以定制化并在节点属性用户界面中配置。
- 决策节点: *定制佣金*或者 *自动计算*节点的值。默认为自动计算。当在完整决策树模型上点击 *运行时*, 决策树节点会随着结果更新。
- 不确定性节点: *事件名*, *可能性*, *设置仿真假设*。您可以在不确定分支创建之后, 添加可能性事件名, 可能性, 仿真假设。
- 终结点: 人工输入, *Excel 连接*, 设置仿真假设。最终事件结果可以手动输入或者连接到一个 Excel 单元格(例如, 如果有 Excel 模型计算结果, 可以将模型与 Excel 模型输出单元格连接), 或者设运行仿真来 设置可能性分布假设。
- *编辑 浏览节点属性窗口* 节点被选择时。节点属性会更新
- 决策树模组有以下高级功能:
  - 决策树下的蒙特拉罗模型
  - 得到次可能性的贝叶斯分析
  - 完全信息的预测值, 最大最小分析, 风险文件, 不完整信息的值
  - 敏感性分析
  - 情景分析
  - 效用函数分析

### 5.27.2 仿真模型

这个工具可以在决策树上运行蒙特卡洛分析 (图 5.62). 可以设置可能性分布作为输入假设来运行仿真。您可以为选择的节点设置假设, 也可以用新假设来设置假设, (或者使用过去创建的假设) 在一个数量等式或者公式中。例如, 您可以设置一个新假设叫做正态分布(平均值为 100, 标准差为 10), 在决策树中运行假设, 在等式(100\*正态+15.25)中使用假设。<L315>在一个数字表达盒中创建自己的模型。您也可以使用基本运算或者在等式中添加变量, 双击变量列表。新的变量可以添加进列表, 可以是数字表达式, 也可以是假设。

### 2.57.3 贝叶斯分析

<L371>贝叶斯定理分析工具(图 5.63)可以用于任意连在一条路径上的两个不确定事件, 在右边的案例中(图 5.63), 不确定的 A 和 B 连在一起, 事件 A 先发生, 事件 B 后发生。事件 A 是市场研究, 有两个输出(有利的或者不利的)。事件 B 是市场条件, 同样有两个输出 (强的和弱的)。工具用于计算连接, 临界, 通过可能性和依赖条件来确定贝叶斯后可能性, 依赖可能性可以通过两个后更新条件可能来计算。选择如下相关分析点击加载案例, 了解案例输入及哪些结果用于决策树中的输入。

#### 快速流程

- **第一步:** 输入第一个和第二个未确定事件的名字, 选择每个事件的可能性数量。
- **第二步:** 输入可能性事件的名称或输出
- **第三步:** 输入第二个事件的前期可能性, 每个事件或者输出的条件, 可能性的和为 100%

### 5.27.4 精确信息的预计数值, 最大最小分析, 风险文件, 不精确信息的数值

工具(图 5.64)计算精确信息(EVPI)的的预计数值, 最大最小分析, 风险文件, 不精确信息的数值。开始前, 输入决策分支的数量或决策(例如: 构建大型的, 中型的, 小型), 不确定事件的数量, 自然输出的状态 (例如., 好市场, 差市场), 在每种情景下的预测结果。

精确信息的预测数值 (EVPI), 假设您有预见性, 知道需要做什么(通过市场研究或其他方法来更好地识别概率结果), 计算信息中是否有增值(例如, 市场研究是否会增值) 来比较更多简单的概率状态预测。输入决策分支或者决策(例如:构建大型的, 中型的, 小型) 以及未知事件或者输出状态的数量 (例如好市场, 差市场), 在每种情景下的预测结果。

最大最小 (最小化最大遗憾) 以及最大最小(最大化最小结果) 是找到优化决策路径的两种方法。这两种方法并不常用, 但依然可以提供为决策制定流程提供参考。输入决策分支的数量或者存在的路径 (例如: 构建大型的, 中型的, 小型的), 或者不确定性的事件或者每个路径的状态 (例如: 好经济, 坏经济)。然后, 完成不同情景下结果表并计算最大及最小结果。您也可以点击登录案例来查找一个计算。

### 5.27.5 敏感性分析

敏感性分析 (图 5.65)用来决定输入对于不同决策路径数值的影响。首先, 选择一个决策节点进行分析, 然后从列表选择一个可能性事件来测试。当不同可能性事件具有同一可能性时, 能够分析出是独立还是同时发生。

敏感性图展示了在不同可能性程度下, 决策路径的价值。结果表中展示了数值。交叉线的位置, 代表了概率事件一个决策路径变成主导。

### 5.27.6 情景表

情景表 (图 5.66) 能够在输入变动时, 生成决定输出数值。您可以选择一个或多个决策路径来分析(每个选择路径的结果将表示成一个单独的表或者图), 一二个不确定或者终止节点作为情景表的输入变量。

#### 快速流程

- 从下列表中选择一个或多个决策路径来分析
- 选择一个或者两个不确定事件或终止结果来模型
- 决定是否改变事情的可能性, 在同一可能性事件下
- 输入情景区间

### 5.27.7 生成效用函数

效用函数(图 5.67), 或者  $U(x)$ , 用于代替决策树中预期终止结果的预测值。 $U(x)$  可以有两种方法产生: 使用繁琐和详细的实验的每一个可能的结果或指数外推方法 (用于这里)。他们可以被模拟为一个风险厌恶的决策者 (负面影响比潜在下降更具有灾难性), 风险中性 (两面有平等的吸引力), 或风险喜好 (更希望具有上涨潜力)。输入最大或最小值的终端结果, 一些数据点之间计算完整曲线表。

如果你有一个 50 :50 的赌博, 你可以赚  $X$  美元或失去  $X/2$  美元, 或者不玩得得到 0 美元的报酬,  $X$  应该是什么? 举个例子, 如果打赌, 你或者可以赢得 100 美元或失去 50 美元, 或者完全不玩, 那么  $X$  就是 100 美元。在正收益的下面输入  $X$ 。请注意,  $X$  越大, 风险平均的可能性就越小, 一个小的  $X$  代表你更偏风险性。输入所需的输入, 选择  $U(x)$  类型, 然后单击计算函数获得结果。你也可以使用计算出的 ( $x$ ) 值来重新运算, 或恢复树到使用预期价值的成果。

# Risk Simulator

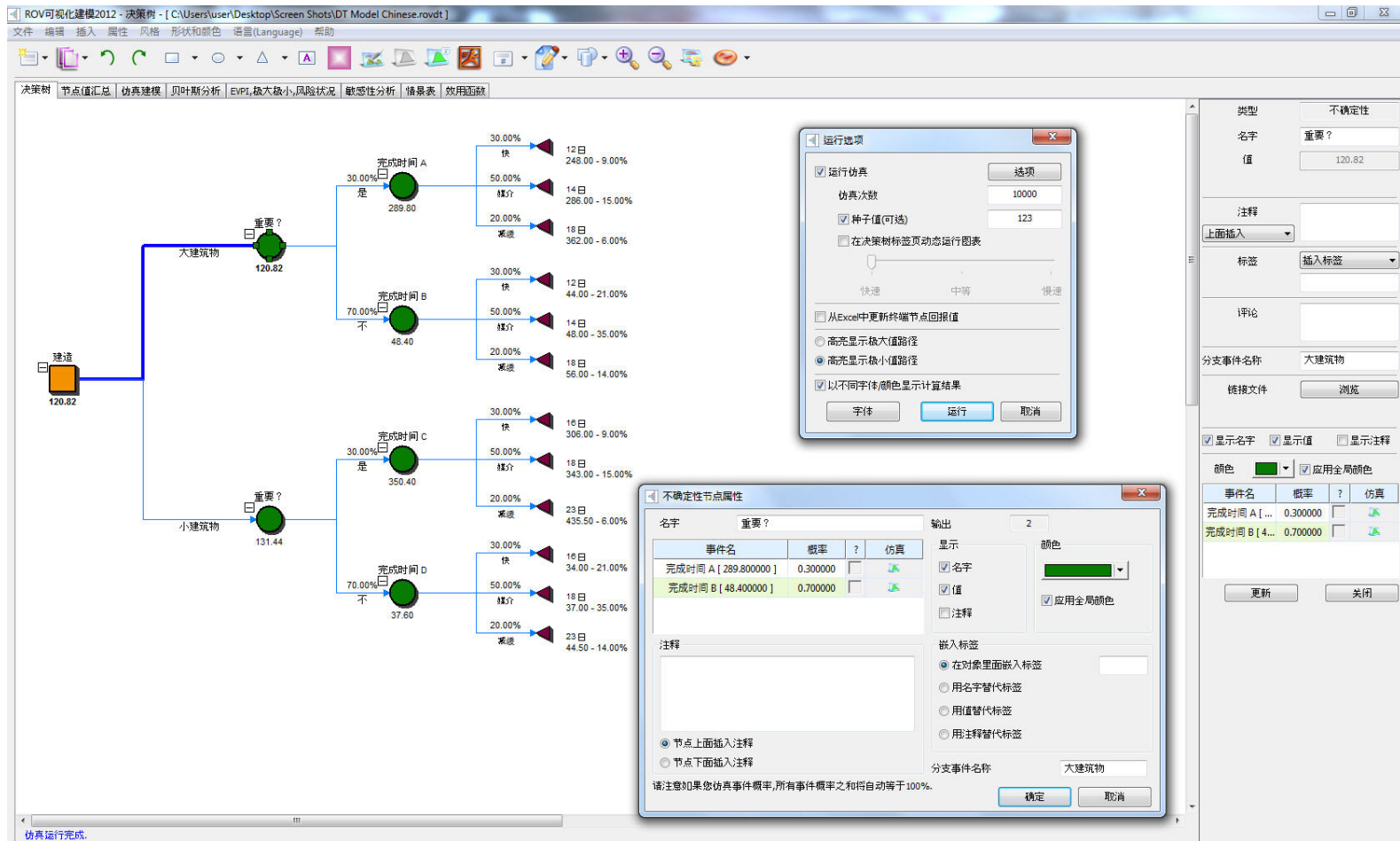


图 5.61 – ROV 决策树(决策树)

# Risk Simulator

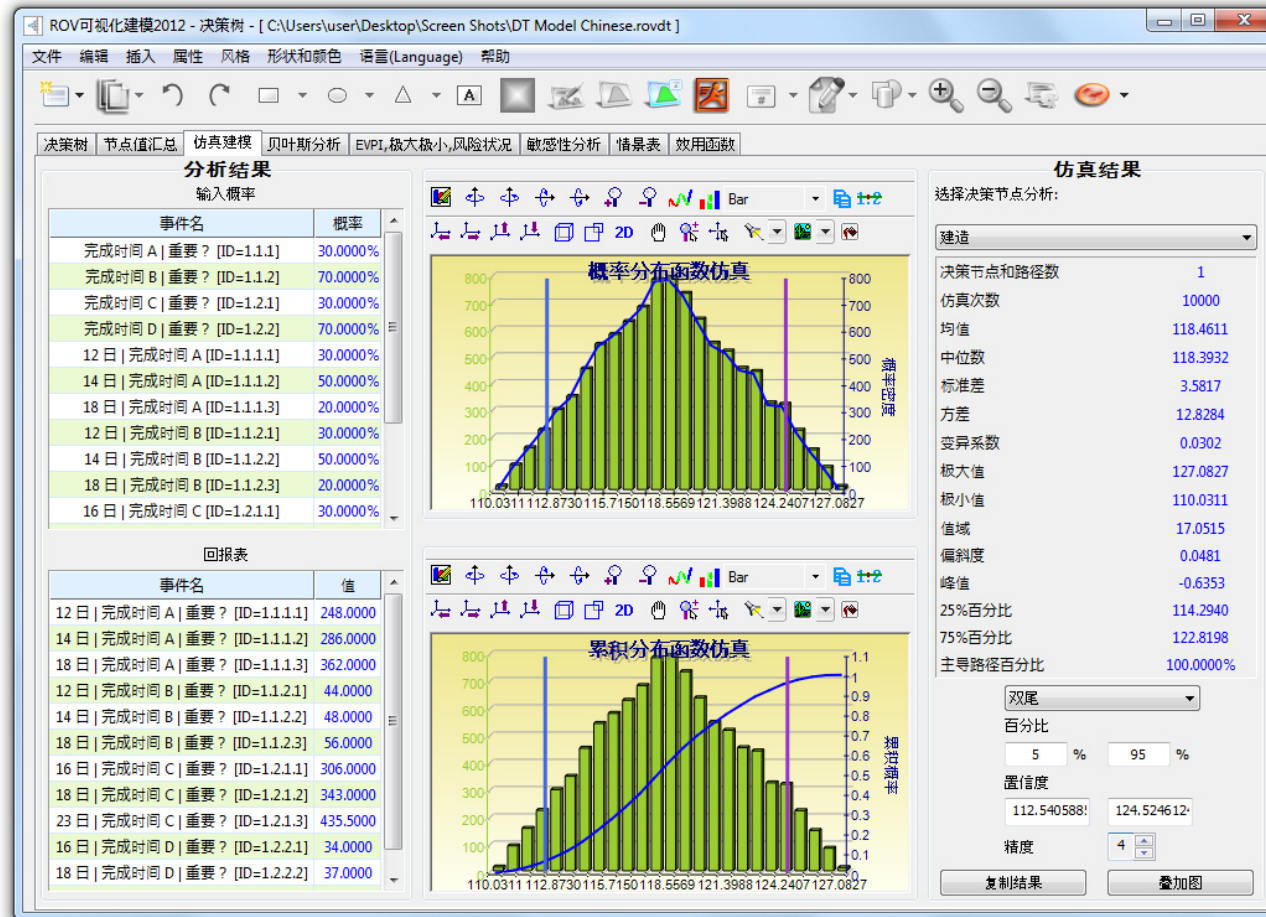


图 5.62 – ROV 决策树 (仿真结果)

# Risk Simulator

ROV可视化建模2012 - 决策树 - [ C:\Users\user\Desktop\Screen Shots\DT Model.rovdt ]

文件 编辑 插入 属性 风格 形状和颜色 语言(Language) 帮助

决策树 节点值汇总 仿真建模 贝叶斯分析 EVPI,极大极小,风险状况 敏感性分析 情景表 效用函数

贝叶斯分析工具可以执行任何两个沿路径链接的不确定性事件。例如,在右边的例子中,不确定性 A 和 B 链接。时间序列上事件 A 先发生,事件 B 后发生。第一事件 A 是Market Research(两个输出Favorable或Unfavorable)。第二事件 B 是市场条件(也是两个输出Strong和Weak)。该工具通过输入先验概率和可靠条件概率来计算联合、边缘、和贝叶斯后验更新概率;或输入后验条件概率计算可靠性概率。选择以下相关分析并点击加载示例来查看相对应的样例输入和右边表格中显示的结果,还有哪个结果是用来作为在决策树中的输入。

给定先验和可靠联合概率来计算贝叶斯更新后验概率(更多使用)  
 给定先验和后验概率来计算可靠联合概率(较少使用)

步骤一: 输入第一和第二不确定性事件的名称和选择每个事件有多少概率事件(自然状态或结果)。

第一事件名称: Market Research 概率事件或状态: 2  
 第二事件名称: Market Conditions 概率事件或状态: 2

步骤二: 输入每个概率事件或结果的名字。

状态	Market Research	Market Conditions
1	Favorable	Strong
2	Unfavorable	Weak

步骤三: 输入第二事件的先验概率和每个事件或结果的条件概率。概率之和应该等于100%。

事件	先验 P(x)	Favorable	Unfavorable	总计
Strong	45.00%	60.00%	40.00%	100.00%
Weak	55.00%	30.00%	70.00%	100.00%
总计	100.00%			

条件概率(可靠性) 保存的模型 名字

加载示例 计算

贝叶斯分析结果

先验概率和可靠条件概率

概率 (Strong)	45.00%
概率 (Weak)	55.00%
概率 (Favorable   Strong)	60.00%
概率 (Favorable   Weak)	30.00%
概率 (Unfavorable   Strong)	40.00%
概率 (Unfavorable   Weak)	70.00%

联合和边缘概率

概率 (Favorable)	43.50%
概率 (Unfavorable)	56.50%
概率 (Strong ∩ Favorable)	27.00%
概率 (Weak ∩ Favorable)	16.50%
概率 (Strong ∩ Unfavorable)	18.00%
概率 (Weak ∩ Unfavorable)	38.50%

后验或更新概率

概率 (Strong   Favorable)	62.07%
概率 (Weak   Favorable)	37.93%
概率 (Strong   Unfavorable)	31.86%
概率 (Weak   Unfavorable)	68.14%

图 5.63 - ROV 决策树(贝叶斯分析)



# Risk Simulator

ROV可视化建模2012 - 决策树 - [ C:\Users\user\Desktop\Screen Shots\DT Model.rovdt ]

文件 编辑 插入 属性 风格 形状和颜色 语言(Language) 帮助

决策树 | 节点值汇总 | 仿真建模 | 贝叶斯分析 | EVPI,极大最小,风险状况 | 敏感性分析 | 情景表 | 效用函数

完善信息的预期值, 最小极大和最大最小分析, 风险配置, 和不完全信息值

该工具计算完善信息的预期值(EVPI), 最小极大和最大最小分析, 和风险配置以及不完全信息值。输入决策分支或考虑的战略数(如, 构建一个大, 中, 小设施)和不确定性事件数或结果的自然状态(如, 畅销, 滞销), 和输入各种情况下的预期回报开始。

**输入假设**

决策分支: 3

不确定性或状态: 2

概率	状态 1	状态 2	总计
	80%	20%	100%

回报	状态 1	状态 2	均值
决策路径 1	8	7	7.80
决策路径 2	14	5	12.20
决策路径 3	20	-9	14.20
极大值	20.00	7.00	

加载示例

计算

**最小极大和最大最小分析**

最小极大(极大值中的最小值)和最大最小(最小回报中的最大值)是两种交替的找出最优决策路径的方法。这两种方法不经常使用但是仍然提供以更深入理解决策过程。输入决策分支数或存在的路径(如, 构建一个大, 中, 小设施), 和每条路径里的不确定性事件数或结果的自然状态(如, 经济良好 vs. 经济不景气)。完成不同情况的回报表和计算最小极大和最大最小值。您还可以点击加载示例来查看样例计算。

回报	状态 1	状态 2	极小值
决策路径 1	8	7	7.00
决策路径 2	14	5	5.00
决策路径 3	20	-9	-9.00

遗憾	状态 1	状态 2	极大值
决策路径 1	12.00	0.00	12.00
决策路径 2	6.00	2.00	6.00
决策路径 3	0.00	16.00	16.00

最小极大	极大最小
6.00	7.00

路径 2 是最优的  
路径 1 是最优的

**风险配置**

战略 1 风险配置	
回报	概率
248.00	9.00%
286.00	15.00%
362.00	6.00%
44.00	21.00%
48.00	35.00%
56.00	14.00%
概率总计	100.00%
预期值	120.82

战略 2 风险配置	
回报	概率
306.00	9.00%
343.00	15.00%

不完善信息的预期值: 10.62

保存的模型

名字:

添加 删除

**完善信息的预期值**

完善信息的预期值(EVPI), 如, 假设您有完美预见和精确知道做什么(通过市场调研或其他更好的理解概率结果), 相对于更不成熟的概率自然状态的估计, EVPI计算如果在这种信息里有添加值(如, 如果市场调研将添加值), 输入决策分支或考虑的战略数(如, 建立大, 中, 小设施)和不确定性事件数或结果的自然状态(如, 畅销, 滞销), 和输入各种情况下的预期回报开始。

自然状态的完善信息的预期值: 17.40

无自然状态完善信息的预期值: 14.20

完善信息的预期值: 3.20

图 5.64 – ROV 决策树 (EVPI, 最大最小, 风险文件)

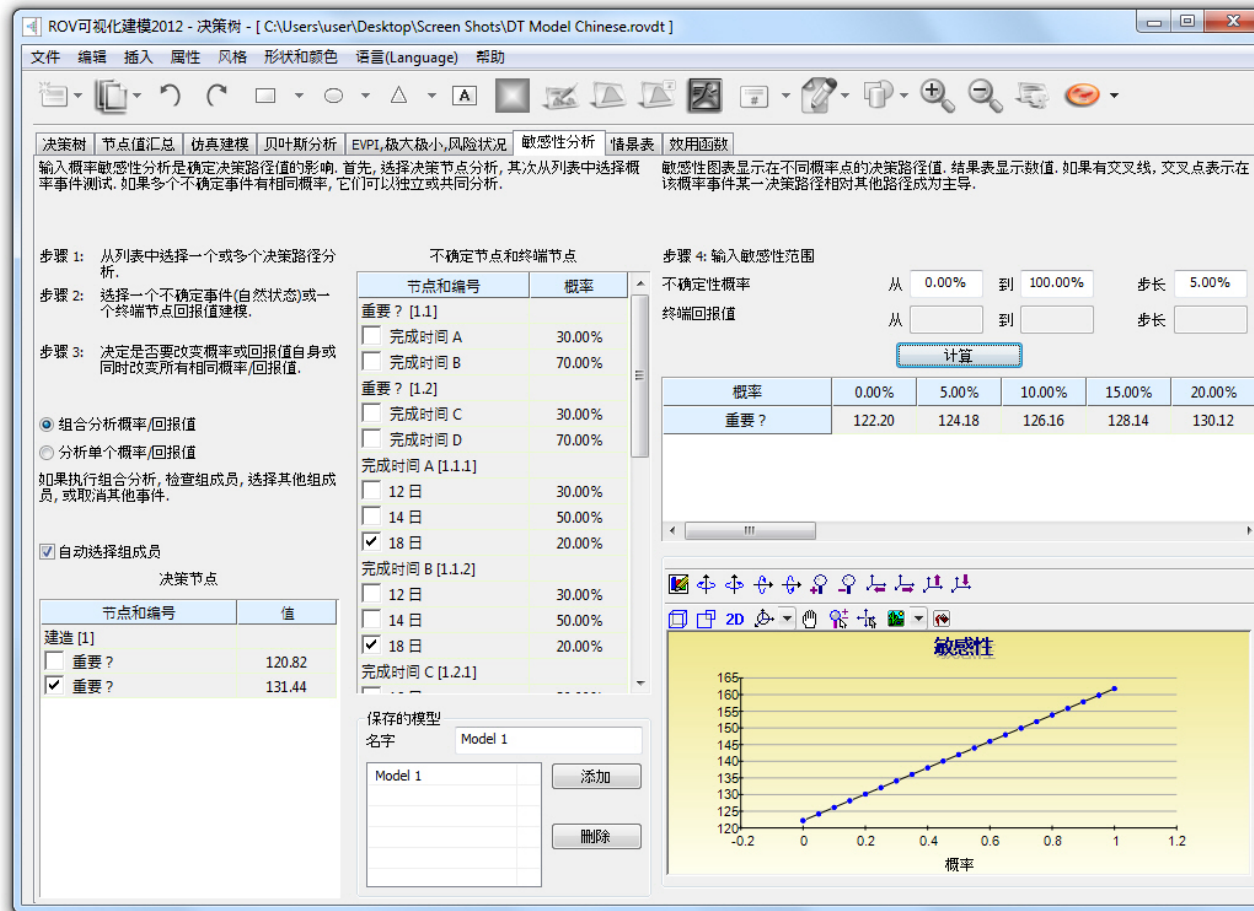


图 5.65 – ROV 决策树 (敏感性分析)

# Risk Simulator

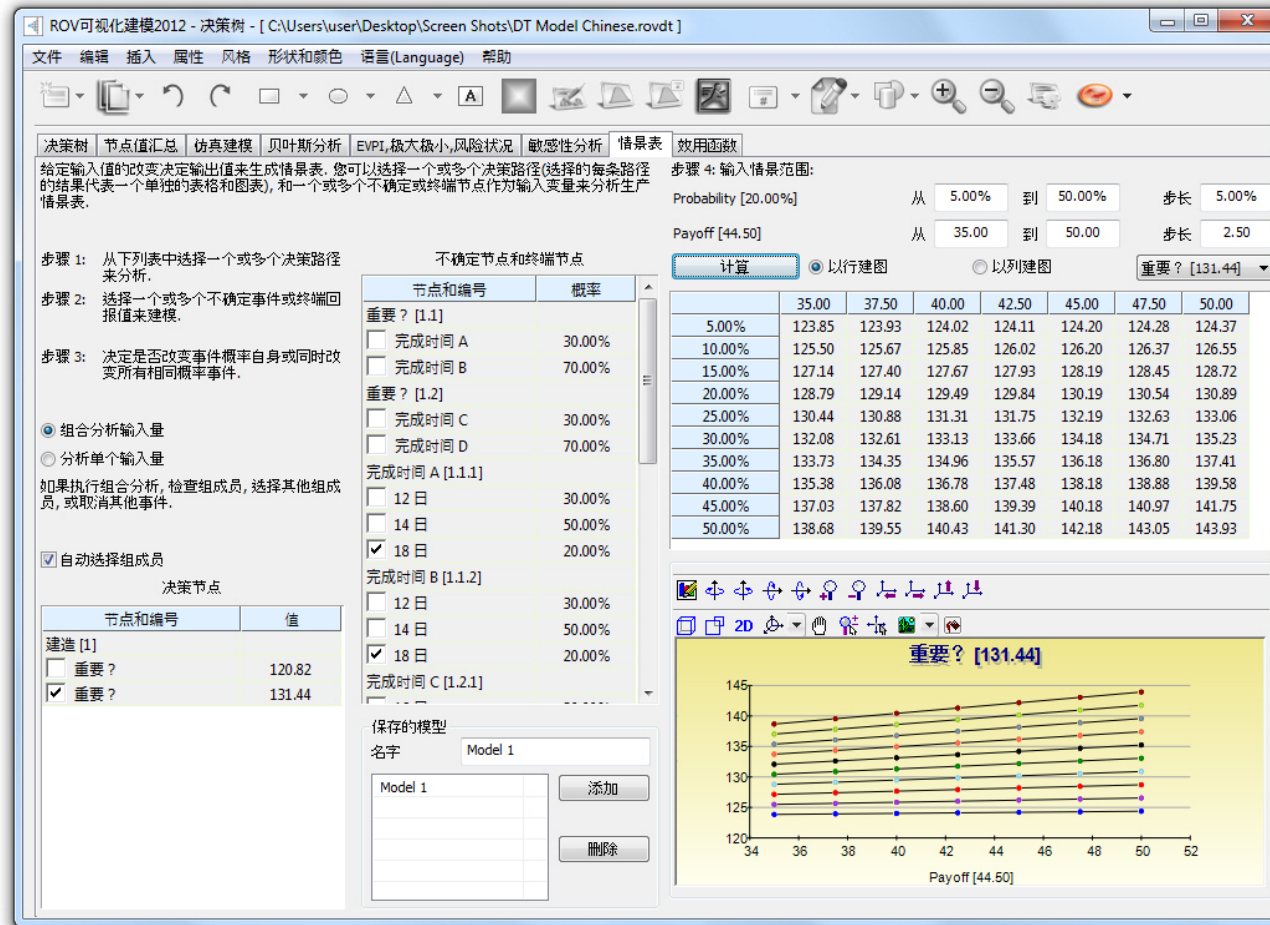


图 5.66 – ROV 决策树 (情景表格)

# Risk Simulator

ROV可视化建模2012 - 决策树 - [ C:\Users\user\Desktop\Screen Shots\DT Model.rovdt ]

文件 编辑 插入 属性 风格 形状和颜色 语言(Language) 帮助

决策树 节点值汇总 仿真建模 贝叶斯分析 EVPI,极大极小,风险状况 敏感性分析 情景表 效用函数

效用函数生成

效用函数, 或 $U(x)$ , 有时用在决策树中终端回报的期望值中.  $U(x)$ 可以发展成两种方式: 用于每种可能的回报的冗长而详细的实验过程或指数外插法(这儿用到的). 它们可以仿照一个厌恶风险的(损失比同等升值潜力更严重或痛苦), 风险适中的(上升和下降有相同吸引力), 爱好冒险的(升值潜力更具吸引力)决策者. 输入最小和最大终端回报期望值和之间的数据点数来计算效用曲线和表格.

最小期望值 34.00 最小期望回报值来生成 $U(x)$ 曲线的起点

最大期望值 435.50 最大期望回报值来生成 $U(x)$ 曲线的终点

$U(x)$ 计算点数 50 在最小和最大 $U(x)$ 区间的步长数

如果您有50:50风险赌率, 相对于不赌只得到¥0回报, 或者得到¥ $x$ 或失去-¥ $x/2$ , 那么¥ $x$ 应该是多少? 例如, 相对于不玩如果您不关心在相同概率下您可以得到¥100或失去-¥50, 这时 $x$ 是¥100. 在下面正收益框中输入 $x$ . 注意 $x$ 越大, 风险规避越小, 反之 $x$ 越小表示风险规避越大.

正收益 217.75 输入需求量, 选择 $U(x)$ 类型, 点击计算效用得到结果. 您还可以应用计算的 $U(x)$ 值到决策树中重新运行, 或恢复树中的回报期望值.

可能损失 -108.88

等值 0

期望值	$U1(x)$	终端回报	$U1(x)$
34.0000	0.1446	248.0000	0.6798
42.1939	0.1762	286.0000	0.7311
50.3878	0.2066	362.0000	0.8103
58.5816	0.2359	44.0000	0.1830
66.7755	0.2641	48.0000	0.1978
74.9694	0.2913	56.0000	0.2268
83.1633	0.3175	306.0000	0.7547
91.3571	0.3427	343.0000	0.7930
99.5510	0.3669	435.5000	0.8647
107.7449	0.3903	34.0000	0.1446
115.9388	0.4128	37.0000	0.1563
124.1327	0.4345	44.5000	0.1830

U1: 厌恶风险效用函数(正负效用)  
 U1: 厌恶风险效用函数(正正效用)  
 U2: 厌恶风险效用函数(仅正效用)  
 U3: 厌恶风险效用函数(0到1之间校正)  
 U4: 厌恶风险效用函数(0到100之间校正)  
 U5: 风险适中效用函数(0到1之间校正)  
 U6: 风险适中效用函数(0到100之间校正)  
 U7: 爱好冒险效用函数(仅正效用)  
 U8: 爱好冒险效用函数(仅正效用)  
 U9: 爱好冒险效用函数(0到100之间校正)

删除

图 5.67 – ROV 决策树 (效用功能)



## 6. 帮助贴士和技巧

下面是对于 Risk Simulator 高级用户的一些快速指南和简要的技巧。对于某些特定工具的详细信息，请参考用户手册的相关部分。

### 贴士：假设（设定输入假设用户界面）

- 快速跳跃—输入任何字母选择某个分布，将会直接跳到以该字母开头的分布中（例如，点击正态分布按 W 键，就会直接跳到 Weibull 分布）。
- 右击查看—选择任何分布，右击选择不同的分布形态（大图标，小图标，列表形式）。
- 更新图表栏—再输入新的输入参数（例如，输入了新的均值或者标准差），点击 Tab 键或者点击用户界面任意位置自动查看图表更新。
- 输入相关性—可以直接输入两两相关性（列会根据需要重新调整大小），使用多重分布拟合工具，以及计算两两相关系数，或者设定假设，使用相关性编辑器输入相关矩阵。
- 假设单元格函数—只有空单元格或者包含静态值的单元格可以设定为假设；但是，有时候函数或者公式要求作为假设单元格，可以首先在单元格输入输入假设，然后输入公式或者函数（当仿真运行的时候，仿真的值将取代函数，在仿真完成之后，公式和函数会再次展示）。

### 贴士：复制和粘贴

- 使用 Esc 进行复制和粘贴—当选择一个单元格，使用 Risk Simulator 的复制功能，就会将所有的信息复制到 Windows 的剪切板，包括单元格的数值，函数，方程，颜色，字体，大小，包括 Risk Simulator 的假设，预测，或者决策变量。当应用 Risk Simulator 的粘贴功能，会有两个选项。第一个选项直接应用 Risk Simulator 的粘贴功能，所有的单元格值，颜色，字体，函数和公式以及参数都将粘贴到新单元格中。第二个选项首先点击 Esc 键，然后应用 Risk Simulator 的粘贴，Esc 会告诉 Risk Simulator 只想粘贴假设，预测或者决策变量，而非单元格的值，颜色，公式，函数，字体等等。在点击 Esc 之前，可以保持目标单元格的值和计算，Risk Simulator 的参数只被粘贴过去。
- 在多重单元格中复制和粘贴—选择多个需要复制和粘贴的单元格（包括连续和非连续的假设）。

## 贴士：相关性

- 设置假设—使用输入假设对话框设置相关性（理想情况只可以输入定量的相关性）。
- 编辑相关性—手动输入或者从 Windows 剪切板粘贴输入相关性矩阵。
- 多分布拟合—自动计算和输入配对相关性（当进行多个变量拟合时，自动计算相关性，并且决定哪些具有显著相关性）。

## 贴士：数据诊断和统计分析

- 随机参数估计—在统计分析和数据争端报告中，有一栏关于随机参数估计，基于历史数据可以计算波动率，漂移率，均值恢复率，跳跃扩散率。注意，这些参数结果都是基于历史数据的，参数会随着时间改变而改变。而且，分析的结果并不表明哪个模型是拟合最好的（例如，布朗运动，均值恢复，跳跃扩散，或者混合过程）。这个由用户自己决定根据时间变量进行预测。只允许客户自己决定使用什么过程（例如，布朗运动用于模拟股价波动，但是分析工具并不知道变量的性质）。最后，在这里提示，某个参数在正常范围之外，该过程要求的输入参数可能并不是正确的过程（例如，如果均值恢复率是 110%，均值恢复过程就不是适合的过程，如此类推）。

## 贴士：分布分析，图表

- 分布分析—快速计算 Risk Simulator 中 42 种分布的 PDF，CDF 和 ICDF，并且以表格形式返回这些值。
- 分布图表—用于比较相同分布的不同参数（例如，具有不同 Alpha, Beta 参数[2,2], [3,5], [3.5,8]的 Weibull 分布的 PDF，CDF，ICDF 的形状，并且将彼此重叠到一起）。
- 重叠图—用于比较不同的分布（理论输入假设和试验仿真输出预测）并且将彼此重叠比较。

## 贴士：有效边际

- 有限边际变量—可以查看有效边际变量，首先在设定有效边际设定模型限制。

## 贴士：预测单元格

- 不包含公式的预测单元格—可以在不包含任何公式或者值的单元格中设置预测（只需要跳过警告信息）但是预测图表将不包含任何信息。空值的输出预测单元格一般用于包含宏的单元格。

## 贴士：预测图表

- *Tab* 和空格—输入一些参数后敲击 Tab 更新预测图表或者百分位数和置信值，敲击空格键在预测图表的不同栏目中切换。

- 常规视图和全局视图一点击这些视图在常规视图和全局视图直接切换，全局视图可以一次性显示所有的图表。
- 复制一复制的预测图表取决于当前是常规视图还是全局视图。

### 贴士：预测

- 单元格链接地址一如果选择电子报个中的数据，然后运行预测工具，所选数据的单元格地址会自动输入到用户界面，否则的话需要手动输入单元格地址或者使用链接图标链接相关数据地址。
- 预测 RMSE一多个预测模型使用误差估计，比较每个模型的精度。

### 贴士：预测：ARIMA

- 预测期一外生变量数据必须至少超过预测期数（例如，如果想用 100 个时间序列点预测未来 5 期，那外生变量至少需要 105 个），否则的话运行不包含外生变量的 ARIMA，就可以预测没有任何限制的期间。

### 贴士：预测：基本计量经济学

- 变量通过分号隔开一使用分号隔开自变量。

### 贴士：预测：Logit, Probit 以及 Tobit

- 数据要求一运行 logit 和 probit 检验的因变量必须是二元的（0 和 1），而 tobit 模型可以接受二元和其他浮点数值。自变量对于任何三种模型可取任何值。

### 贴士：预测：随机过程

- 默认样本输入一使用默认样本输入作为模型的起始点。
- 用于参数估计的统计分析工具一使用该工具通过估计原始数据校正随机模型的输入参数。
- 随机过程模型一有时候如果随机过程的用户界面显示了较长的时间，可能是输入不正确或者对应模型指定不对（例如，如果均值恢复率是 110%，可能均值恢复过程不是正确的过程，等等）。请使用其他不同的输入或者使用不同的模型。

### 贴士：预测：趋势线

- 预测结果一拖卷报告的结尾部分查看预测的值。

### 贴士：函数调用

- RS 函数一可以读取假设变量得到预测统计函数。得到这些函数，首先安装 RS 函数（开始菜单，所有程序，Real Options Valuation, Risk Simulator, 工具，安装函数）然后在设定 RS 函数之前运行仿真。引用示例模型 24 作为案例指导如何使用这些函数。

## 贴士：入门练习和视频

- 入门练习— 在开始菜单，所有程序，Real Options Valuation, Risk Simulator 快捷方式，包含有多个入门练习和示例结果解释。这些示例可以让用户快速入门 Risk Simulator。
- 入门视频— 这些视频都可以在我们的官方网站查看 [www.realoptionsvaluation.com/download.html](http://www.realoptionsvaluation.com/download.html) or [www.rovdownloads.com/download.html](http://www.rovdownloads.com/download.html)

## 贴士：硬件 ID

- 右键 HWID 复制— 在安装授权用户界面，选择或者在 HWID 上双击选择该值，右键复制或者点击 Email HWID 链接生成 包含 HWID 的邮件。
- .Troubleshooter— 在开始菜单，所有程序，Real Options Valuation, Risk Simulator 文件夹运行 Troubleshooter，以及运行 Get HWID 获得计算机的 HWID。
- 贴士：拉丁超立方抽样（LHS）vs Monte Carlo 仿真
- .相关性— 设定相互相关假设之后，建议在 Risk Simulator，选项菜单使用 Monte Carlo 设置，拉丁超立方方法与 copula 仿真法不兼容。
- LHS 间距— 间距越多，当进行仿真的时候会降低仿真的速度。
- 随机化— 所有在选项菜单的随机仿真技术都经过了检验，方法和引擎的随机化程度相当高。

## 贴士：在线资源

- 书籍，入门视频，模型，白皮书— 都可以免费在 [www.realoptionsvaluation.com/download.html](http://www.realoptionsvaluation.com/download.html) 或者 [www.rovdownloads.com/download.html](http://www.rovdownloads.com/download.html) 获得。

## 贴士：优化

- 次优解— 如果优化运行返回次优解，尝试将限制条件从等于 (=) 改变为不等于 (>=或者<=) 再试一次。该操作对与运行有效边际分析相同适用。

## 贴士：文档

- .多重文档— 创建并且在一个模型的多个文档间切换。该操作可以进行仿真情景分析，并且可以更改输入参数或者分布类型来查看结果。
- 文档要求— 假设，预测，或者决策变量在无文档的情况下无法被创建。但是，一旦创建了文档，就没必要每次都创建新文档。实际上，如果想运行添加更多假设或者预测的仿真模型，可以在相同的文档中更改。
- 激活文档— 最后适用的文档将随着 Excel 的保存而保存，打开而打开。
- 多重 Excel 文档— 当在多个打开的 Excel 模型中切换，激活的文档也将在不同 Excel 模型中切换。
- 跨工作簿文档— 当有多个 Excel 文档被打开或者只有一个 Excel 文件和激活文档的时候，请注意，突然切换到另外一个 Excel 文档以及文档包含设



定假设和预测，假设和预测将不会运行。

- 删除文档—可以复制当前的文档或者删除当前文档但是注意删除文档的时候必须保证还存在有一个文档最后留下。
- 文档位置—创建的文档（包含假设，越策，和决策变量，目标，限制，等等）都保存在压缩隐藏的工作簿中。这就是每次保存 excel，文档也相应保存的原因。

### 贴士：右键快捷键或者其他快捷键

- 右键—可以在 Excel 单元格中右键打开 Risk Simulator 快捷菜单。

### 贴士：保存

- 保存 Excel 文档—该操作保存了文档的设定，假设，预测，决策变量，以及 Excel 模型（包括 Risk Simulator 报告，图表，和输出的数据）。
- 保存图表设置—该操作保存了预测图表设定例如可以被回复的相同设置，可用于未来的预测图表（在预测图表中使用保存和打开图标）。
- 在 Excel 中保存和输出仿真数据—该操作可以在 Excel 中输出仿真的假设和预测，Excel 本身也将被保存以保存这些数据。
- 在 Risk Simulator 中保存仿真数据和图表—使用 Risk Simulator，数据输出，和保存为\*.RiskSim 格式可以再次打开包含相同数据的动态和预测图表，而不需要再次运行仿真。
- 保存和生成报告—仿真报告和其他分析报告都以单独的工作簿输出，整个 Excel 文件都会保存以保存数据用于未来的研究。

### 贴士：抽样和仿真技术

- 随机数生成器—包含 6 中随机数生成引擎(查看用户手册了解详情)，ROV Risk Simulator 默认的方法和高级负随机移动法是推荐是同的方法。不要使用其他的方法，除非模型或者分析会调用这些方法，建议对这两种方法进行结果测试。

### 贴士：软件卡发工具（SDK）和动态链接库

- SDK，DLL 和 OEM—所有 Risk Simulator 中的分析都可以被外部的软件调用，并且和其他产权的软件相兼容。请[联系 admin@realoptionsvaluation.com](mailto:admin@realoptionsvaluation.com)获得更多详细的内容。

### 贴士：在 Excel 中使用 Risk Simulator

- ROV Troubleshooter—运行 ROV Troubleshooter 获得自己计算机的 HWID 以获得授权，查看计算机的设定，以及解决 Risk Simulator 是否被阻止。
- 和 Excel 一起启动 Risk Simulator—可以让 Risk Simulator 和 Excel 一起自动启动，或者点击开始菜单，所有程序，Real Options valuation，Risk Simulator 快捷方式。这些都可以在 Risk Simulator 的选项中设定。

### 贴士：极速仿真

- 模型开发—如果在模型中想要进行极速仿真，当模型构建完成后，完全可以运行急速仿真。
- 常规速度—毫无疑问，常规速度也是可以进行的。

### 贴士：飓风图分析

- 飓风图分析—飓风图分析不能单单只运行一次。它可以作为模型诊断工具，可以在同一模型中运行多次。例如，在较大的模型中，飓风图可以使用默认设置和所有的自变量（选择显示所有变量）运行一次。否则的话，该工具提供了起始点用来确定哪些变量是关键成功因素（例如，飓风图可以显示对输出影响最大的头 5 个变量，剩下的 200 个可能就具有较小的影响），第二次飓风图分析可以运行较少的变量（例如，如果头 5 个影响较大，就选择前 10 个变量，这样就可以创建包含关键和非关键因素的图表）。
- 默认值—默认的测试点可以增加 $\pm 10\%$ 到某个更大的值来检验非线性（蜘蛛图将显示非线性以及飓风图可以偏向一边如果变量是非线性的话）。
- 零值和整型值—包含零和整数的输入需要在飓风图取消选择，才能进行运算。否则，进行扰动就没有效果（例如，如果模型用户使用 lookup 表格，当 Jan = 1, Feb = 2, Mar = 3 等等，围绕 1 扰动 $\pm 10\%$ 可以得到 0.9 和 1.1，这对于模型没有意义）。
- 图表选项—使用多个图表选项获得最优的打开或者关闭模型的选项。

### 贴士：Troubleshooter

- ROV Troubleshooter—运行 Troubleshooter 或者电脑的 HWID 用于授权，查看电脑的设置，以及重新激活 Risk Simulator 如果在偶然条件下被阻止。