RISK SIMULATOR
用户手册

乔纳森 文 **(Dr. Johnathan Mun) Ph.D.** 博士，**MBA**，注册风险分析师，注册金融分析师，
金融风险管理师

王立峰，注册风险分析师

李倩，注册风险分析师

黎毅，注册风险分析师

张元昊，注册风险分析师

Real Options Valuation

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 or visit www.realoptionsvaluation.com.cn

目录

对数正态分布
正态分布
Pareto 分布
Student's t 分布
三角分布
均匀分布
Weibull(Rayleigh 分布)

3．预测

不同类型的预测方法
运行 Risk Simulator 中的预测工具
时间序列分析
多元回归
随机预测
非线性外推
Box-Jenkins ARIMA 高级时间序列
自回归求和移动平均模型(Box-Jenkins ARIMA 高级时间序列)
基本计量经济学模型
J-S 曲线预测
GARCH 波动率预测
马尔可夫链
最大似然估计模型（MLE）
样条模型 (三次样条内插和外插模型)

4．优化

优化方法
连续型决策变量的优化
离散变量的优化

5．风险仿真分析工具

仿真中的飓风图和敏感性工具
敏感性分析
分布拟合：单变量和多变量
Bootstrap 仿真
假设检验
数据输出和保存仿真结果
创建报告
回归和预测诊断工具
统计分析工具
分布分析工具

1. 系统简介

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 2006；*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。

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 24 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）

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 Excel XP, 2003, 2007 版及以上
- Microsoft .NET Framework 2.0/3.0
- 80M 空间
- 推荐 1GB 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

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 or 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。

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 和 Excel 2007：

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 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 使用 Excel XP，Excel 2003 或 Excel 2007：

首先，在 Windows Vista 中打开 Excel2007，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 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。

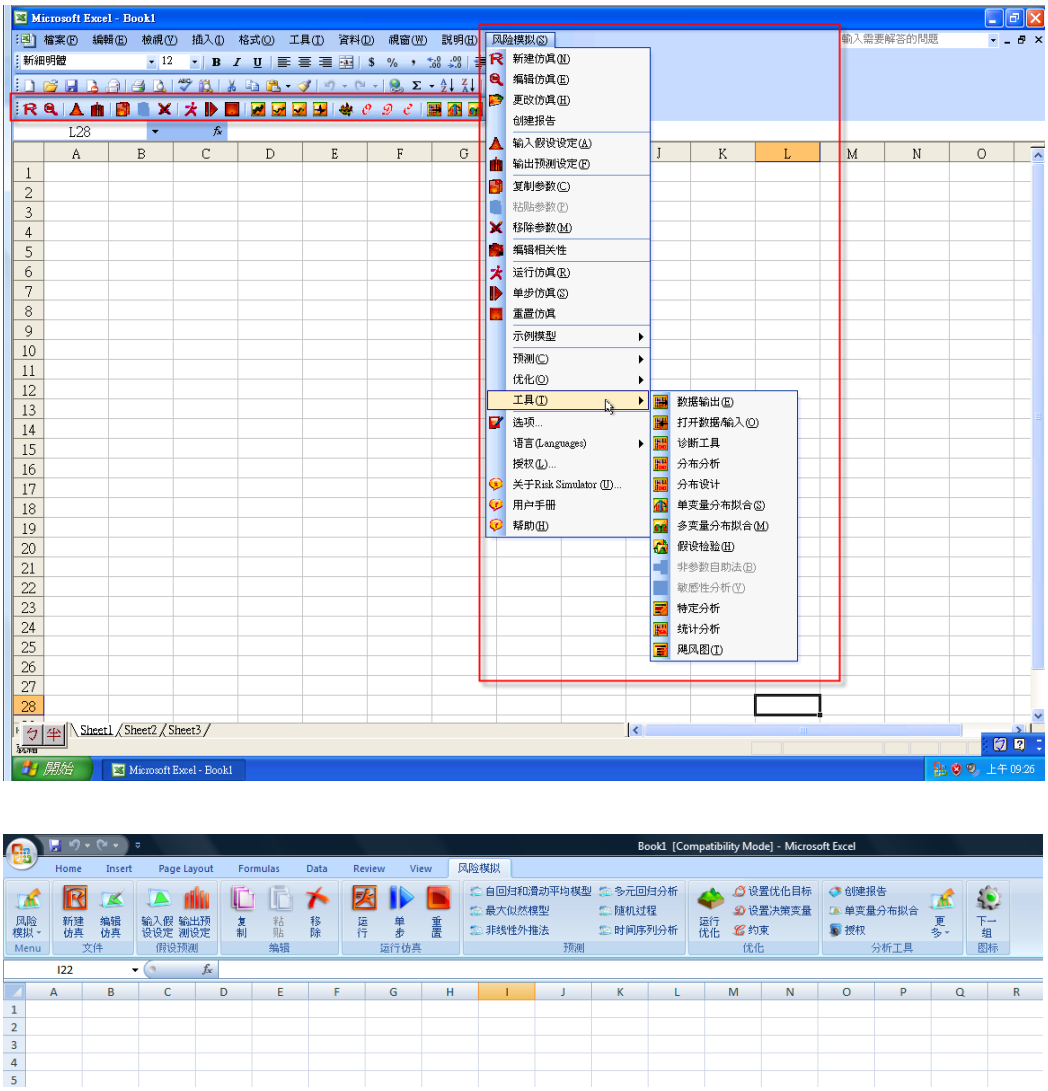


图 1.1 ——Risk Simulator 菜单和图标

图 1.2 ——Risk Simulator 启动画面

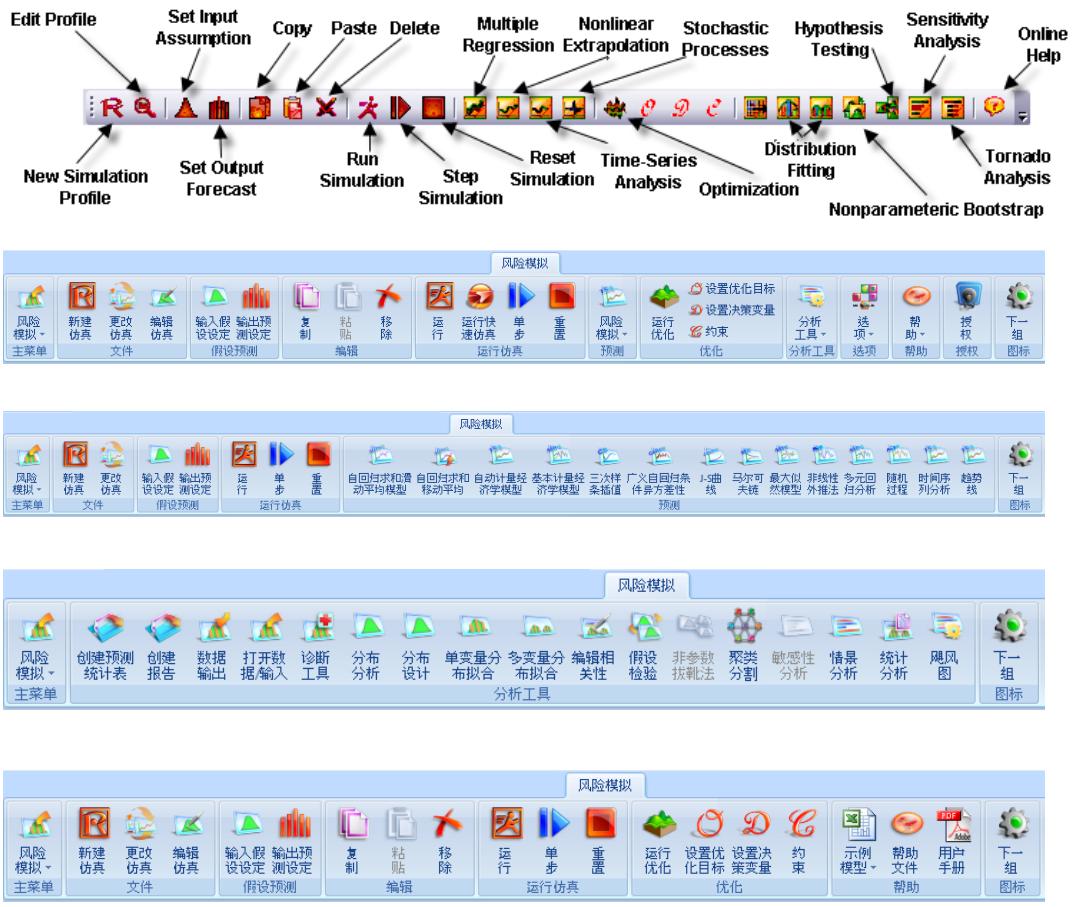


图 1.3 ——Risk Simulator 工具栏

风险模拟 5 有哪些新增功能？

最新版的**风险模拟**和之前的版本相比，新增了很多功能，并且对之前版本软件的很多功能也进行很大了改善，详细介绍如下：

- **第五版新增功能：**
  - **超快仿真：**这个新功能可以让你以超快的速度运行仿真。首先通过分析你的 Excel 模型，然后将此模型汇编成纯数学代码然后以非常快的速度运行仿真。对于某些不能进行汇编的模型将以常规的速度运行仿真（例如，带有 VBA 函数或宏的模型，连接到外部数据或文件的模型，软件不支持的或错误的函数，或带有错误的模型）。
  - **修改了在 Excel2007 中的图标：**对于 Excel 2007 的用户来说，你将看到一个全新的更加直观和易于使用的图标工具栏。有四套图标将保证软件与大部分的屏幕分辨率相匹配（1280 x 760 和以上）。
  - **改善了的预测图：**预测图现在有了以下几项改善：
    - **带截断的统计量：**如果你在预测图中进行数据过滤（在选项标签中设置数据滤波段），统计量标签将基于截断的数据显示更新后的统计量。如果你不截断预测数据，所有数据的统计量将不会发生变化。
    - **图表控制器（3D/斜体/移动，颜色，拟合，概率密度函数/累积分布函数）：**这是在预测图上的一个新标签，在这里你可以修改现有预测图，包括对预测数据运行分布拟合分析，创建 PDF/CDF/ICDF 图，更改图表选项（图表类型，3D 旋转，颜色，缩放，小数位，坐标轴上的最小值和最大值，标题名，和很多其它选项，包括可以保存修改设定和以多种格式打印图表）。
  - **计量经济学自动分析功能：**通过测试线性，非线性，滞后数据，先导数据，相关影响，嵌套和其它模型，这个新的预测工具可以通过使用智能优选法来运行上百甚至上千种模型的组合和排列来找到最拟合数据的模型。这个工具和 ROV Risk Modeler 软件的计量经济学自动分析功能相似，计量经济学自动分析功能可以对一个很大的数据集合运行成千上万，甚至上百万种模型。
  - **计量经济学基本功能：**这个现有的预测工具的功能已经被大大地提升，新的功能包括可以创建新的变量和函数，例如 TIME（一个线形时间序列变量），DIFF(对时间序列数据一阶微分)，RESIDUAL（数据来自你指定的一个预测方程的误差项），RATE（时间序列数据的一阶比率），和 FORECAST（数据来自你指定的一个预测方程的误差项）。
  - **趋势线分析：**这个新的工具可以运行大部分常用的趋势线模型包括线形，非线形，指数，幂函数，移动平均，和多项式模型。运行模型后，可以得到一系列图表和每个模型的最优统计量。
  - **覆盖图：**通过将假设变量和/或预测变量以时间序列或截面数据以覆盖图的方式在图上标定出来，这个新的绘图工具可以用于比较多个假设变量和/或预测变量。这样可以使你快速地观看假设量或预测量之间的相似性和差异性，更容易阅读图表。
  - **分割聚类：**通过应用一些智能算法和优选法，这个工具可以对一个大的数据集合进行隔离，聚类或者分成具有不同统计特性的组。

- **创建预测统计量表格：**这个新工具可以对一些关键的预测统计量（例如，均值，中值，众数，标准差，方差，变异系数，斜度，峰度）和置信区间，还有你选定的输出预测变量的概率创建报告。结果是一个比较图表，这个比较图表上列示了你从多个预测变量中选定的统计量。
- **灵活的授权方式：**获取软件授权的方式有了以下的改善：模块
  - 对于进入控制设置为开启或者没有管理登陆限制的 Vista 用户来说，仍然可以无须投入额外的努力就可以成功地安装软件授权，享受**风险模拟**软件的全部功能。如果需要安装一个你收到的新的授权文件，只需要简单地启动 Excel，点击**风险模拟**，授权，安装授权，浏览你的授权文件，将此授权文件导入即可。这样你就可以永久地激活此软件或享有使用此软件的一段时间。
  - **风险模拟**现在可以让你根据你的风险分析经验对某些功能进行开启或关闭。例如，如果你只对**风险模拟**的预测工具感兴趣，你可以获取一个特别的只激活预测工具的授权码，不激活和使用其它的模块，这样你就可以以更低的成本使用本软件。四个可以开启或者关闭的模块分别是仿真模块，预测模块，优化模块和分析工具模块。另外，每个模块中的特定工具和分析方法也可以被开启或关闭，这种定制的服务仅对超过 10 台电脑的定点授权购买适用。

- **高级预测模型：**把 5 版本上新的预测工具和技术包含在内，**风险模拟**软件现在共包含以下的预测方法：
  i. **ARIMA (自回归求和滑动平均模型)**
  ii. **自动 ARIMA**
  iii. **计量经济学自动分析功能**
  iv. **计量经济学基本功能**
  v. **三次样条插值法**
  vi. **GARCH (广义自回归条件异方差)**
  vii. **J-曲线**
  viii. **马尔可夫链**
  ix. **最大似然法**
  x. **非线形外推**
  xi. **回归**
  xii. **S-曲线**
  xiii. **随机过程**
  xiv. **时间序列分析**
  xv. **趋势线**

- **风险模拟 第四版或之前版本软件功能的全面改进：**
  - **Excel RS 函数：**你可以在 Excel 电子表格的任何地方通过点击**插入函数**然后卷动选择以“RS”开始的函数进入**风险模拟** 函数。在这里，你可以设置假设变量和获得预测变量的预测统计量。例如，你可以运行 RSAssumptionNormal 函数来为一个单元格设置正态分布假设，或者运行 RSForecastStatistic 来获取一个预测单元格的统计量。在设置假设预测的时候，你可以设置一个占位符或临时值（这个值在运行仿真之前和运行仿真之

后都不变），假设名（变量名），分布参数（例如，均值，标准差），和其它的选项例如百分位数，相关性，最小和最大边界。对于结果，你还可以使用*RSForecastStatistic(A1, "Percentile99.9")*来获取单元格 A1 的 99.90 百分位数，这个单元格有一个预测参数集。可以使用的函数包括"PercentileXXX", "CertaintyXXX", "Mean", "Median", StandardDeviation", "Variance", "Skewness", 和"Kurtosis"。

- **在 Excel 右点击：**现在你可以在 Excel 里通过右点击鼠标来快速进入**风险模拟**进行操作，例如设置假设变量，设置预测变量，和运行防真。
- **百分位数和条件平均数：**在随机优化中，我们现在可以获得额外的统计量信息，包括百分位数和条件平均数，例如只要某个值大于 A 或者小于 A 就能获得平均值，这在计算条件在险价值（VaR）的时候非常关键。
- **变异系数（CV）：**在预测图的统计量中，绝对离差均值已经更改为变异系数（CV），CV 是标准差用均值来除，有时候可以作为波动性的一个近似替代，在比较不同规模的项目的时候，CV 作为一个相对风险的测量指标非常有用，也作为一个风险-回报比率。
- **情景分析：**这个新工具用于计算你的模型中的不同情景，通过同时更改一个或者两个输入变量，对于输入变量一定范围的变化，可以确定对输出结果的影响。
- **强大的飓风图：**额外的检查清单和选项，还有更加稳定和强大的飓风图分析，可以帮助你在多个工作表中运行飓风图分析。你还可以进行全局设定（改变一个变量，例如测试 10%上升和下降，你可以控制单个引用单元改变还是所有的引用单元同时改变），强调或忽略可能的整数值（有时候在一个模型中整数值作为一个标记，这个选项可以帮助你辨析那些在运行飓风图分析的时候你可能希望忽略的某些引用单元），现在在敏感性分析表格中还包括工作表名，这可以帮助用户更容易识别某些变量，还包括很多其它的改善。
- **有效前沿：**这个优化工具可以对变化的约束条件运行多种优化。你可以在优化选项中通过设置约束对话框来进入并使用这个工具。这个技术还可以同时运行静态，动态和随机优化。
- **重新启用风险模拟：**这个工具可以在菜单***开始 | 所有程序 | Real Options Valuation | 风险模拟***中启动。当 Windows 或者 Excel 暂时禁用本软件的时候（当你运行仿真的时候电源中断，你的电脑中有病毒或者木马程序，或者你错误地删除了关键的文件等等），你可以重新启用**风险模拟**。
- **多阶段优化：**此模块现在配备在多阶段优化应用中，在"高级"选项（当你运行优化的时候会出现这个选项）中还有一个局部和全局最优化测试。当一起使用这两个新特性的时候还有一个高级的特性，可以使用户对优化的运行有更好的控制，和增加准确度和结果之间的依存关系。

- **统计分析工具：**选定你希望进行分析的数据，包括标题，然后启动此工具（位于***风险模拟 | 工具 |统计分析***），你将可以获得以下分析结果：
  - **描述性统计**，包括分布的四矩和其它的置信区间测量。
  - **分布拟合，**测试数据是否可以拟合成某种分布。
  - **假设检验，**检验数据是否与某个特定的值在统计上是否显著相似或显著不同。
  - **非线形外推，**测试一个时间序列数据在本质上是否是非线形的。

- **正态性，**测试数据是否在统计上接近一个正态分布。和假设检验一样，这是一个非常重要的统计属性，因为很多建模技术都需要进行正态性假设。
- **随机参数估算，**确定一个随机游走过程，均值回复过程，或者一个跳跃扩散过程的输入参数，和决定被解释变异是否足够来证实使用随机过程进行预测是合理的。
- **自相关性测试，**可以对数据进行测试来确定是否可以使用这些时间序列数据的历史来预测未来。
- **时间序列预测**，测试时间序列数据基准水平的变动，趋势和季节性的影响。
- **趋势分析，**测试数据是否符合线性时间趋势，如果符合，可预测性的程度是多少

- **高级数据诊断工具：**选择你希望进行分析的数据，包括标头，然后启动此诊断工具（位于***风险模拟 | 工具 | 诊断工具***），你将可以获得以下分析：
  - 异方差性
  - 多重共线性
  - 微数缺测性
  - 非线形
  - 异常数据
  - 自相关性
  - 偏自相关性
  - 分布滞后性
  - 正态性和球性
  - 非平稳性
  - 随机特性
  - 线形和非线形相关
  - 方差膨胀因子
  - 可视图表

  在启动任何类型的预测或者数据分析的时候，这些测试都是很重要的。每个测试包括一个易于理解的详细报告，所以不需要找一位高级计量经济学家或者统计学家对结果进行阐述。
- **最大似然模型：**可以从以下路径（***风险模拟 | 预测 | 最大似然估计***）开启这个功能，这里，最大似然估计的迭代和内部优化过程可用于对二元回应变量进行建模（因变量是一个二元值，取 0 或 1）。这是一个重要具有多种用途的判别分析法（例如，给定一些特性如年龄，吸烟数量，血压来判别病人是否患有癌症；或者给定公司的资产，资产波动率，或者个人的年龄，教育水平，工作年限等来确定信用限额或者个人是否会违约）
- **多语言支持：**软件可支持多种语言，包括英语（美国），中文（简体），西班牙文，和日语，即将发布的版本还包括更多的语言支持。用户使用软件的时候，如果需要切换语言，可以通过简单地点击**风险模拟**和**语言菜单**图标，然后重启 Excel 即可。
- **Microsoft .NET Framework 2.0/3.0：**我们已经完全升级我们的源代码来使我们的软件与 Microsoft .NET Framework 2.0/3.0 完美地结合起来，这将使软件可以更快速地运行并且更好地与较新的电脑兼容。

RS 5.2 新功能

- 在预测方法和分析工具中添加了高速引擎。同样的分析结果运行的更快速。
- 支持多个 Excel 版本。对于加载的 Excel 2007 和 Excel 2003/XP，可以运行 Risk Simulator 在任何 Excel 版本在一台电脑中。软件中具有 Excel 转化工具可以让你决定 使用哪个 Excel 启动 Risk Simulator。
- 注释单元格。可以在 Risk Simulator 开启或者关闭单元格注释，选项菜单，以决定是否在所有的单元格注释中显示输入假设，输出预测以及决策变量。
- 右键菜单。就可以得到所有的 Risk Simulator 的工具菜单。
- 多个计量经济学模型。可以输入一个或者多个函数公式运行基本的计量经济学模型。
- 以及一次自动生成预先设置好的多个个模型。
- 高速仿真可以运行在动态和随机优化中。只需点击高级选项的按钮选择高速仿真运行优化。
- 微小的更新
    - o 动态的敏感性分析图可以获得单元格名称和地址。
    - o 动态和随机优化的规则可以支持条件均值和半标准差用于 CVaR 计算
    - o 更新了案例可以进行有效前沿和多个计量经济学模型的展示
    - o 更新了日语用户手册

- 在西班牙语报告和用户界面中更新了语言的编辑和纠正。

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？

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。

开始学习 Risk Simulator

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。

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

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**。

1. 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

1. 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**）。
2. 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）。

3. 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。
4. 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）。
5. 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；或是当与别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。

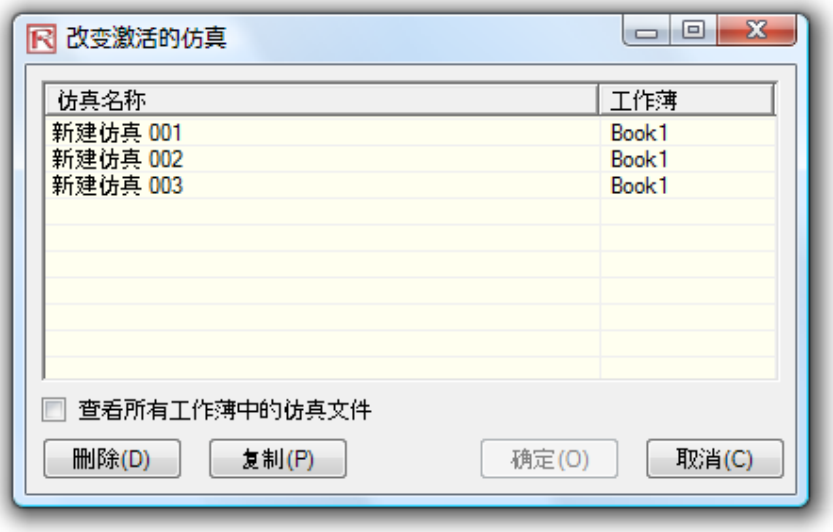


图 2.2 ——修改当前仿真

2. 定义输入假设

接下来这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）

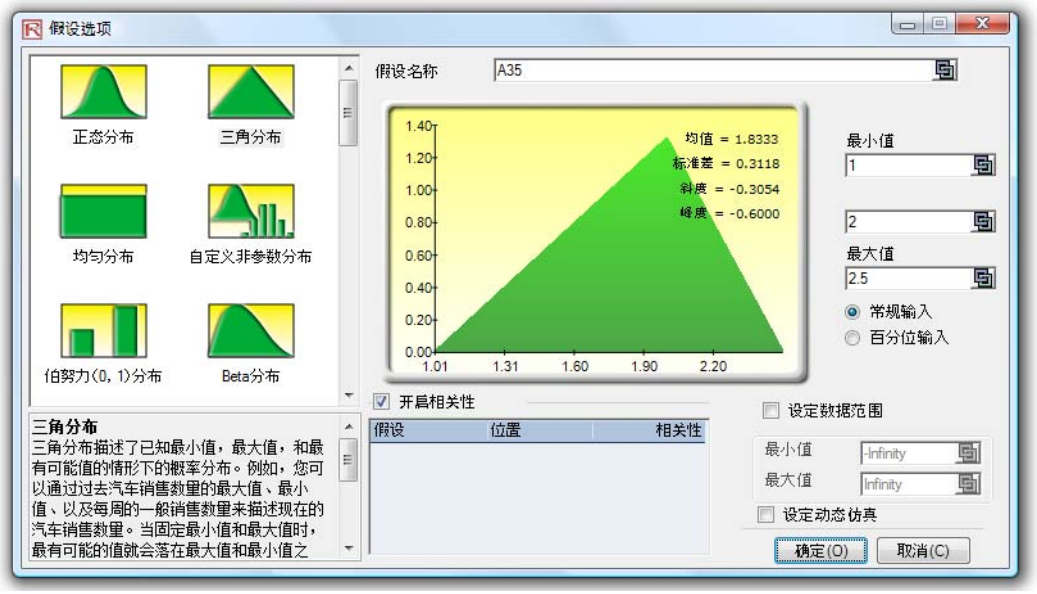


图 2.3 设定输入假设

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：**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。
- **分布图库：**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。
- **输入参数：**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。
- **分布边界：**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。
- **相关性：**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**仿真|编辑仿真**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。
- **简短描述：**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*了解 Monte Carlo 仿真中的概率分布*章节关于软件中可用的分布类型的详细介绍。

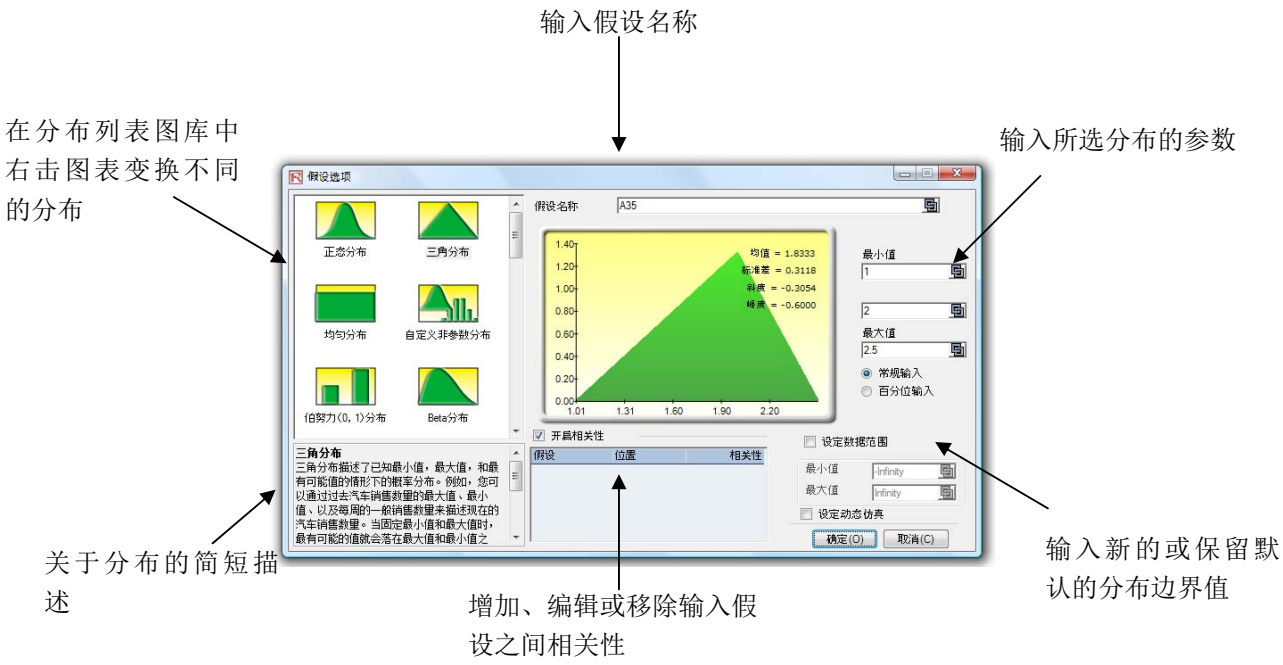


图 2.4 假设属性

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，在下个步骤中定义输出预测。

3. 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。

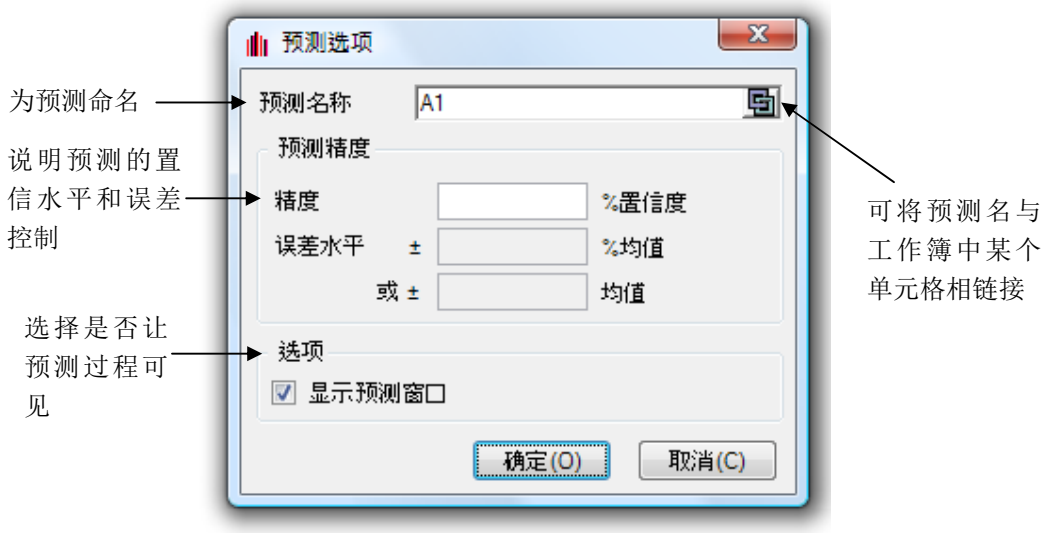


图 2.5　输出预测设定

4. 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）

5. 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。

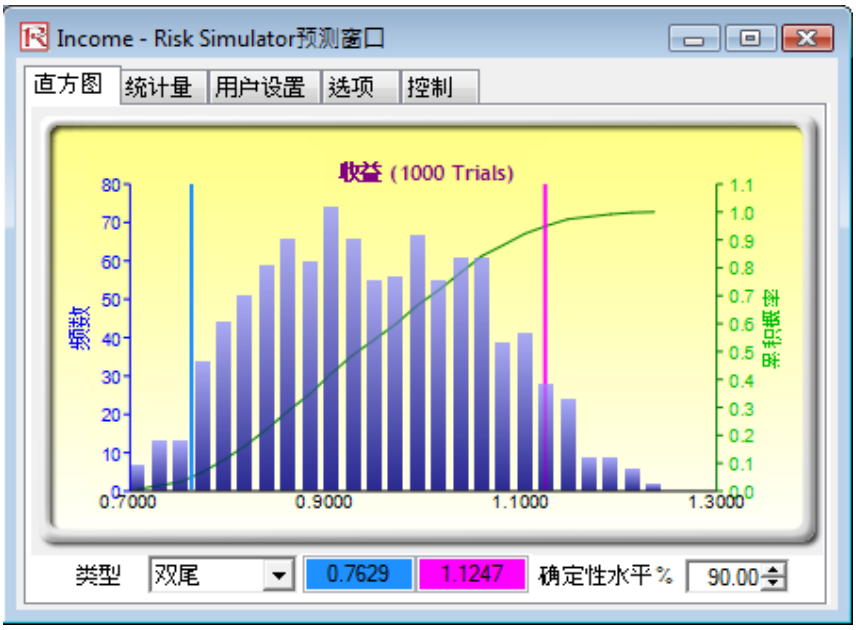


图 2.6　预测图

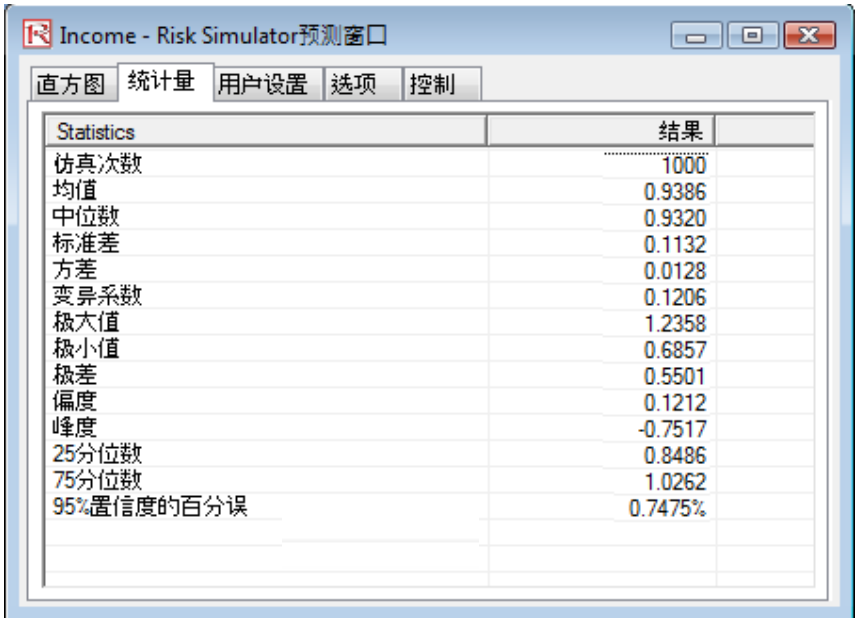


图 2.7　预测统计表

- 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。

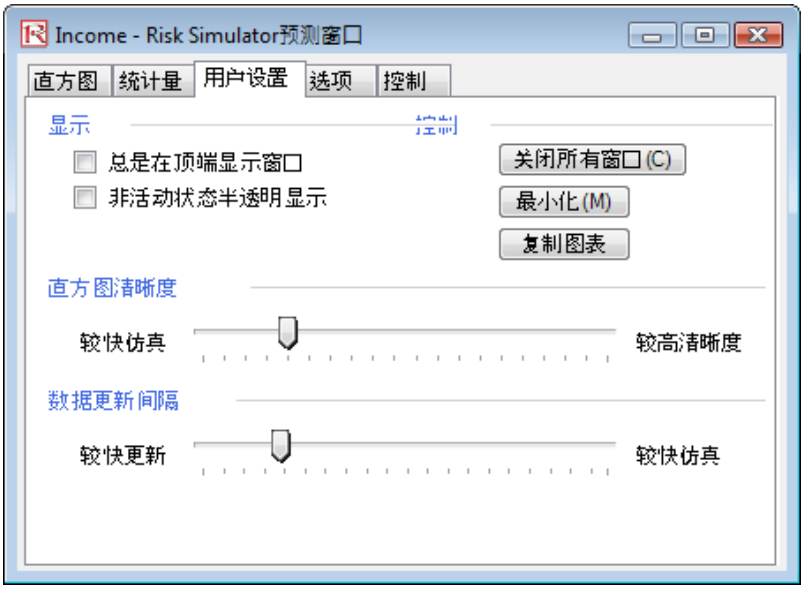


图 2.8 预测图参数选择

- 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。

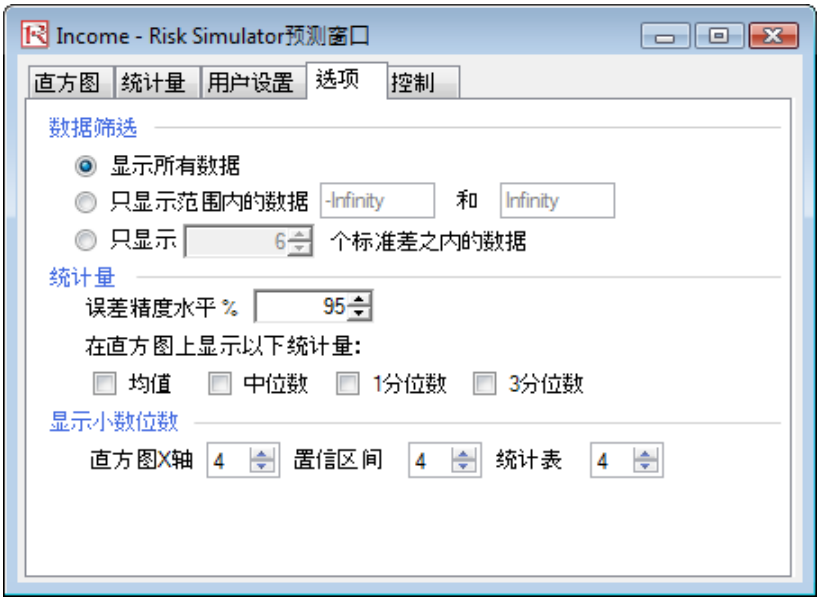


图 2.9 预测图表选项

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 健，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。

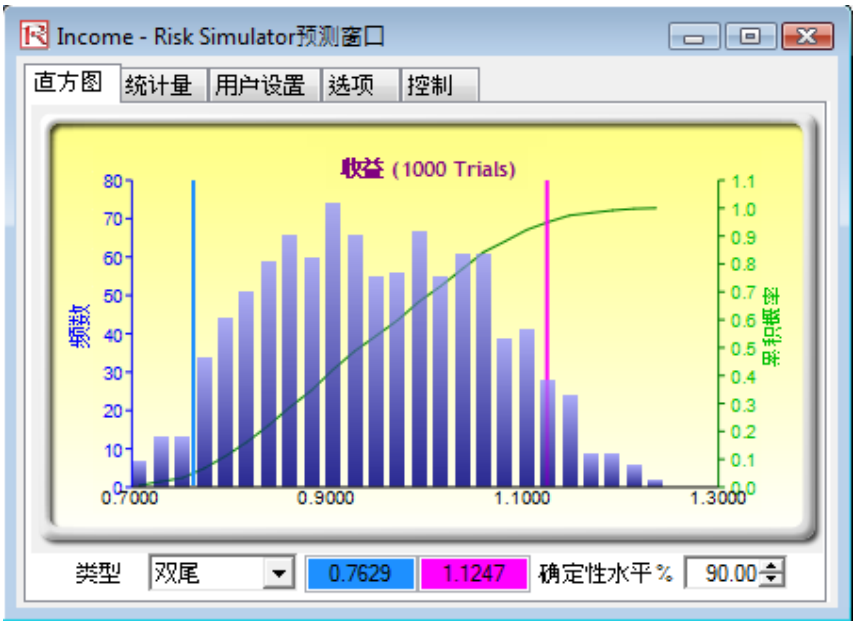


图 2.11 单尾置信区间预测图

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%）。

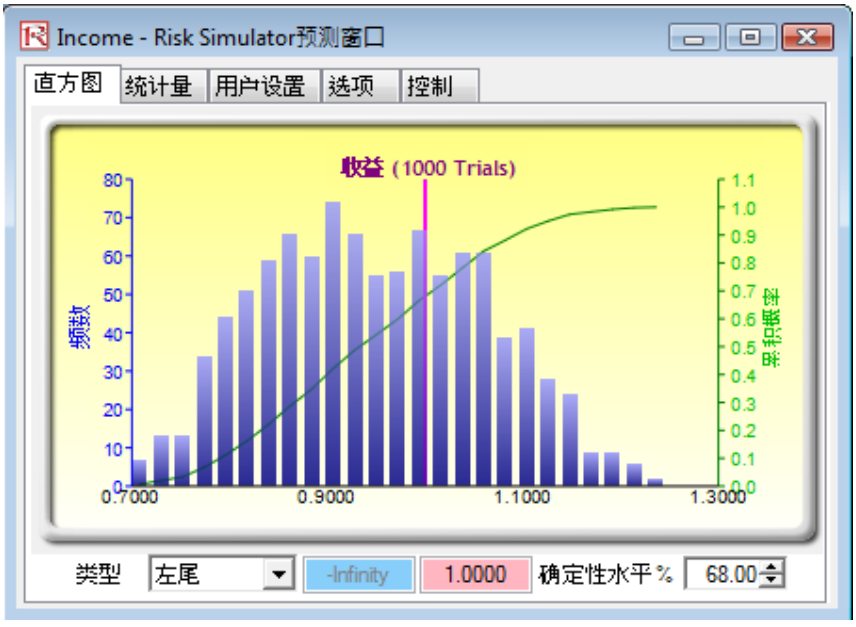


图 2.12 预测图概率估计

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%）。

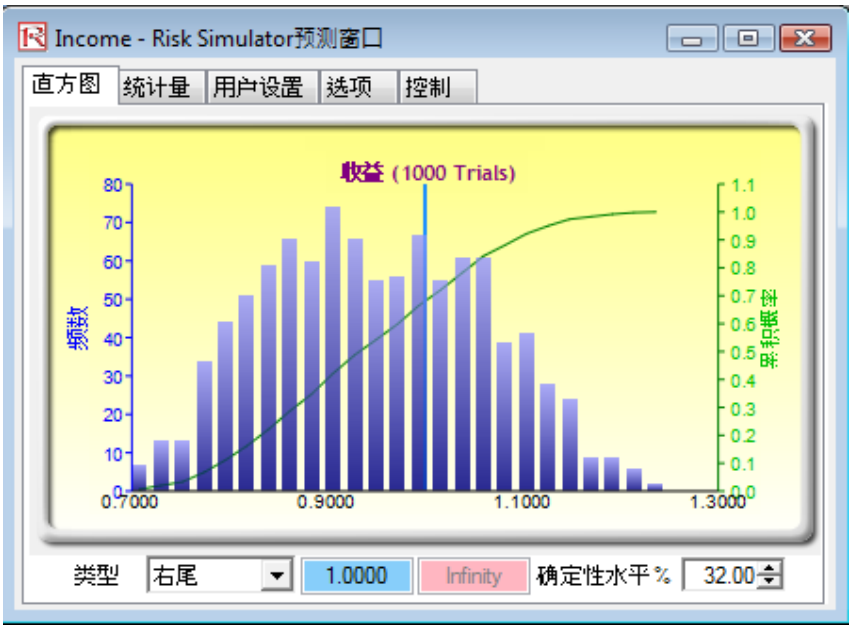


图 2.13 预测图概率估计

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。

相关性和精度控制

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：

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

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。

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距阵网格中。
- 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|编辑相关系性**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）。

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

Price Negative Correlation	*Quantity Negative Correlation*	*Correlation*	*Price Positive Correlation*	*Quantity Positive Correlation*	*Correlation*
676	145	**-0.90**	102	158	**0.89**
368	452		461	515	
264	880		515	477	
235	877		874	833	
122	711		769	792	
490	641		481	471	
336	638		627	446	
495	383		82	190	
241	568		659	674	
651	571		188	286	
854	59		458	439	
66	950		981	972	
707	262		528	569	
943	186		865	812	

图 2.16 相关系数恢复

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

对预测统计量的理解

大多数分布都可以通过四个距的计算来描述。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（μ）或平均数来测量的。

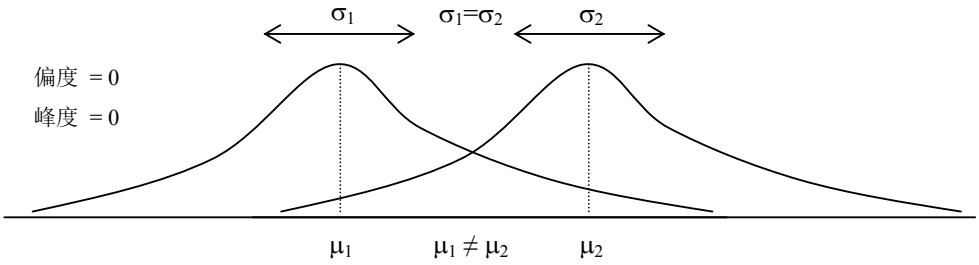


图 2.19——第一矩

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。

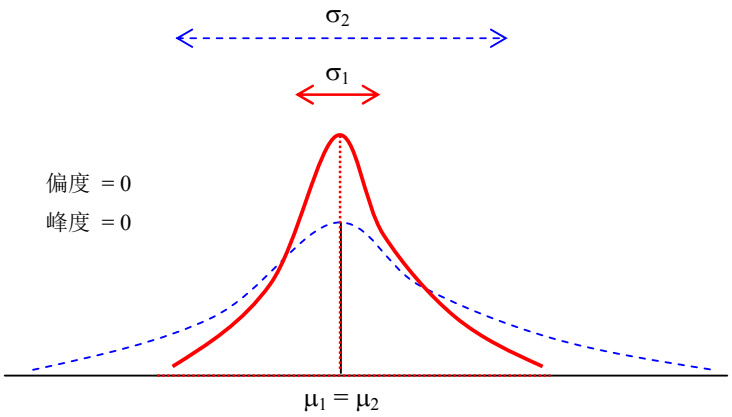


图 2.20——第二矩

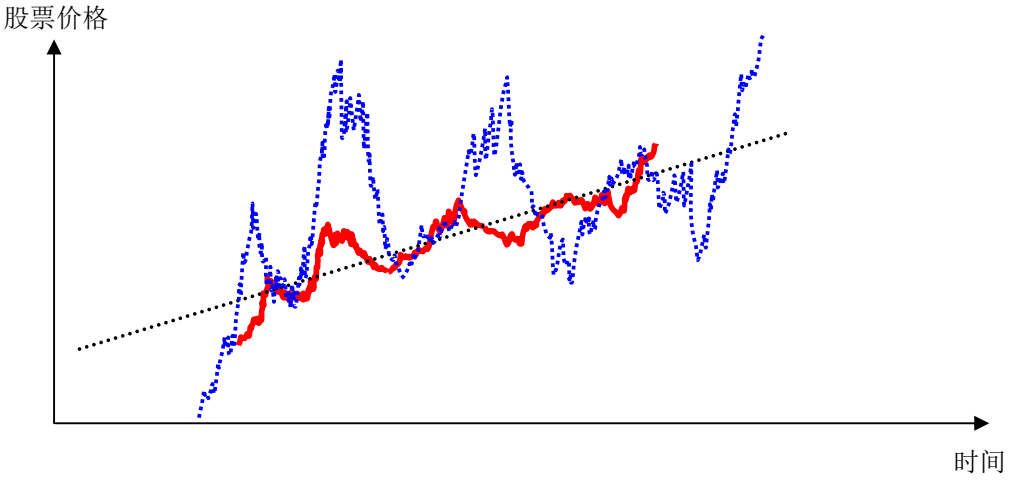


图 2.21 股票价格波动

测量分布的偏度——第三矩

分布的第三矩测量了分布的偏度，就是分布偏方和幅度。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）。

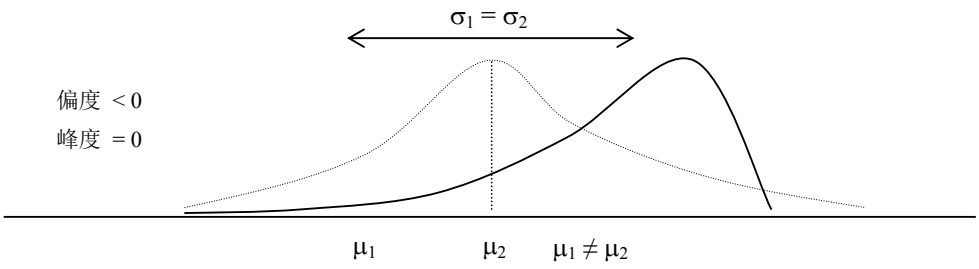


图 2.22 第三个矩（左偏）

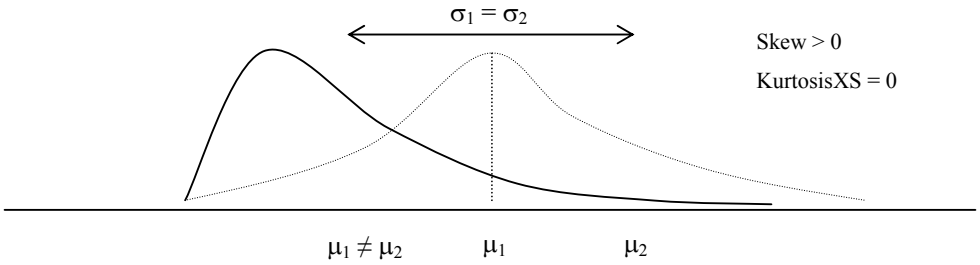


图 2.23 第三矩（右偏）

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）。

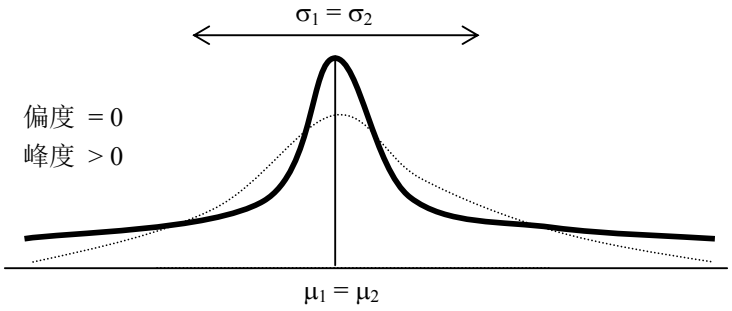


图 2.24 第四矩

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）。

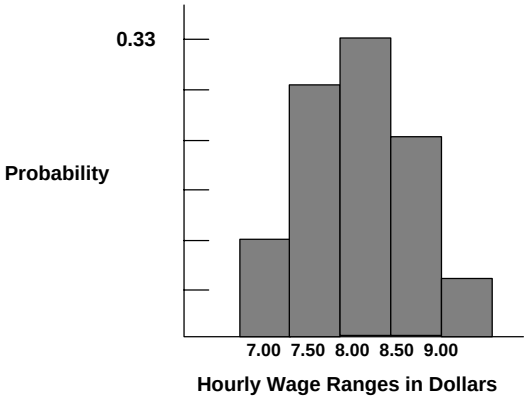


图 2.26 频数方柱图 II

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。

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

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

均值 $=p$

标准差 $=\sqrt{p(1-p)}$

偏度 $=\dfrac{1-2p}{\sqrt{p(1-p)}}$

超峰度 $=\dfrac{6p^2-6p+1}{p(1-p)}$

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≤P≤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：

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

均值 $=np$

标准差 $=\sqrt{np(1-p)}$

偏度 $=\dfrac{1-2p}{\sqrt{np(1-p)}}$

超峰度= $\dfrac{6p^{2}-6p+1}{np(1-p)}$

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≤P≤0.9999）

试验次数≥1 且≤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：

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

均值 $=\dfrac{N+1}{2}$ ×（最大值－最小值）

标准差 $=\sqrt{\dfrac{(N-1)(N+1)}{12}}$ ×（最大值－最小值）

偏度 $=0$ (分布是完全对称的)

超峰度 $=\dfrac{-6(N^{2}+1)}{5(N-1)(N+1)}$ ×（最大值－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：

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

$$\text{均值} = \frac{1}{p} - 1$$

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

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

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

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≤P≤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器零部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：

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

均值 = $\frac{N_x n}{N}$

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

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≥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：

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

方差 = $\sqrt{\frac{r(1-p)}{p^2}}$

均值 = $\frac{r(1-p)}{p}$

偏度= $\frac{2-p}{\sqrt{r(1-p)}}$

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。

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：

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

均值 $=\lambda$

标准差 = $\sqrt{\lambda}$

偏度= $\frac{1}{\sqrt{\lambda}}$

超峰度 = $\frac{1}{\lambda}$

唯一的分布参数是比率或（λ）。

参数输入要求：

比率>0，≤1000（也就是，0.0001≤比率≤1000）

连续分布

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：

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

均值 $=\frac{\alpha}{\alpha+\beta}$

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

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

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 α 和 β，Γ 是 γ 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：

α 和 β >0，可以取任意正值

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（α）和尺度（β）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 α 可以取任意值

尺度参数 β 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：

$$Z_1^2 + Z_2^2 + ... + Z_k^2 \overset{d}{\sim} \chi_k^2$$

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 $= \sqrt{2k}$

偏度 $= 2\sqrt{\frac{2}{k}}$

超峰度 $= \frac{12}{k}$

Γ 代表 ɣ 函数。自由度 k 是唯一的分布参数。

ɣ 分布可以通过一定的设置变为卡方分布：

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

参数输入要求：

自由度为大于 1 小于 300 的整数

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：

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

均值 $= \frac{1}{\lambda}$

标准差= $\frac{1}{\lambda}$

偏度=2（这个值适用于所有成功率 λ）

超峰度=6（这个值适用于所有成功率 λ）

成功率（λ）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：

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

均值 = $\alpha+0.577215\beta$

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

偏度= $\frac{12\sqrt{6}(1.2020569)}{\pi^3}=1.13955$（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（α）和尺度（β）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 α 可以取任意值

尺度 β >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：

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

均值 = $\dfrac{m}{m-2}$

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

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

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

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的三个条件有：

- 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：

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

均值 = $\alpha\beta$

标准差 = $\sqrt{\alpha\beta^2}$

偏度 = $\dfrac{2}{\sqrt{\alpha}}$

超峰度 = $\dfrac{6}{\alpha}$

形状参数（α）和尺度参数（β）是分布参数，Γ是 Gamma 函数。

参数α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：

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

参数输入要求：

尺度参数β可以为大于 0 的任意数

形状参数α可以为大于等于 0.5 任意数

位置数可以取任意值

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：

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

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

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

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

$$超峰度\ =\ \exp\left(4\sigma^2\right)+2\exp\left(3\sigma^2\right)+3\exp\left(2\sigma^2\right)-6$$

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：

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

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

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

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

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

均值（μ）和标准差（σ）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：

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

$$\text{均值}=\mu$$

$$\text{标准差}=\sigma$$

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（μ）和标准差（σ）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

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（α）和位置（β）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≥0.05

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）

标准差 = $\sqrt{\frac{r}{r-2}}$

偏度 = 0 (适用于自由度为r的所有值)

超峰度 = $\frac{6}{r-4}$, $r>4$

$t=\frac{x-\bar{x}}{s}$, Γ 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：

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

均值= $\frac{1}{3}(Min+Likely+Max)$

标准差 = $\sqrt{\frac{1}{18}(Min^2+Likely^2+Max^2-Min\,Max-Min\,Likely-Max\,Likely)}$

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

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：

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

$$均值 = \frac{Min + Max}{2}$$

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

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：

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

均值 $=\beta\Gamma(1+\alpha^{-1})$

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

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

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（α）和中心位置尺度（β）是分布的两个参数，Γ是γ函数。

参数输入要求：

形状参数α≥0.05

尺度参数β可以为大于0的任意正值

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。

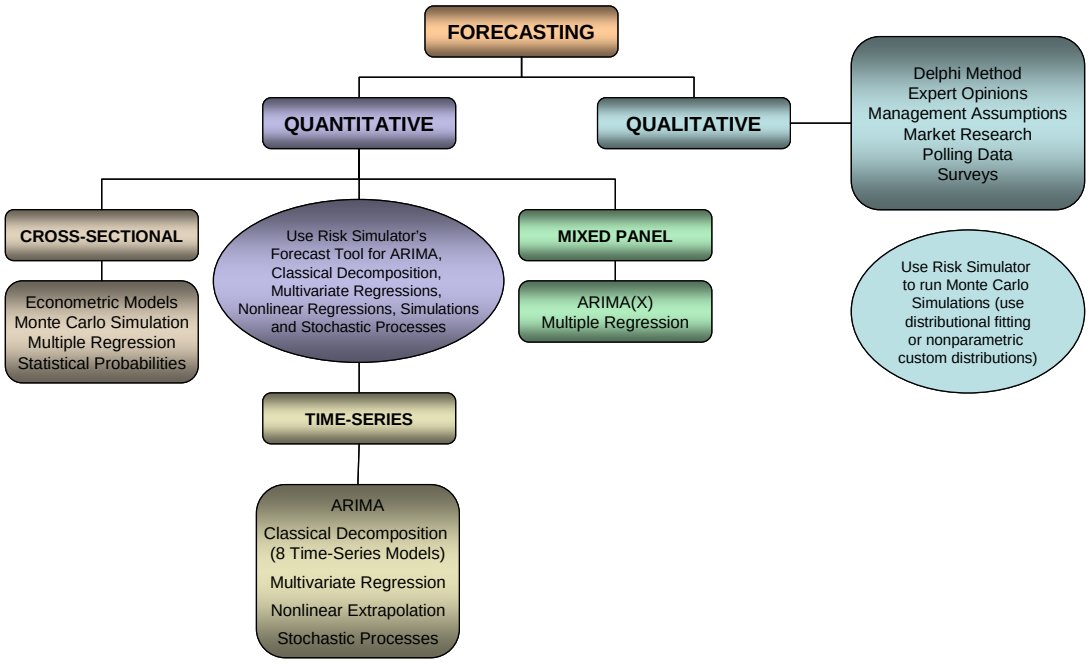


图 3.1 预测方法

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，或内部成员达成共识的预测）、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（自回归求和移动平均）
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. 三次样条插值法
6. GARCH （广义自回归条件异方差）
7. J-曲线
8. 马尔可夫链
9. 最大似然法
10. 非线形外推
11. 回归分析
12. 随机过程
13. 时间序列分析
14. 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。

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**。

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**

历史销售收入

年份	季度	时间点	销售额
2000	1	1	$684.20
2000	2	2	$584.10
2000	3	3	$765.40
2000	4	4	$892.30
2001	1	5	$885.40
2001	2	6	$677.00
2001	3	7	$1,006.60
2001	4	8	$1,122.10
2002	1	9	$1,163.40
2002	2	10	$993.20
2002	3	11	$1,312.50
2002	4	12	$1,545.30
2003	1	13	$1,596.20
2003	2	14	$1,260.40
2003	3	15	$1,735.20
2003	4	16	$2,029.70
2004	1	17	$2,107.80
2004	2	18	$1,650.30
2004	3	19	$2,304.40
2004	4	20	$2,639.40

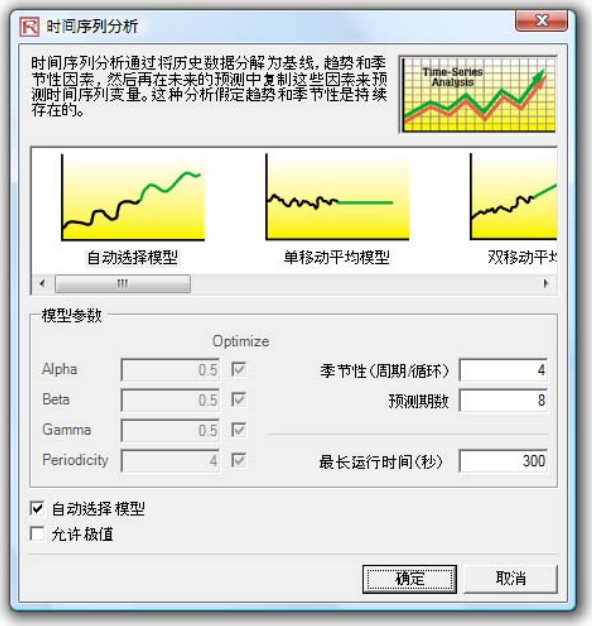


图 3.4 时间序列分析

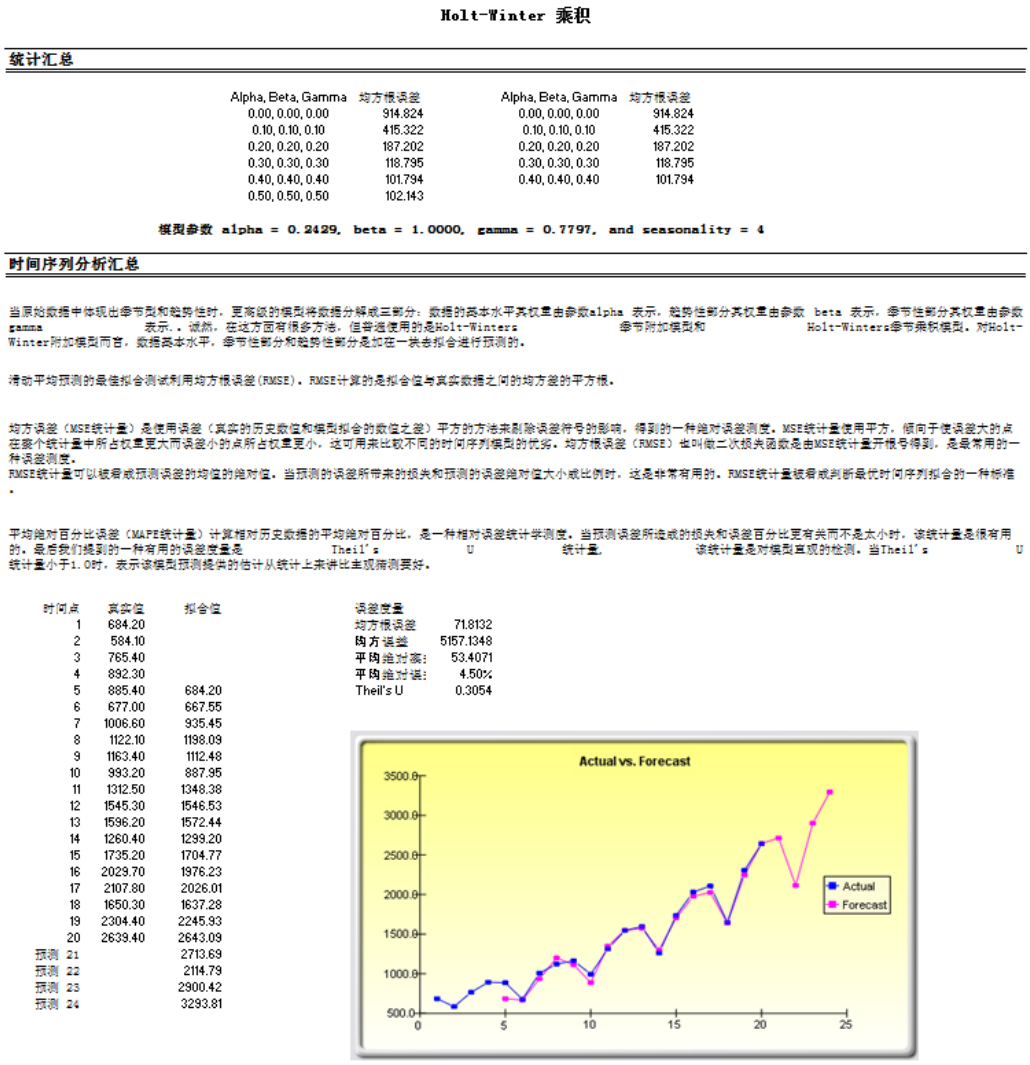
结果解析：

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 α，β，γ 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。α 针对的是基准水平随时间变化的记忆影响，β 是趋势参数，衡量了趋势的强度，γ 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 α，β，γ 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》，Wiley 2006。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 α， β， γ 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		

模型参数 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型和趋势性时，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季节乘积模型。对Holt-Winter附加模型而言，数据基本水平，季节性部分和趋势性部分是加在一块去拟合进行预测的。

滑动平均预测的最佳拟合测试利用均方根误差(RMSE)。RMSE计算的是拟合值与真实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（真实的历史数值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测度。MSE统计量使用平方，倾向于使误差大的点在整个统计量中所占权重更大而误差小的点所占权重更小，这可用来比较不同的时间序列模型的优劣。均方根误差（RMSE）也叫做二次损失函数是由MSE统计量开根号得到，是最常用的一种误差测度。
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统计量被看成判断最优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学测度。当预测误差所造成的损失和误差百分比更有关而不是太小时，该统计量是很有用的。最后我们提到的一种有用的误差度量是 Theil's U 统计量。 该统计量是对模型互现的检测。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主观猜测要好。

时间点	真实值	拟合值
1	684.20	
2	584.10	
3	765.40	
4	892.30	
5	885.40	684.20
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	2107.80	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

误差度量	
均方根误差	71.8132
均方误差	5157.1348
平均绝对误:	53.4071
平均绝对误:	4.50%
Theil's U	0.3054

图 3.5 Holt-Winter 预测报告示例

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：$Y=\beta_0+\beta_1X+\varepsilon$，其中 β_0 代表截距，β_1 代表斜率，ε 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 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\ldots+\beta_nX_n+\varepsilon$ 。在这个例子中，最佳拟合线就位于一个 n+1 维的空间中了。

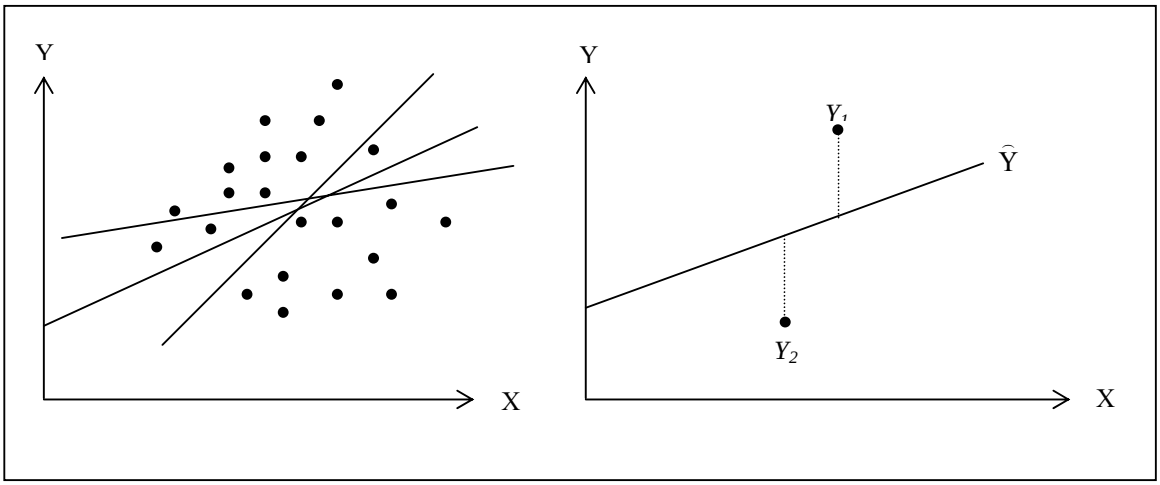


图 3.6 双变量回归

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

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

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：

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

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》（Wiley 2006）。

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》（Wiley 2006）。

多元回归分析数据

暴力袭击事件	学士学位	人均警察支出	人口数（百万）	人口密度（人/平方米）	失业率
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	12683	288	2.842	29.6	6
222	62184	229	11.992	265.1	6.9
280	9153	287	1.003	960.3	8.5
759	14250	224	3.487	115.8	6.2
114	3680	161	0.696	9.2	3.4
419	18063	221	4.877	118.3	6.6
435	65112	237	16.987	64.9	6.6
186	11340	220	1.723	21	4.9
87	4553	185	0.563	60.8	6.4
188	28960	260	6.187	156.3	5.8
303	19201	261	4.867	73.1	6.3
102	7533	116	1.793	74.5	10.5
127	26343	268	4.892	90.1	5.4
251	1641	300	0.454	4.7	5.1

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	人口数（百万）	（人/平方米）	失业率
系数	57.9555	-0.0035	0.4644	25.2377	-0.0086	16.5579
标准差	108.7901	0.0035	0.2535	14.1172	0.1016	14.7996
t统计量	0.5327	-1.0066	1.8316	1.7877	-0.0843	1.1188
p值	0.5969	0.3197	0.0738	0.0807	0.9332	0.2693
下限5%	-161.2966	-0.0106	-0.0466	-3.2137	-0.2132	-13.2687
上限95%	277.2076	0.0036	0.9753	53.6891	0.1961	46.3845

自由度		假设检验	
回归自由度	5	临界t值（99%的致信度，自由度为44）	2.6923
残差自由度	44	临界t值（95%的致信度，自由度为44）	2.0154
总自由度	49	临界t值（90%的致信度，自由度为44）	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 = b0 + b1X1 + b2X2 + ... + bnXn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。

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验，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。如果平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)

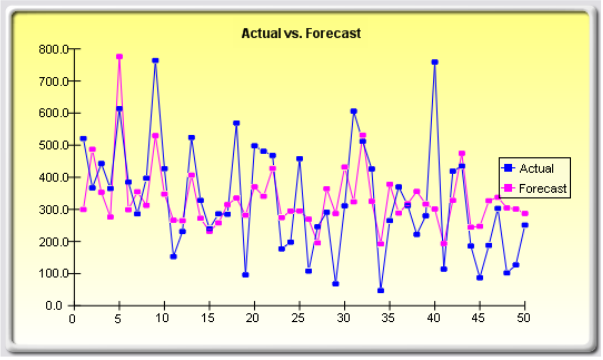


图 3.8 多元回归的结果

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: 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 2006 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。

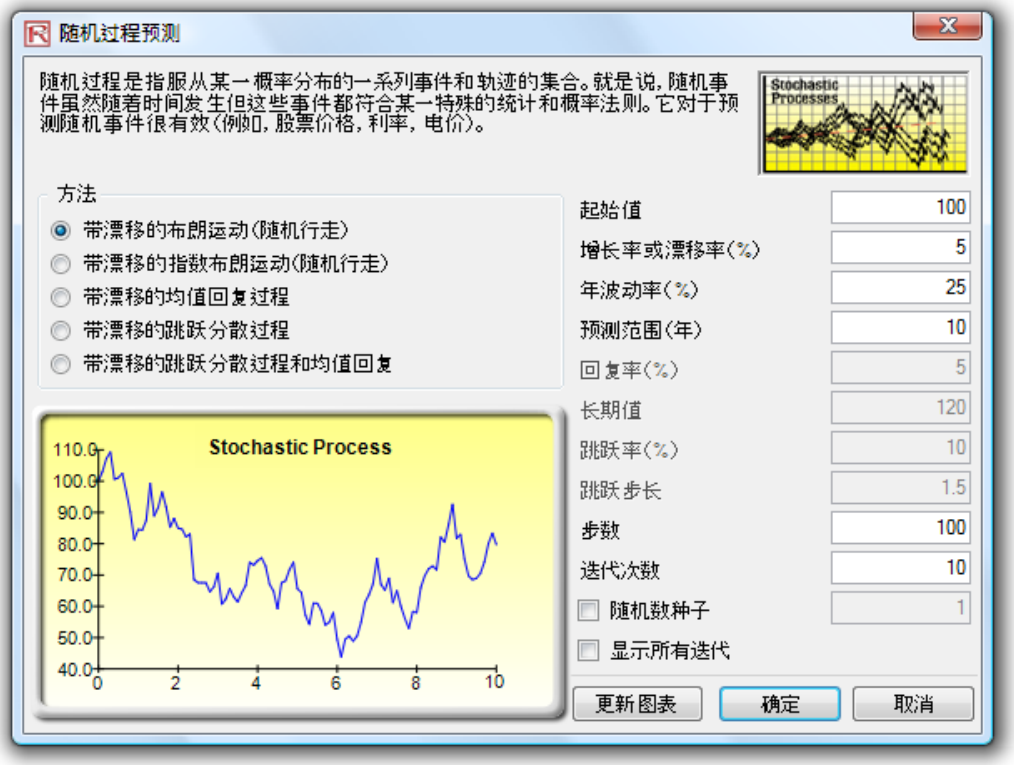


图 3.9 随机过程预测

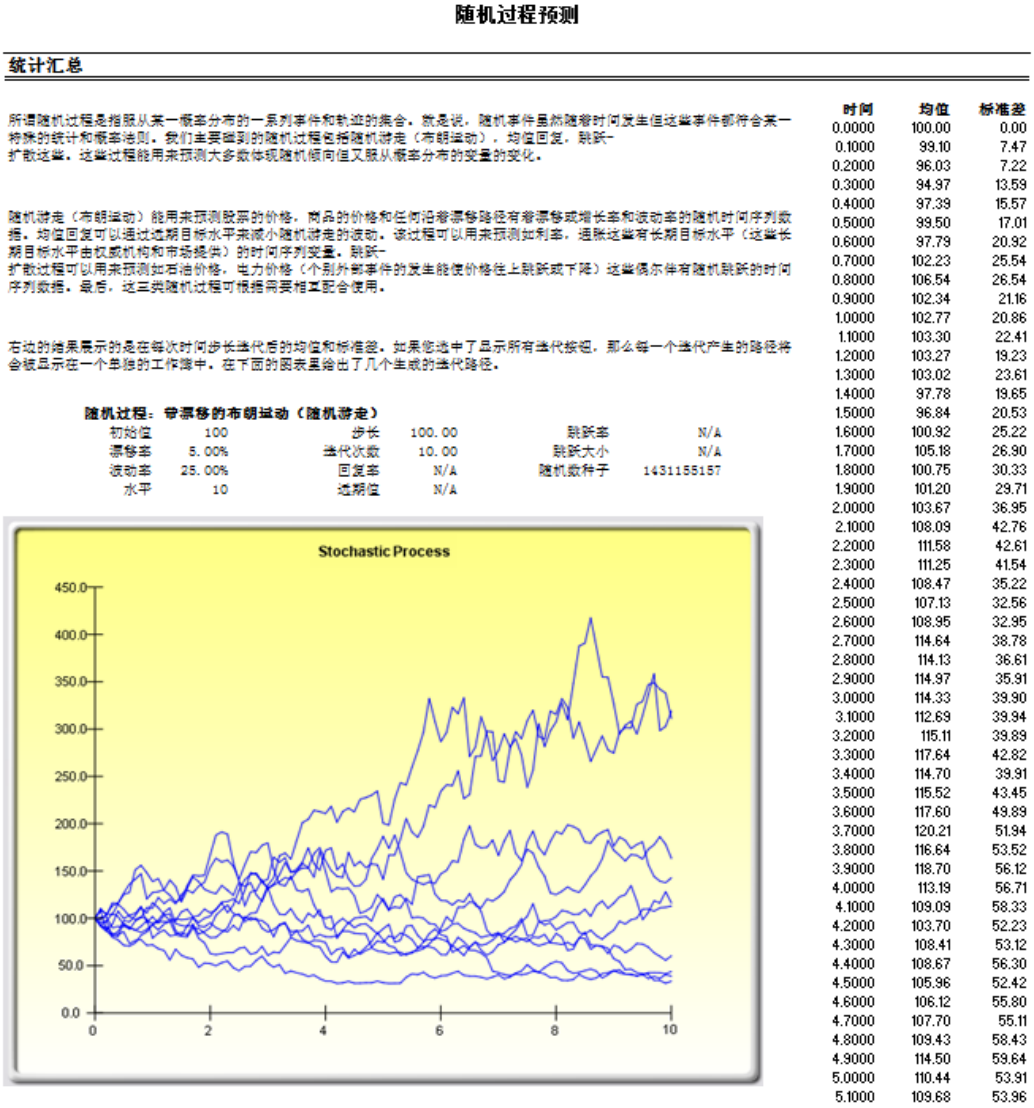


图 3.10 随机预测结果

非线性外推法

理论：

外推法是通过对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

年份	月份	时期	销售
2004	1	1	$1.00
2004	2	2	$6.73
2004	3	3	$20.52
2004	4	4	$45.25
2004	5	5	$83.59
2004	6	6	$138.01
2004	7	7	$210.87
2004	8	8	$304.44
2004	9	9	$420.89
2004	10	10	$562.34
2004	11	11	$730.85
2004	12	12	$928.43

历史净收益
正弦增长率

年份	月份	时期	收益
2004	1	1	$84.15
2004	2	2	$90.93
2004	3	3	$14.11
2004	4	4	($75.68)
2004	5	5	($95.89)
2004	6	6	($27.94)
2004	7	7	$65.70
2004	8	8	$98.94
2004	9	9	$41.21
2004	10	10	($54.40)
2004	11	11	($100.00)
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)

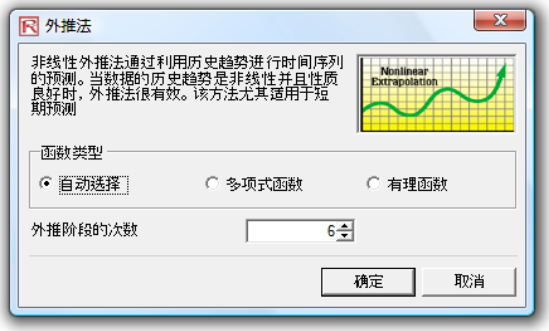


图 3.11 运行非线性外推

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，多变量回归法是更为恰当的。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，相对简单的以及耗费小的。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）。

时间点	真实值	预测拟合值	估计误差
1	1.00		
2	6.73	1.00	
3	20.52	-1.42	-8.15
4	45.25	99.82	119.36
5	83.59	55.92	-46.67
6	138.01	136.71	14.39
7	210.87	211.96	1.69
8	304.44	304.43	-0.41
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

误差度量	
均方根误差	19.6799
均方误差	387.2974
平均绝对离:	10.2095
平均绝对误:	31.56%
Theil's U	1.1210

函数类型：有理

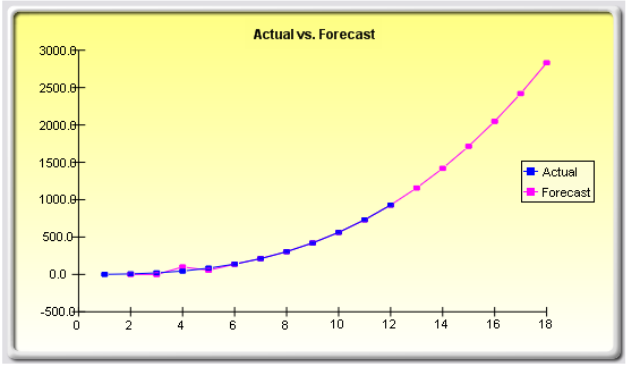


图 3.12 非线性外推的结果

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合成一种综合性模型。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：$y_t = a_1 y_{t-1} + ... + a_p y_{t-p} + e_t$。第二个组成成分是求和法（I）。每一种求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：$y_t = e_t + b_1 e_{t-1} + ... + b_q e_{t-q}$，最后 ARIMA（p,d,q）模型的综合形式是：
$y_t = a_1 y_{t-1} + ... + a_p y_{t-p} + e_t + b_1 e_{t-1} + ... + b_q e_{t-q}$

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。

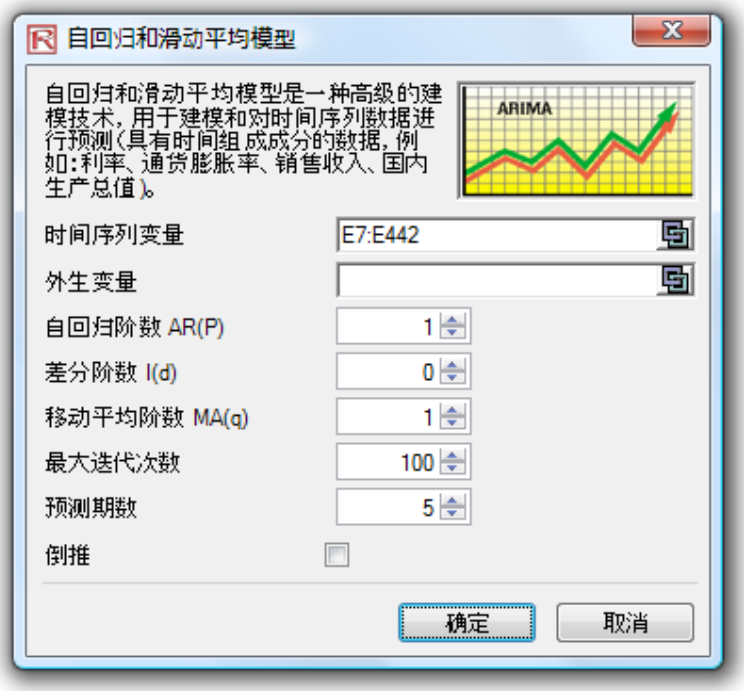


图 3.13 **Box-Jenkins ARIMA 预测工具**

ARIMA（自回归求和滑动平均）

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

自回归求和滑动平均过程简称ARIMA（p,d,q）模型扩展了AR模型，使用了三个部分来对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)+...+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。1(1)为为使用一阶差分。1(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)+...+b(q)*e(t-q)。最后，ARMA(p,q)的形式为：y(t)=a(1)*y(t-1)+...+a(p)*y(t-p)+e(t)+b(1)*e(t-1)+...+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差。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	0.0006	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		**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 = b0 + b1X1 + b2X2 + ... + bnXn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值标明其对应的系数在任何 alpha水平下都不是统计上有效的。

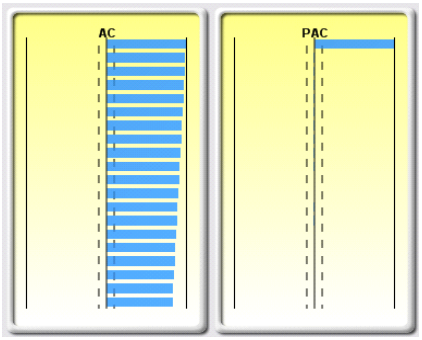
方差分析

	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，备择假设：所有系数不同时为0。如果平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的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(1)不为0,怎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 Q-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)

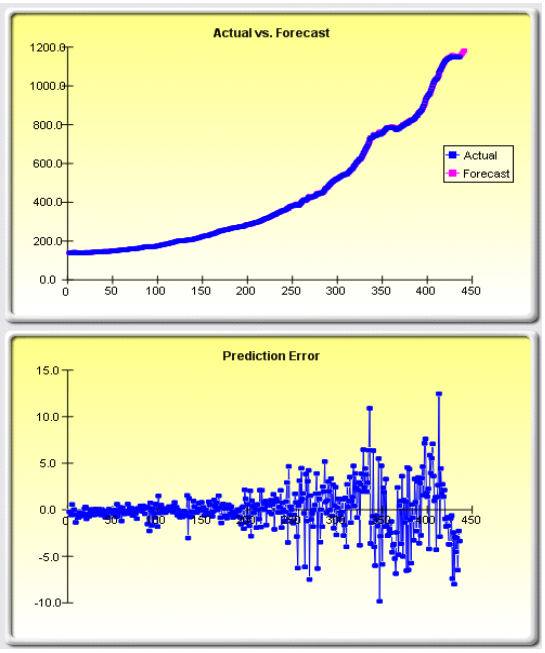


图 3.13 Box Jenkins ARIMA 预测报告

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***

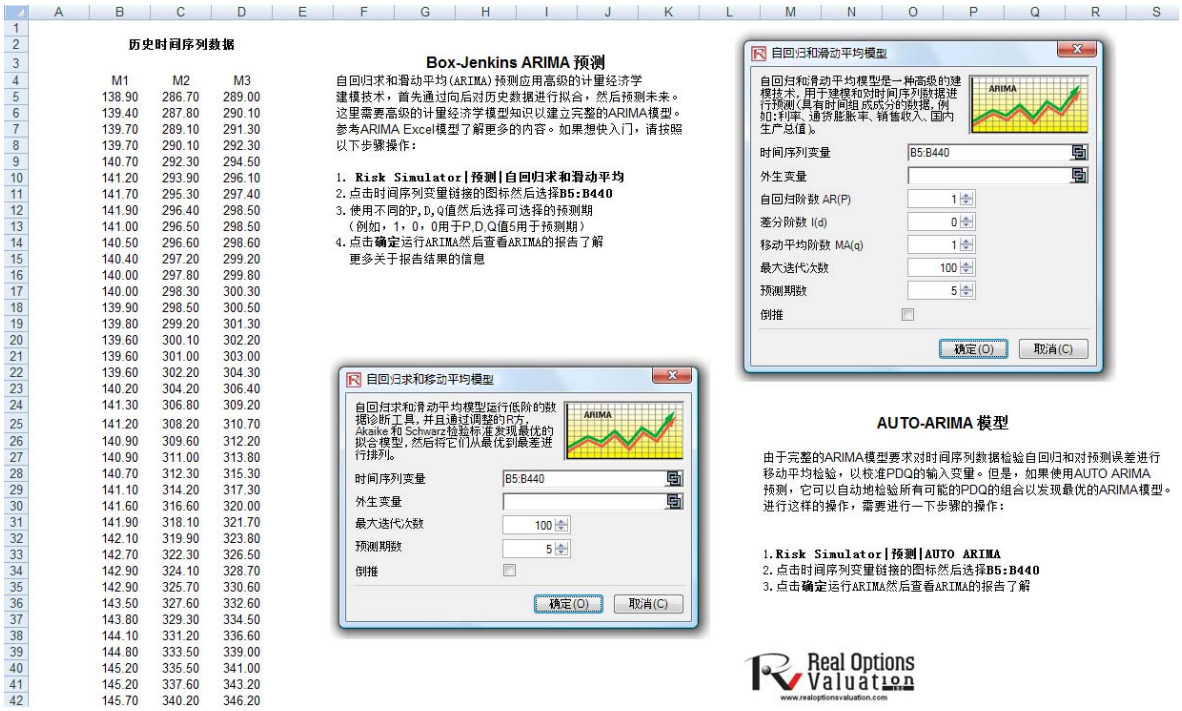


图 3.14 自回归求和移动平均模型

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。

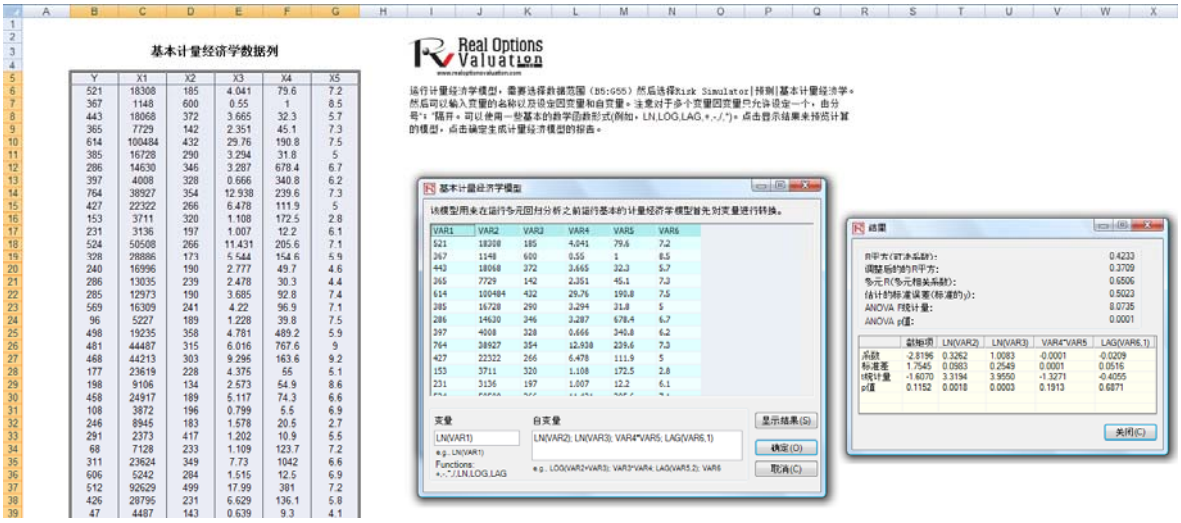


图 3.15 基本计量经济学模块

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。指数增长的背后的原理是，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

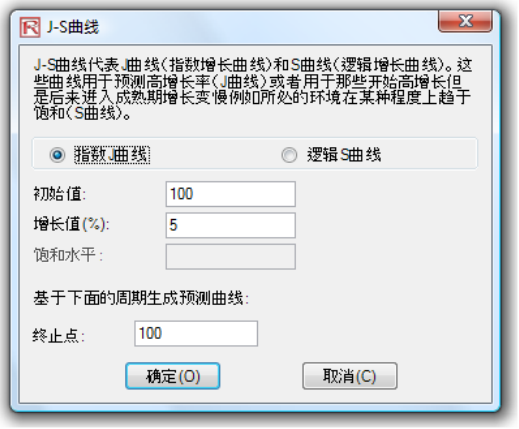


图 3.16 J-曲线预测

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，受精软分裂，细胞数快熟增长：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。

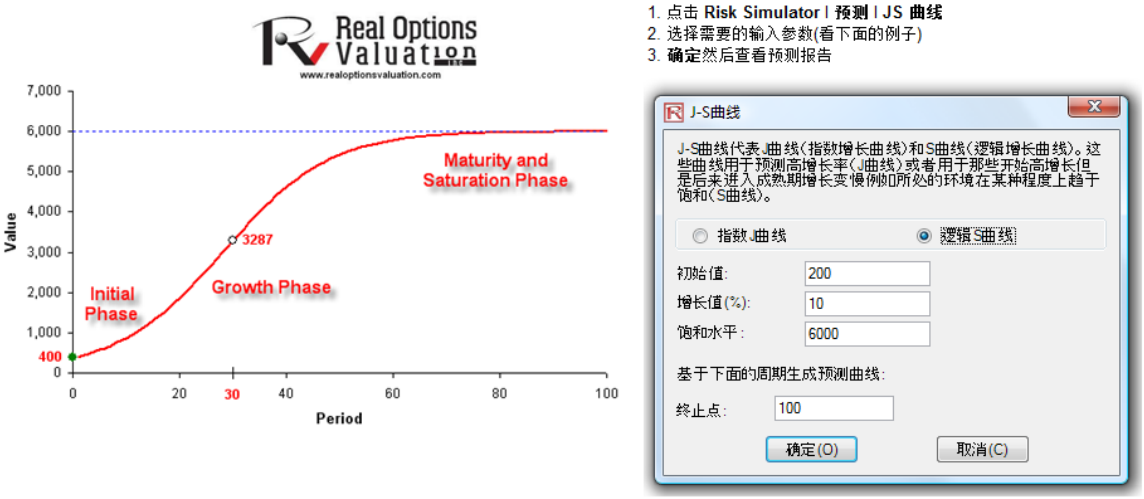


图 3.17 S-曲线预测

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。

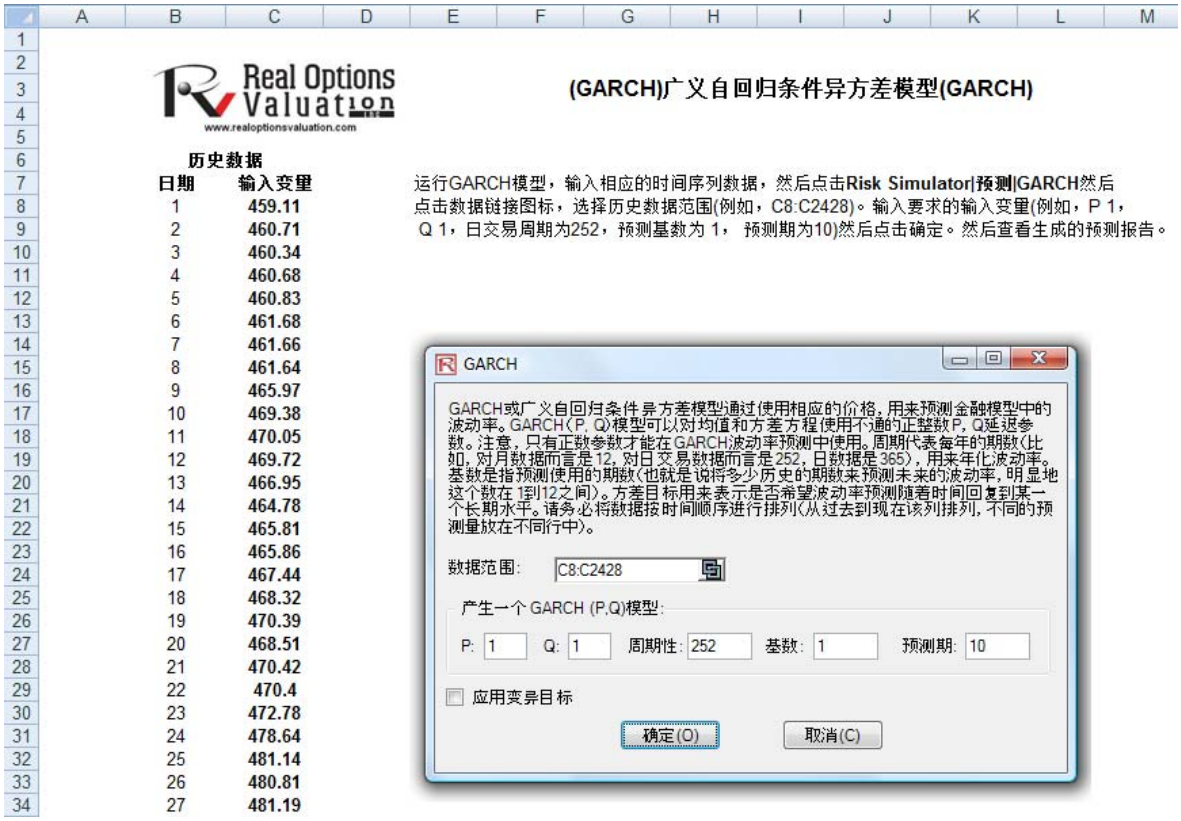


图 3.18 GARCH 波动率预测

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。需要输入的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。

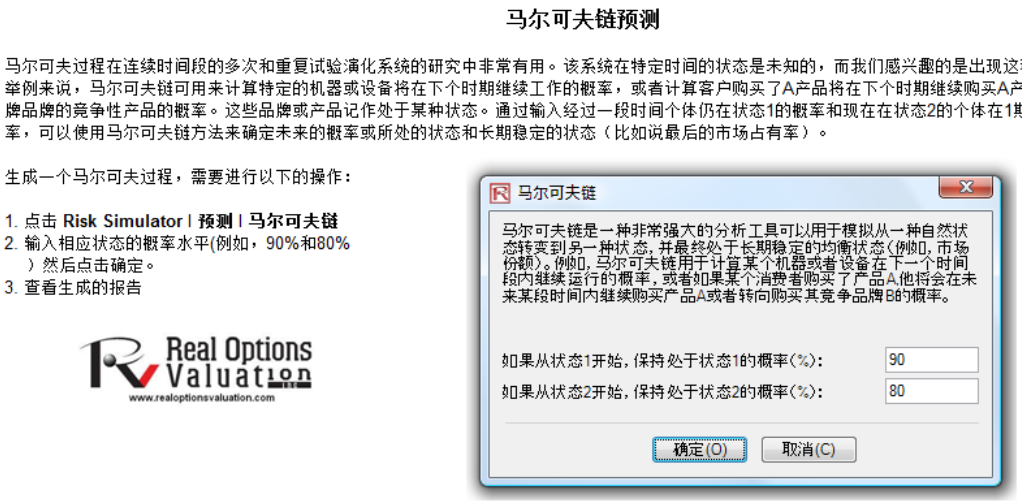


图 3.19 马尔可夫链（转换体系）

最大似然估计模型（MLE）

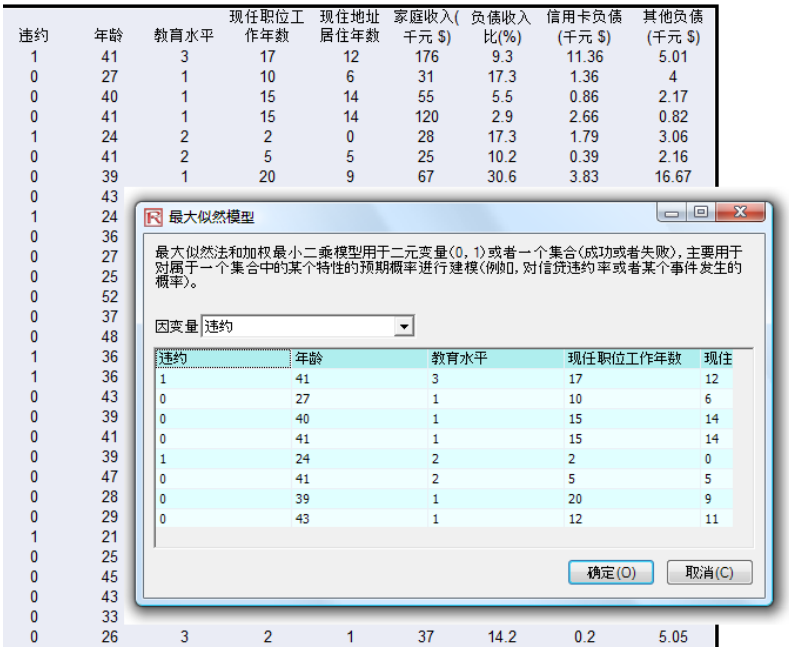
理论:

最大似然估计（MLE）模型用于在给定某些自变量的条件下，预测某些事件发生的概率。例如 MLE 用于预测在给定债务人的特征（如 30 岁，单身，年收入为 100,000 美元，信用卡负债额度为 10,000 美元）的条件下，债务人违约的概率；或者给定病人的特征，男性，50-60 岁，每个月或者每年抽 5 包香烟，等条件下患肺癌的概率。因变量的值是二元的（0 表示不具有这种特征，1 表示具有这种特征），结果估计的系数是对数奇数比率的，不能够直接用概率的形式解释。事先需要进行某个计算。方法非常简单。估计某个特征种群成功的概率（例如，在给定每年的吸烟量，吸烟者是否患有胸部并发症），使用 MLE 系数计算估计的 Y 值。例如，如果模型是 Y = 1.1 + 0.005 (Cigarrettes) 然后某人每年吸 100 包烟 Y 就可以是 1.1 + 0.005(100) = 1.6。然后，计算奇数比率的反对数：*EXP(估计的 Y)/[1 + EXP(估计的 Y)] = EXP(1.6)/(1+ EXP(0.6))* = 0.8320。所以这个人在有生之年得胸部并发症的概率为 83.20%。

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

二元最大似然逻辑预测



这里的数据显示了各种各样的贷款，信贷，或者负债问题。数据显示了每笔贷款会发生违约现象，包括贷款申请人的年龄，教育水平(1-3说明高中，大学，或者研究生教育背景)，现任职位任职年数等等。现在对现有的历史数据进行建模以发现哪些变量会对个体的违约行为发生影响，使用Risk Simulator的最大似然估计模型。结果模型会帮助银行或者信贷投放者计算具有不同特点的信贷申请者的期望的违约率。

运行该分析，选择左侧的数据或者任何其他的数据列(包括数据列的标题)然后确定数据列具有相同的长度，不会有任何的缺失或者无效的数据。然后，点击**Risk Simulator|预测|最大似然模型**。一系列的结果会显示在MLE工作表中，完成上述操作，对个体的违约率进行计算。

图 3.20 最大似然估计模型

样条模型 (三次样条内插和外插模型)

理论:

有时候在时间序列数据列中会有缺失值。例如，1 到 3 年的利率存在，下面就是 5 到 8 年的数据，和第 10 年的数据。样条曲线可以用来预测或者外推未来某个期间的数据。数据可以是线性也可以是非线性的。如图 3.21 显示了三次样条是如何运行的，图 3.22 显示了这个模块的预测报告。已知 X 值显示在图表的 x 轴（在本例中，是年份），已知的 Y 显示在 y 轴（本例中，是已知的利息率）。

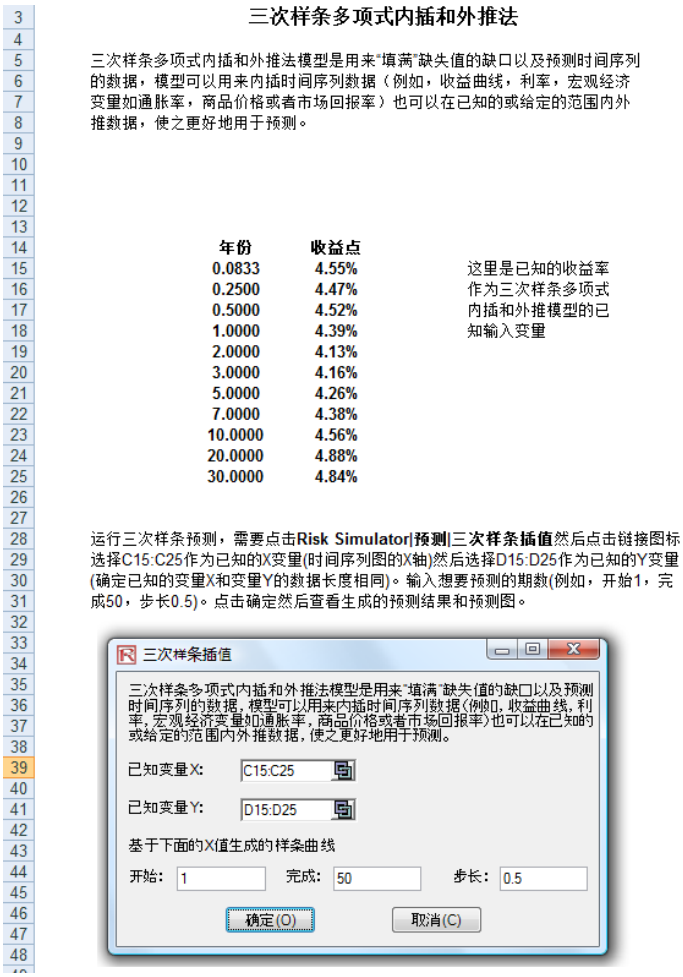


图 3.21 三次样条模块

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

Real Options Valuation
www.realoptionsvaluation.com

下面的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%

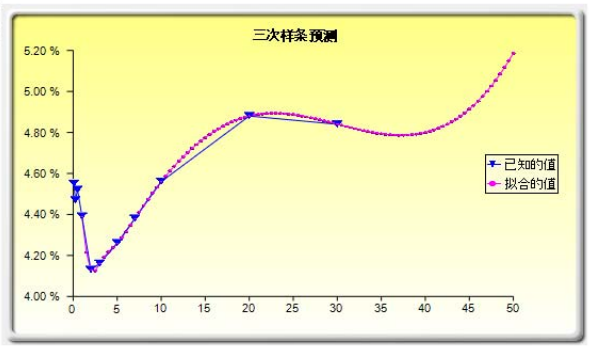


图 3.22 样条预测的结果

4．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。

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侯）。

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。

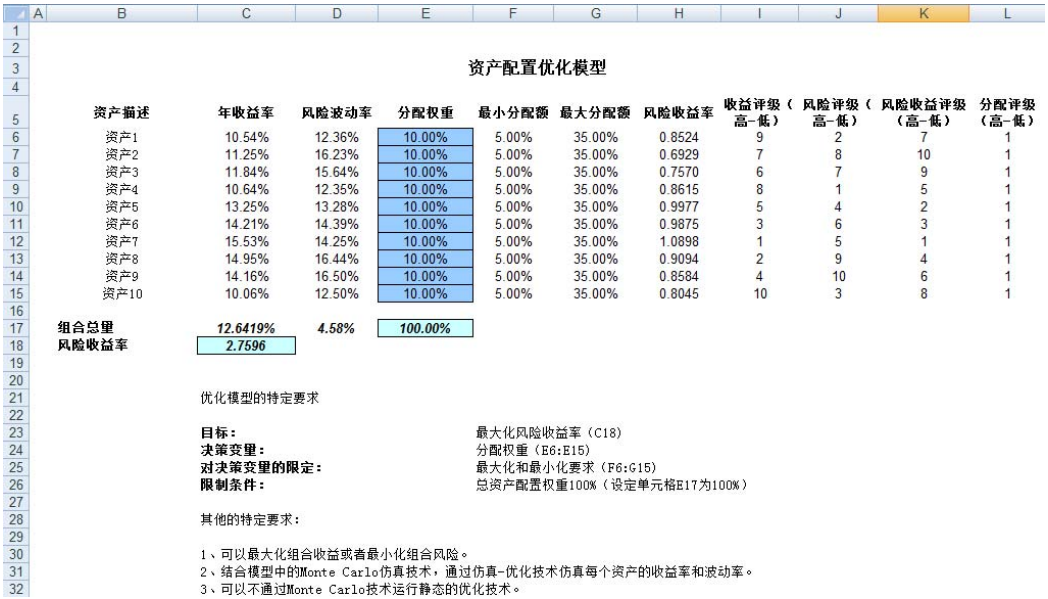
有一点值得注意。有其它一些软件在表面上可以进行随机优化，但实际上它们是不可以的。例如，它们可能是在运行一次仿真后，生成了一个优化过程的迭代，再运行另一次仿真，生成第二个优化迭代，依此类推，这样非常浪费时间和资源。在优化过程中，模型要经过一系列严格运算法则的检验，此时我们就需要运用多重迭代（范围从几次到上万次迭代）来得到最佳的结果。因此，每一次只生成一次迭代非常浪费时间和资源。与这种需要几个小时的方法相比，运用 Risk Simulator 可以在一分钟内得到同样的资产组合结果。同时这种仿真-优化方法一般会得到不好的结果，并不是一种随机的优化方法。当在我们的模型中应用优化方法时一定要非常注意这类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

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。

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
组合总里	12.6419%	4.58%	100.00%							
风险收益率	2.7596									

优化模型的特定要求

目标：	最大化风险收益率（C18）
决策变里：	分配权重（E6:E15）
对决策变里的限定：	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

E 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 Rp 代表资产总的收益率，$R_{A,B,C,D}$是项目的个体收益率，$w_{A,B,C,D}$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：

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

来计算。这里，ρ_{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。

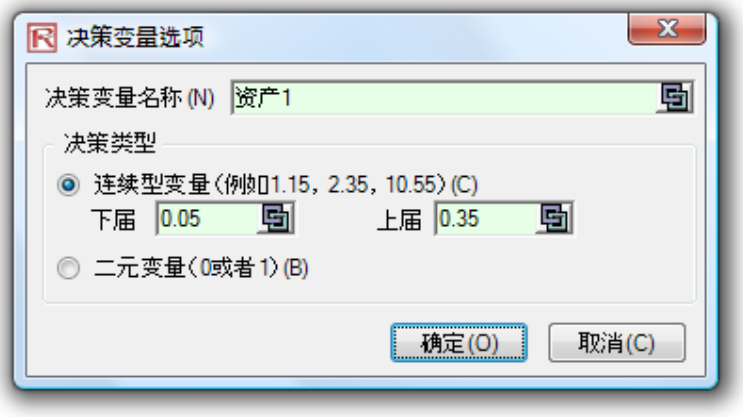
最后还要计算资产的风险收益率或夏普指数（Sharpe-Ratio）。这个值位于 C18 单元格，代表本优化模型中的最大化目标。总结一下，本模型包含如下一些属性：

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。



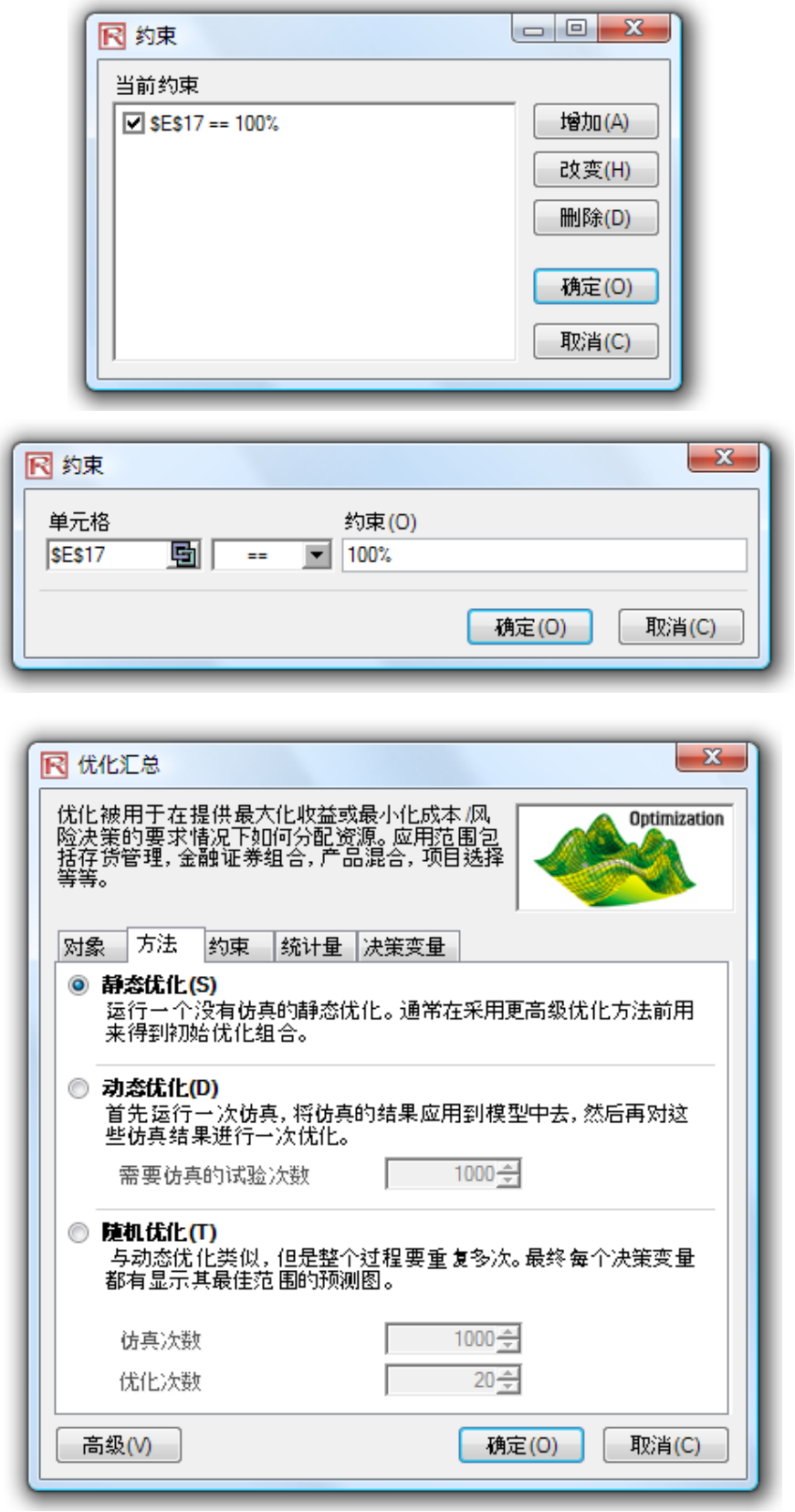


图 4.2 在 Risk Simulator 中运行连续型优化

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
组合总里	*12.6919%*	*4.52%*	*100.00%*							
风险收益率	*2.8091*									

图 4.3 连续型优化结果

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<=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。

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
项目1	$458.00	$1,732.44	$54.96	12.00%	8.33	1.26	1.00
项目2	$1,954.00	$859.00	$1,914.92	98.00%	1.02	3.27	1.00
项目3	$1,599.00	$1,845.00	$1,551.03	97.00%	1.03	1.87	1.00
项目4	$2,251.00	$1,645.00	$1,012.95	45.00%	2.22	2.37	1.00
项目5	$849.00	$458.00	$925.41	109.00%	0.92	2.85	1.00
项目6	$758.00	$52.00	$560.92	74.00%	1.35	15.58	1.00
项目7	$2,845.00	$758.00	$5,633.10	198.00%	0.51	4.75	1.00
项目8	$1,235.00	$115.00	$926.25	75.00%	1.33	11.74	1.00
项目9	$1,945.00	$125.00	$2,100.60	108.00%	0.93	16.56	1.00
项目10	$2,250.00	$458.00	$1,912.50	85.00%	1.18	5.91	1.00
项目11	$549.00	$45.00	$263.52	48.00%	2.08	13.20	1.00
项目12	$525.00	$105.00	$309.75	59.00%	1.69	6.00	1.00
加总：	$17,218.00	$8,197.44	$7,007.00	40.70%			12.00
目标：	最大化	<=$5000					<=6
Sharpe比率：	2.46						

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

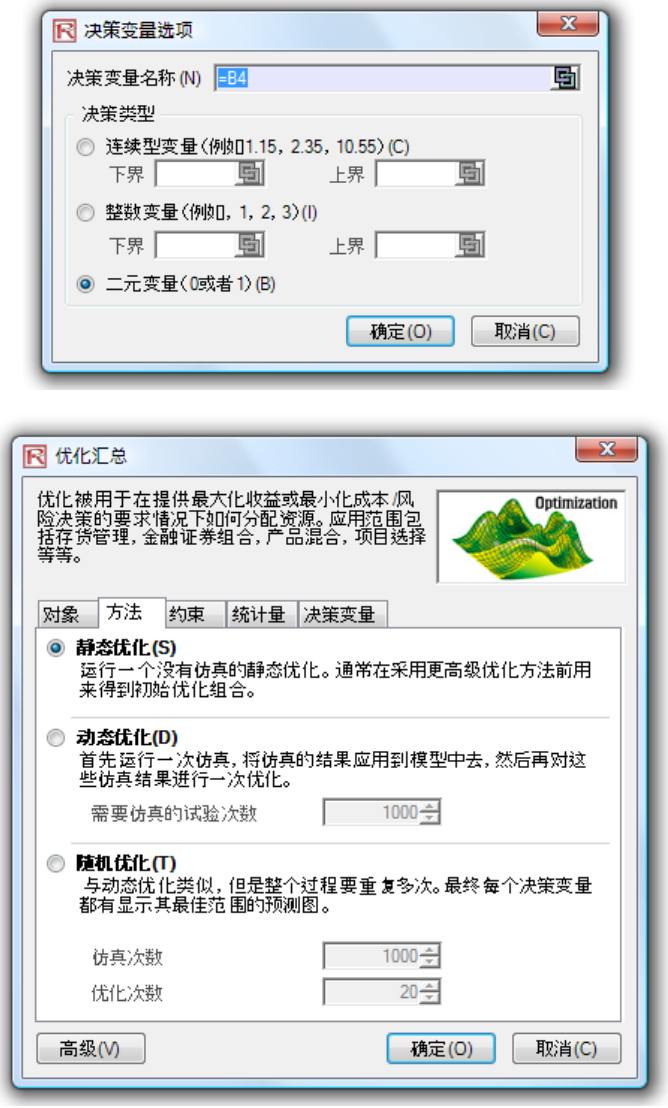
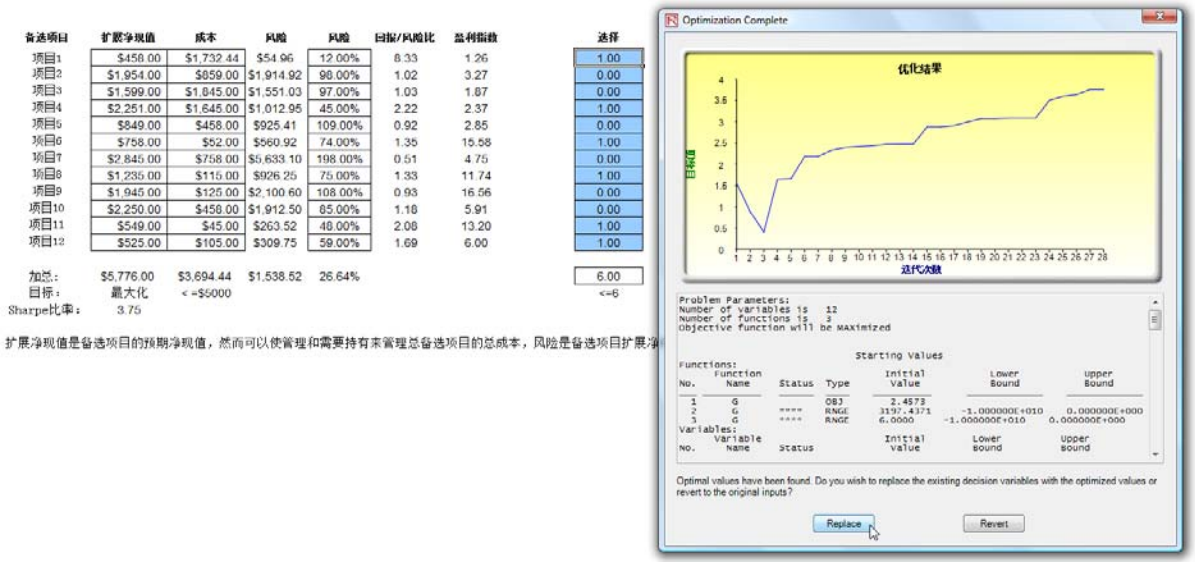


图 4.5 在 Risk Simulator 中运行离散型优化

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。



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
项目1	$456.00	$1,732.44	$54.96	12.00%	8.33	1.26	1.00
项目2	$1,954.00	$859.00	$1,914.92	98.00%	1.02	3.27	0.00
项目3	$1,599.00	$1,645.00	$1,551.03	97.00%	1.03	1.87	0.00
项目4	$2,251.00	$1,645.00	$1,012.95	45.00%	2.22	2.37	1.00
项目5	$849.00	$458.00	$925.41	109.00%	0.92	2.85	0.00
项目6	$758.00	$52.00	$560.92	74.00%	1.35	15.58	1.00
项目7	$2,845.00	$758.00	$5,633.10	198.00%	0.51	4.75	0.00
项目8	$1,235.00	$115.00	$926.25	75.00%	1.33	11.74	1.00
项目9	$1,945.00	$125.00	$2,100.60	108.00%	0.93	16.56	0.00
项目10	$2,250.00	$458.00	$1,912.50	85.00%	1.18	5.91	0.00
项目11	$549.00	$45.00	$263.52	48.00%	2.08	13.20	1.00
项目12	$525.00	$105.00	$309.75	59.00%	1.69	6.00	1.00
加总：	$5,776.00	$3,694.44	$1,538.52	26.64%			6.00
目标：	最大化	<=$5000					<=6
Sharpe比率：	3.75						

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

图 4.6 最大化夏普指数（Sharpe-Ratio）条件下的最佳项目选择

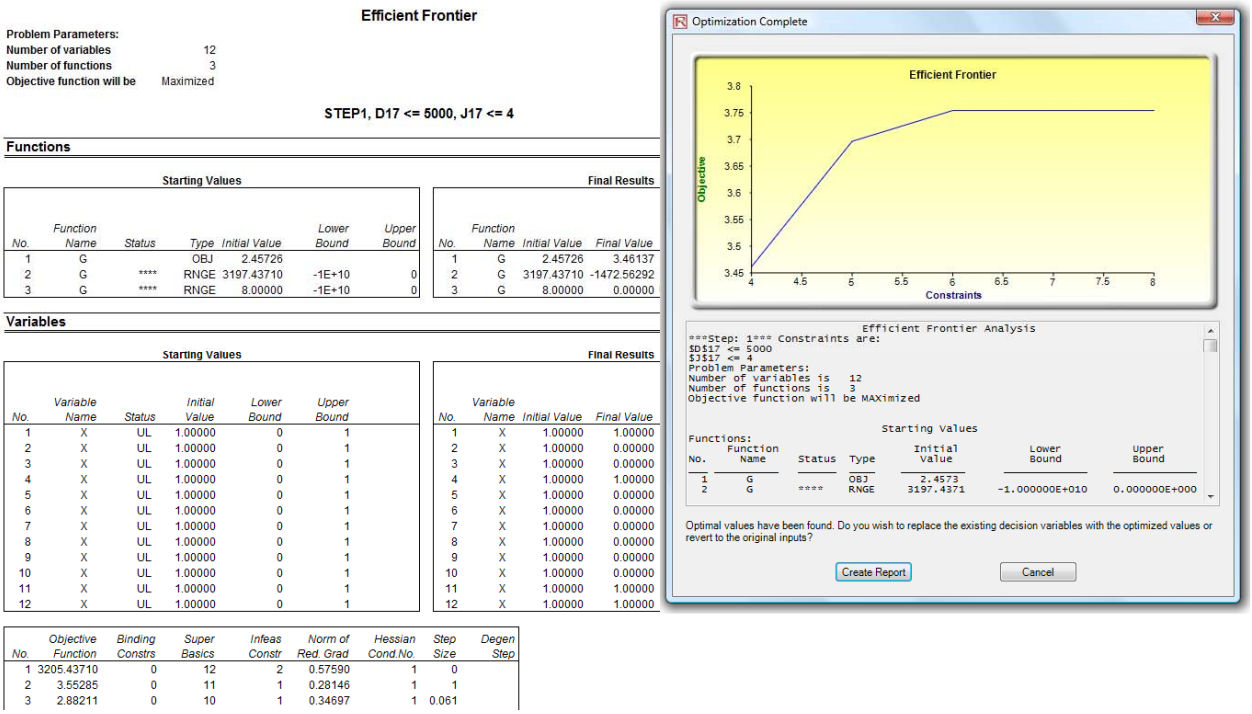
如果想要了解多关于优化的例子，可以参考《实物期权分析：工具和方法》，第二版一书的第十一章综合风险分析，（Wiley Finance, 2005）。这个案例学习解释了如何得到一个有效前沿，以及如何将预测，仿真，优化和实物期权融入到分析过程之中。

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 <= 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 <= 4, J17 <= 5, J17 <= 6, J17 <= 7 和 J17 <= 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



Efficient Frontier

Problem Parameters:

Number of variables	12
Number of functions	3
Objective function will be	Maximized

STEP1, D17 <= 5000, J17 <= 4

Functions

Starting Values

No.	Function Name	Status	Type	Initial Value	Lower Bound	Upper Bound
1	G		OBJ	2.45726		
2	G	****	RNGE	3197.43710	-1E+10	0
3	G	****	RNGE	8.00000	-1E+10	0

Final Results

No.	Function Name	Initial Value	Final Value
1	G	2.45726	3.46137
2	G	3197.43710	-1472.56292
3	G	8.00000	0.00000

Variables

Starting Values

No.	Variable Name	Status	Initial Value	Lower Bound	Upper Bound
1	X	UL	1.00000	0	1
2	X	UL	1.00000	0	1
3	X	UL	1.00000	0	1
4	X	UL	1.00000	0	1
5	X	UL	1.00000	0	1
6	X	UL	1.00000	0	1
7	X	UL	1.00000	0	1
8	X	UL	1.00000	0	1
9	X	UL	1.00000	0	1
10	X	UL	1.00000	0	1
11	X	UL	1.00000	0	1
12	X	UL	1.00000	0	1

Final Results

No.	Variable Name	Initial Value	Final Value
1	X	1.00000	1.00000
2	X	1.00000	0.00000
3	X	1.00000	0.00000
4	X	1.00000	1.00000
5	X	1.00000	0.00000
6	X	1.00000	0.00000
7	X	1.00000	0.00000
8	X	1.00000	0.00000
9	X	1.00000	0.00000
10	X	1.00000	0.00000
11	X	1.00000	1.00000
12	X	1.00000	1.00000

No.	Objective Function	Binding Constrs	Super Basics	Infeas Constr	Norm of Red. Grad	Hessian Cond.No.	Step Size	Degen Step
1	3205.43710	0	12	2	0.57590	1	0	
2	3.55285	0	11	1	0.28146	1	1	
3	2.88211	0	10	1	0.34697	1	0.061	

图 4.8

5．风险仿真分析工具

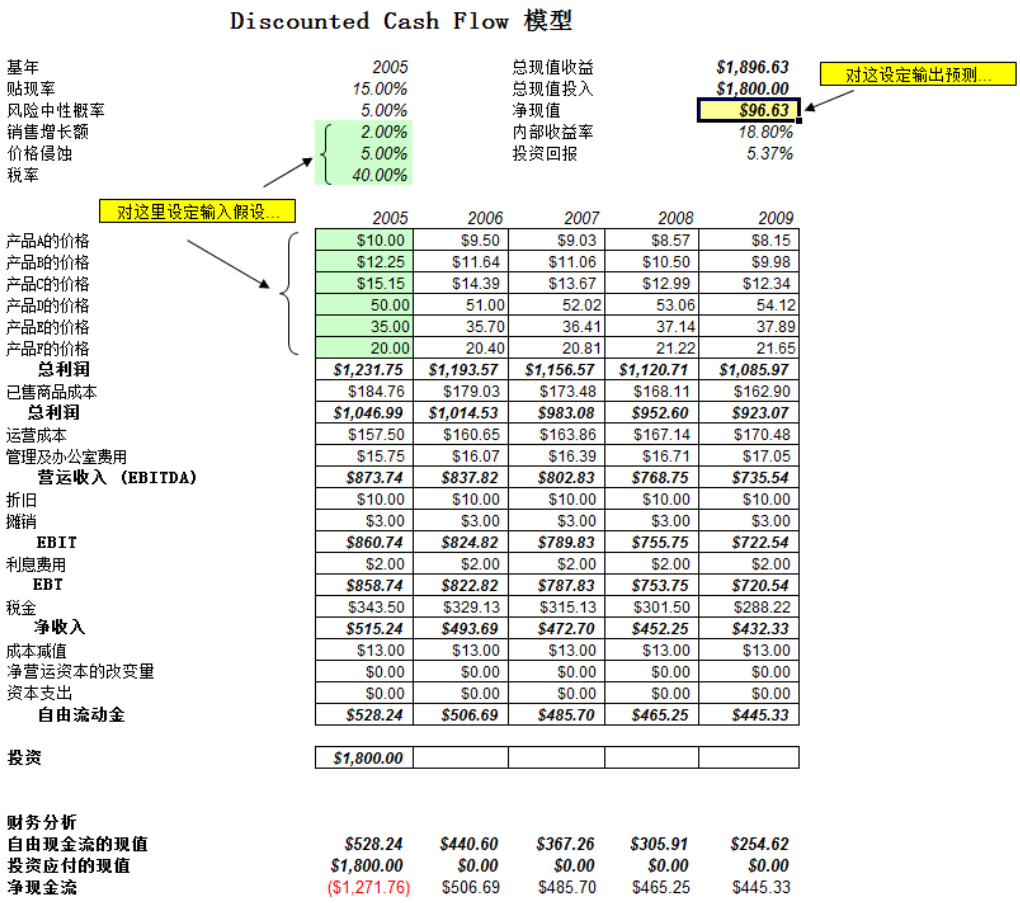
本章涉及的是 Risk Simulator 的分析工具。我们会通过逐步分析的方法来讨论 Risk Simulator 软件中的一些案例应用，以确定分析工具的作用。这些工具对于风险分析领域的分析者来说是非常有价值的。本章将会详细介绍每种工具的应用。

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±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%		

对这设定输出预测...

对这里设定输入假设...

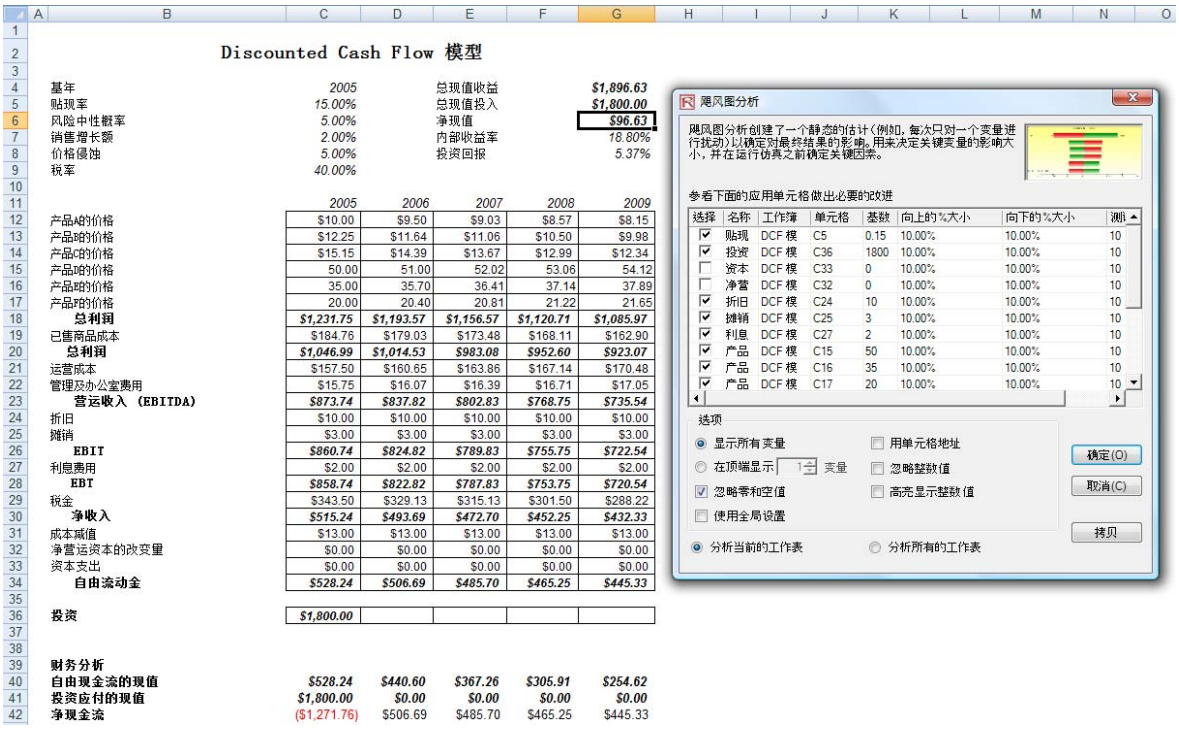
	2005	2006	2007	2008	2009
产品A的价格	$10.00	$9.50	$9.03	$8.57	$8.15
产品B的价格	$12.25	$11.64	$11.06	$10.50	$9.98
产品C的价格	$15.15	$14.39	$13.67	$12.99	$12.34
产品D的价格	50.00	51.00	52.02	53.06	54.12
产品E的价格	35.00	35.70	36.41	37.14	37.89
产品F的价格	20.00	20.40	20.81	21.22	21.65
总利润	***$1,231.75***	***$1,193.57***	***$1,156.57***	***$1,120.71***	***$1,085.97***
已售商品成本	$184.76	$179.03	$173.48	$168.11	$162.90
总利润	***$1,046.99***	***$1,014.53***	***$983.08***	***$952.60***	***$923.07***
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
营运收入 (EBITDA)	***$873.74***	***$837.82***	***$802.83***	***$768.75***	***$735.54***
折旧	$10.00	$10.00	$10.00	$10.00	$10.00
摊销	$3.00	$3.00	$3.00	$3.00	$3.00
EBIT	***$860.74***	***$824.82***	***$789.83***	***$755.75***	***$722.54***
利息费用	$2.00	$2.00	$2.00	$2.00	$2.00
EBT	***$858.74***	***$822.82***	***$787.83***	***$753.75***	***$720.54***
税金	$343.50	$329.13	$315.13	$301.50	$288.22
净收入	***$515.24***	***$493.69***	***$472.70***	***$452.25***	***$432.33***
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
自由流动金	***$528.24***	***$506.69***	***$485.70***	***$465.25***	***$445.33***
投资	***$1,800.00***				

财务分析					
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***
净现金流	($1,271.76)	$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**



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
总利润	***$1,231.75***	***$1,193.57***	***$1,156.57***	***$1,120.71***	***$1,085.97***
已售商品成本	$184.76	$179.03	$173.48	$168.11	$162.90
总利润	***$1,046.99***	***$1,014.53***	***$983.08***	***$952.60***	***$923.07***
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
营运收入 (EBITDA)	***$873.74***	***$837.82***	***$802.83***	***$768.75***	***$735.54***
折旧	$10.00	$10.00	$10.00	$10.00	$10.00
摊销	$3.00	$3.00	$3.00	$3.00	$3.00
EBIT	***$860.74***	***$824.82***	***$789.83***	***$755.75***	***$722.54***
利息费用	$2.00	$2.00	$2.00	$2.00	$2.00
EBT	***$858.74***	***$822.82***	***$787.83***	***$753.75***	***$720.54***
税金	$343.50	$329.13	$315.13	$301.50	$288.22
净收入	***$515.24***	***$493.69***	***$472.70***	***$452.25***	***$432.33***
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
自由流动金	***$528.24***	***$506.69***	***$485.70***	***$465.25***	***$445.33***
投资	***$1,800.00***				

财务分析					
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***
净现金流	($1,271.76)	$506.69	$485.70	$465.25	$445.33

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）。

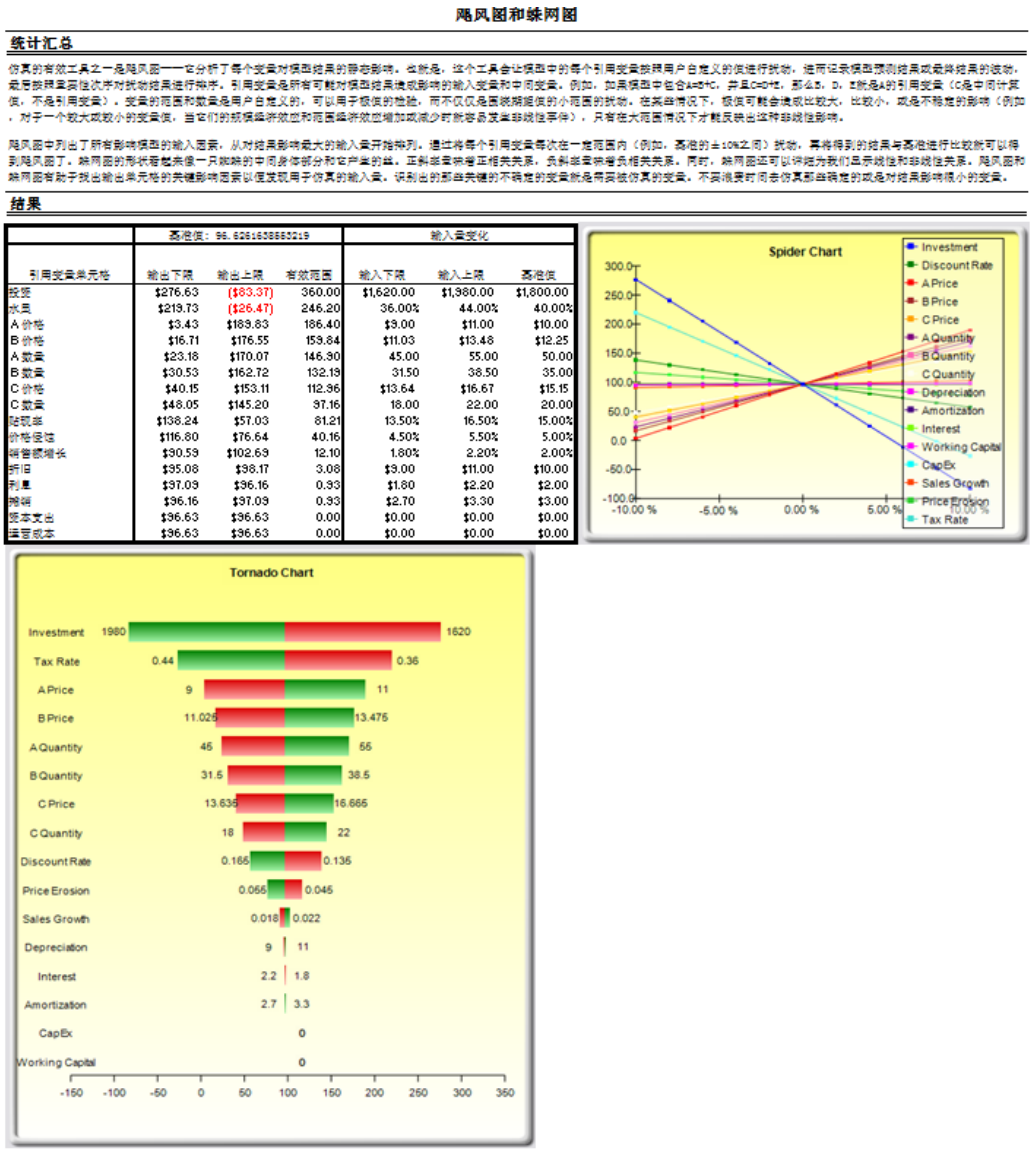


图 5.3——飓风图分析报告

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。

	基准值: 96.6261638553219			输入量变化		
引用变量单元格	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	($63.37)	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 数里	$23.18	$170.07	146.90	45.00	55.00	50.00
B 数里	$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 数里	$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

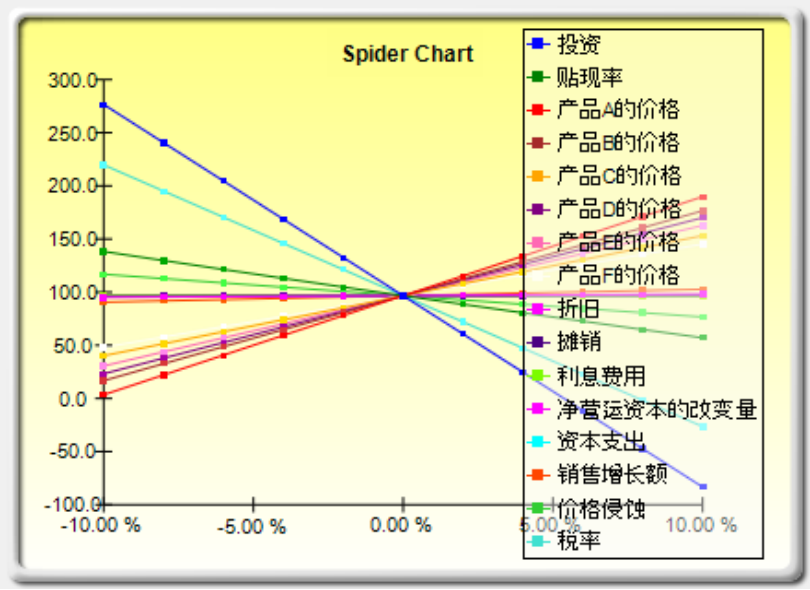


图 5.5——蛛网图

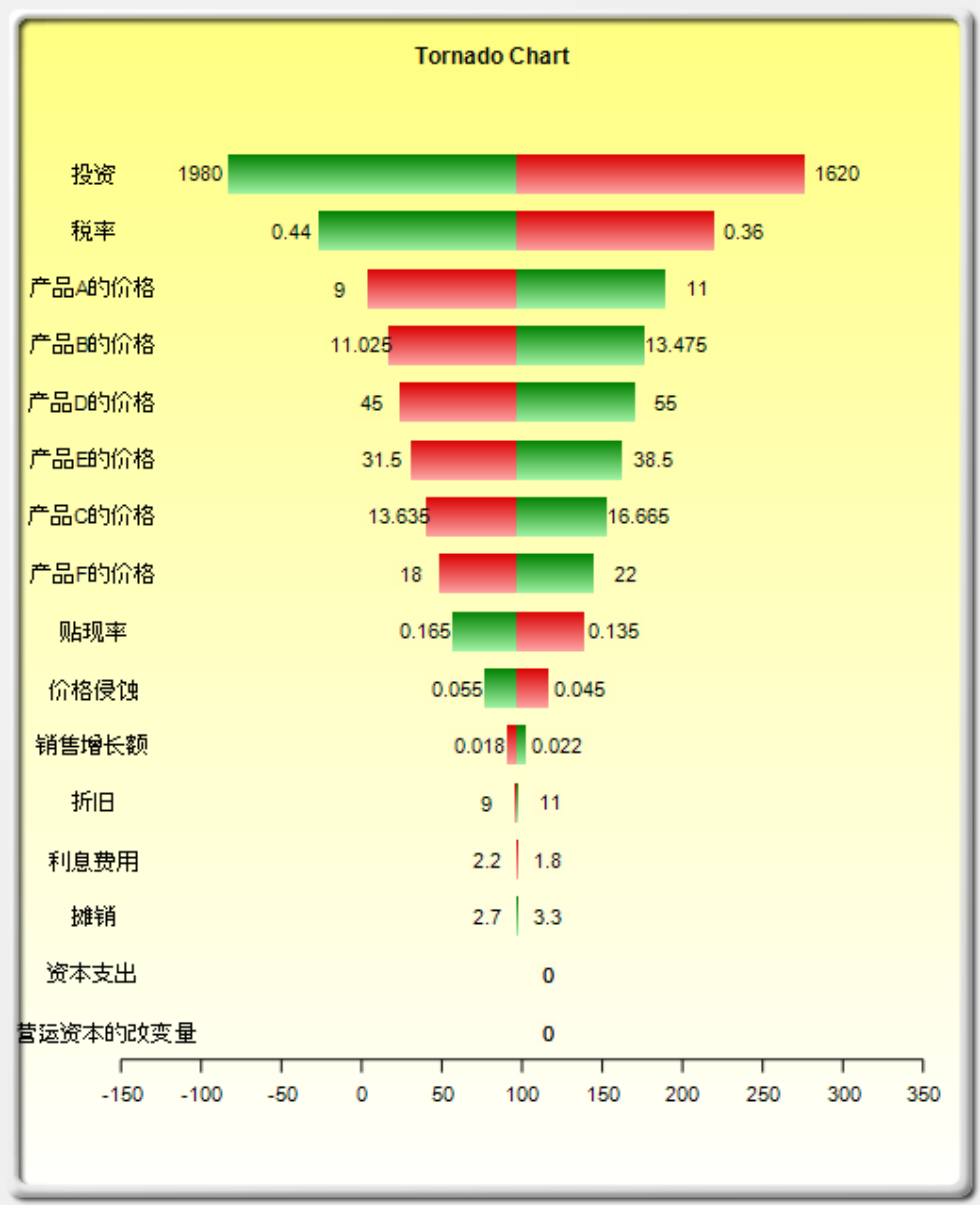


图 5.6——飓风图

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。

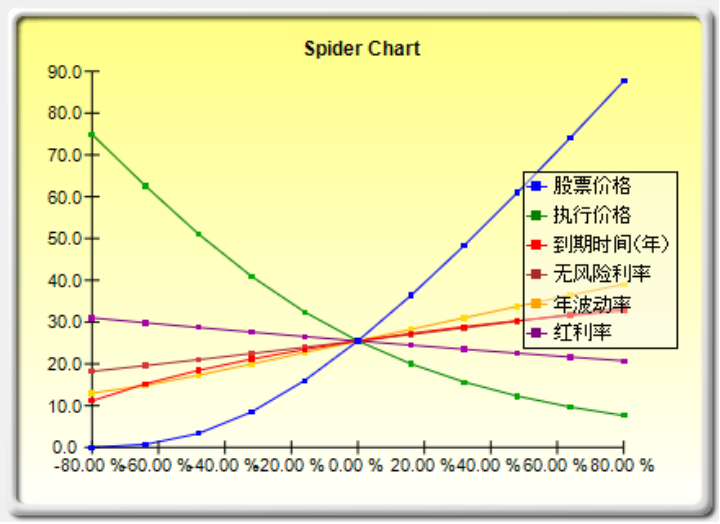


图 5.7——非线性蛛网图

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±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±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，±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。

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。

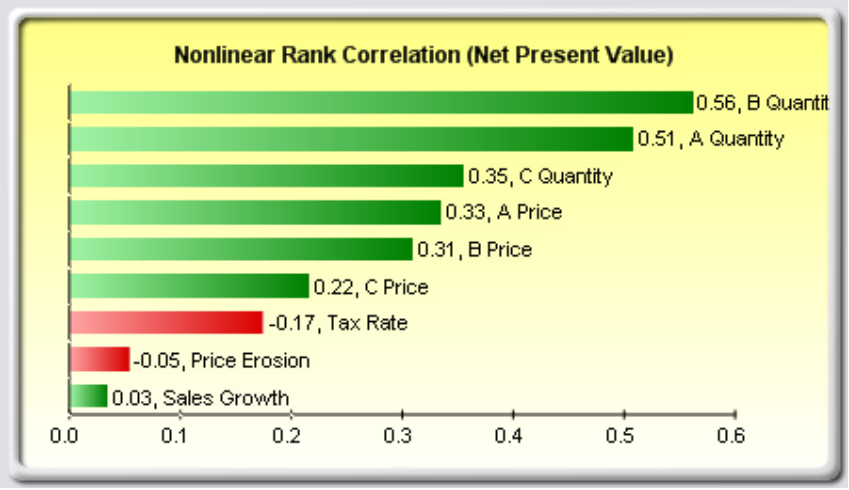


图 5.8——不存在相关性的敏感性图

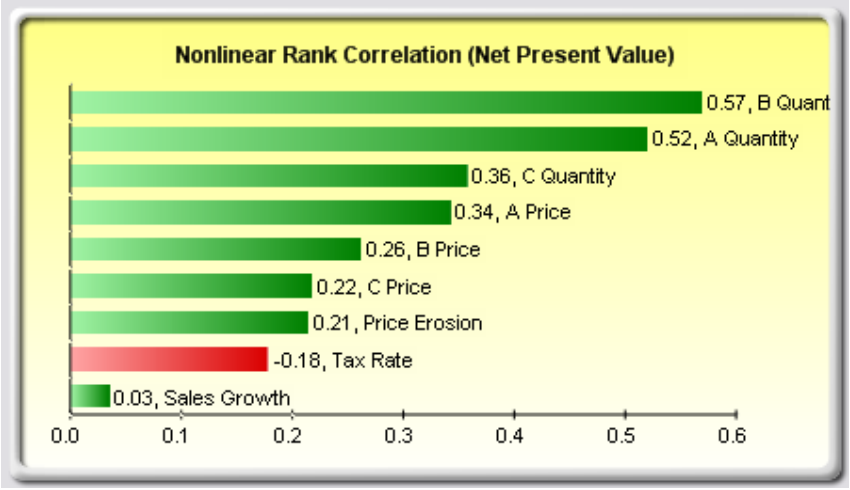


图 5.9——存在相关性的敏感性图

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）

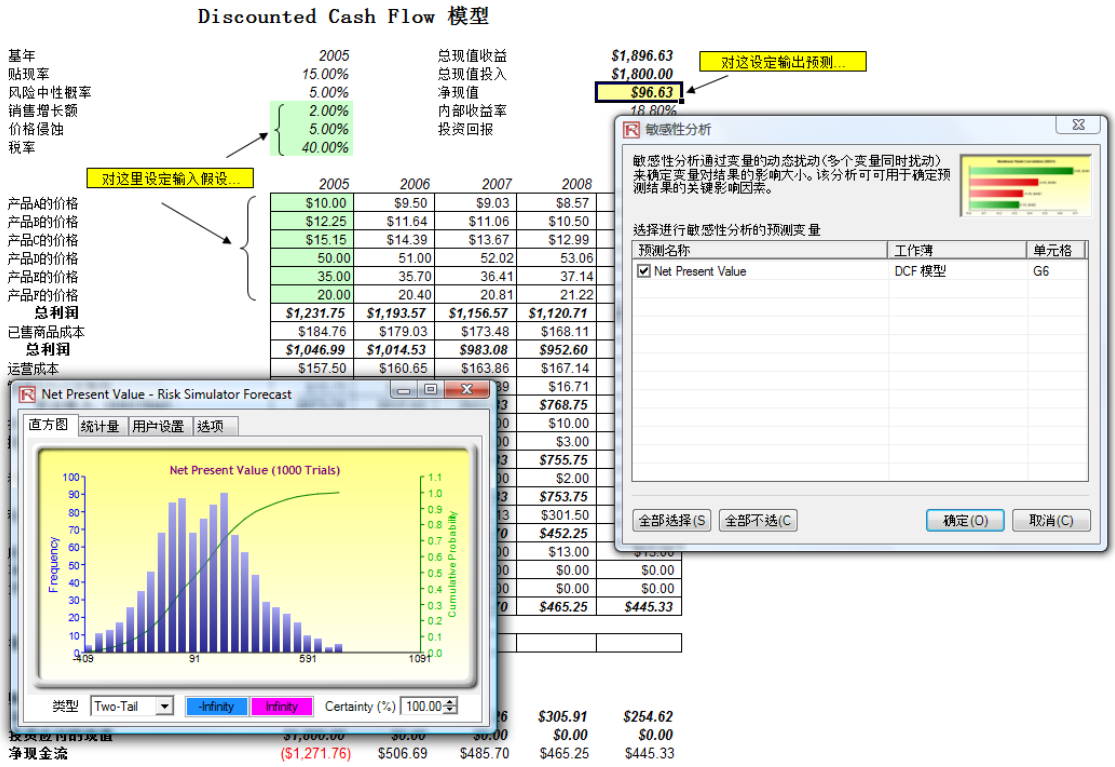


图 5.10——运行敏感性分析

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 β 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出），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）。

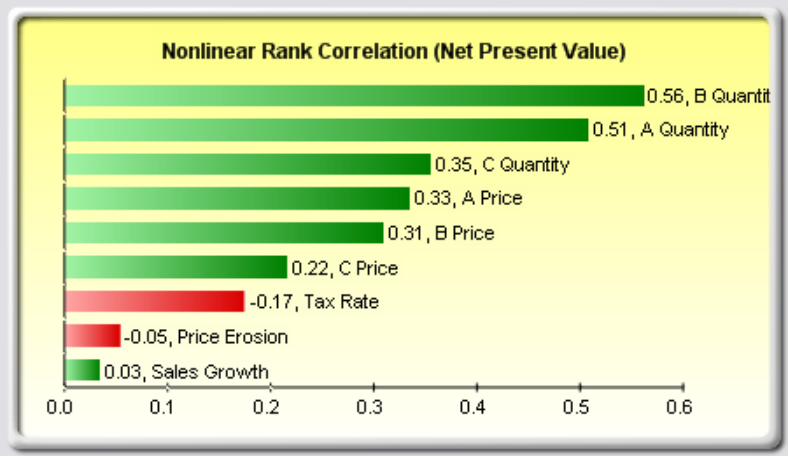


图 5.11——相关性排列图

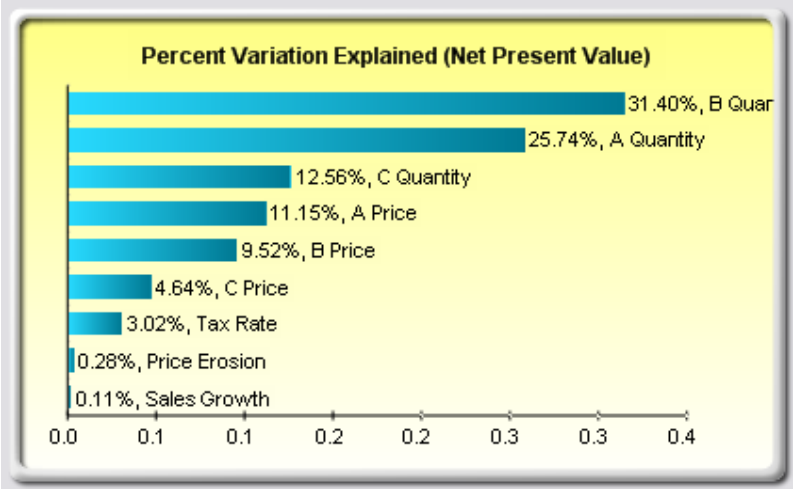


图 5.12——方差贡献图

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。

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）

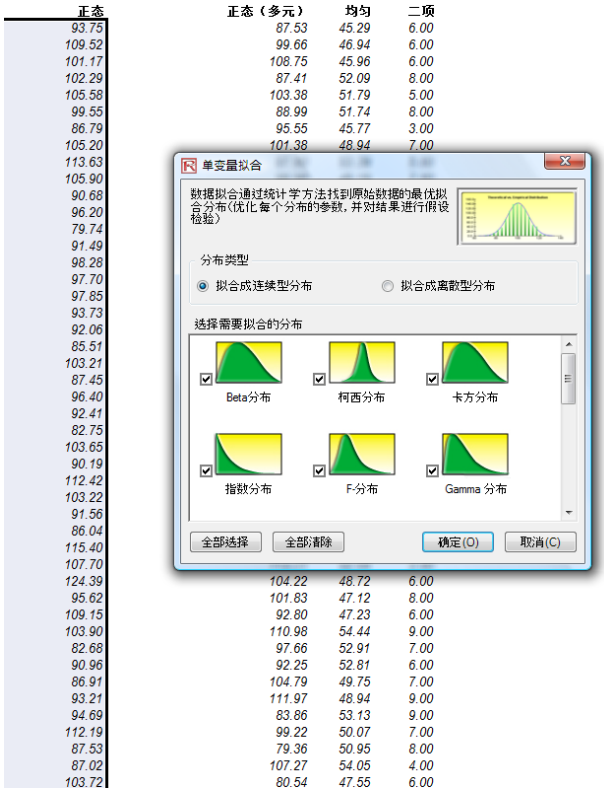


图 5.13——单变量分布拟合

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 α 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。

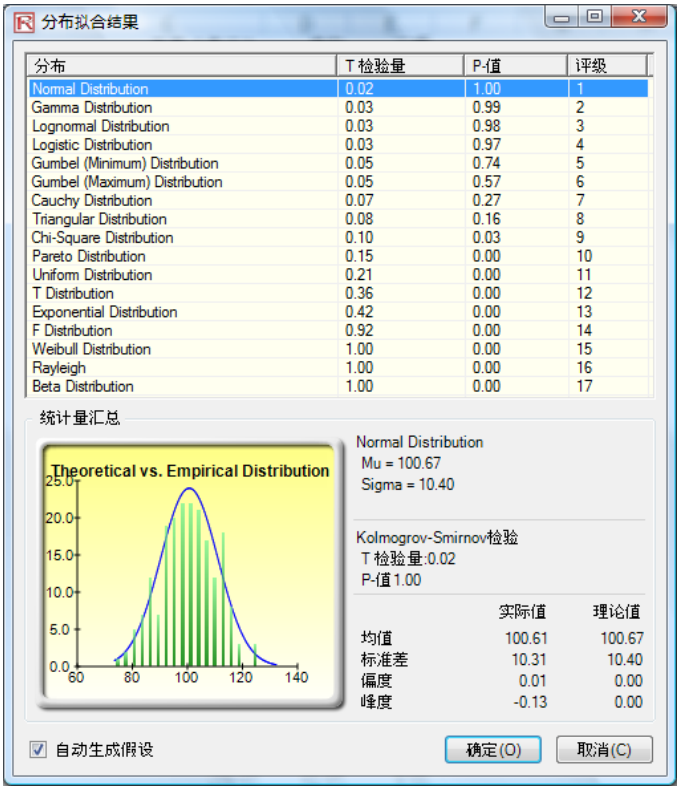


图 5.14：分布拟合结果

单变量分布拟合

统计结果

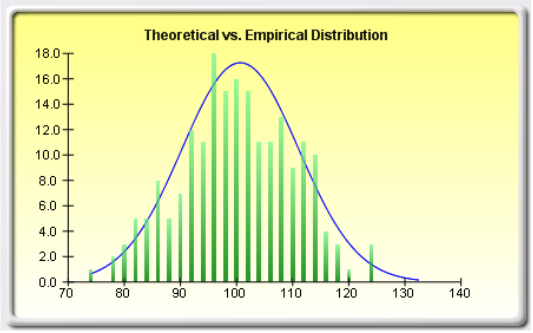
拟合假设 *100.61*

拟合的分布 ***Normal***

Mu	*100.67*
Sigma	*10.40*

Kolmogorov-Smirnov 统计量	*0.02*
P-Value 统计量	*0.9996*

	实际	理论
期望	*100.61*	*100.67*
标准差	*10.31*	*10.40*
偏度	*0.01*	*0.00*
超峰度	*-0.13*	*0.00*



用于拟合的原始数据

73.53	*78.21*	*78.52*	*79.50*	*79.72*	*79.74*	*81.56*	*82.08*	*82.68*	*82.75*	*83.34*	*83.64*	*84.09*
84.66	*85.00*	*85.35*	*85.51*	*86.04*	*86.79*	*86.82*	*86.91*	*87.02*	*87.03*	*87.45*	*87.53*	*87.66*
88.05	*88.45*	*88.51*	*89.95*	*90.19*	*90.54*	*90.68*	*90.96*	*91.25*	*91.49*	*91.56*	*91.94*	*92.06*
92.36	*92.41*	*92.45*	*92.70*	*92.80*	*92.84*	*93.21*	*93.26*	*93.48*	*93.73*	*93.75*	*93.77*	*93.82*
94.00	*94.15*	*94.51*	*94.57*	*94.64*	*94.69*	*94.95*	*95.57*	*95.62*	*95.71*	*95.78*	*95.83*	*95.97*
96.20	*96.24*	*96.40*	*96.43*	*96.47*	*96.81*	*96.88*	*97.00*	*97.07*	*97.21*	*97.23*	*97.48*	*97.70*
97.77	*97.85*	*98.15*	*98.17*	*98.24*	*98.28*	*98.32*	*98.33*	*98.35*	*98.65*	*99.03*	*99.27*	*99.46*
99.47	*99.55*	*99.73*	*99.96*	*100.08*	*100.24*	*100.36*	*100.42*	*100.44*	*100.48*	*100.49*	*100.83*	*101.17*
101.28	*101.34*	*101.45*	*101.46*	*101.55*	*101.73*	*101.74*	*101.81*	*102.29*	*102.55*	*102.58*	*102.60*	*102.70*
103.17	*103.21*	*103.22*	*103.32*	*103.34*	*103.45*	*103.65*	*103.66*	*103.72*	*103.81*	*103.90*	*103.99*	*104.46*
104.57	*104.76*	*105.20*	*105.44*	*105.50*	*105.52*	*105.58*	*105.66*	*105.87*	*105.90*	*105.90*	*106.29*	*106.35*
106.59	*107.01*	*107.68*	*107.70*	*107.93*	*108.17*	*108.20*	*108.34*	*108.42*	*108.43*	*108.49*	*108.70*	*109.15*
109.22	*109.35*	*109.52*	*109.75*	*110.04*	*110.16*	*110.25*	*110.54*	*111.05*	*111.06*	*111.44*	*111.76*	*111.90*
111.95	*112.07*	*112.19*	*112.29*	*112.32*	*112.42*	*112.48*	*112.85*	*112.92*	*113.50*	*113.59*	*113.63*	*113.70*
114.13	*114.14*	*114.21*	*114.91*	*114.95*	*115.40*	*115.58*	*115.66*	*116.58*	*116.98*	*117.60*	*118.67*	*119.24*
119.52	*124.14*	*124.16*	*124.39*	*132.30*								

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。

Bootstrap 仿真（拔靴法）

理论：

Bootstrap 仿真是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）

	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
出预

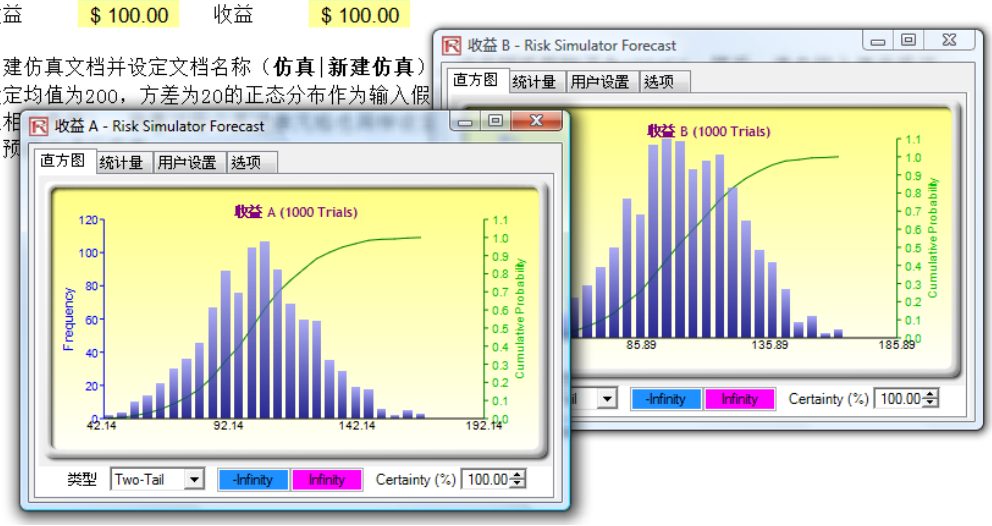


图 5.16——非参数 Bootstrap

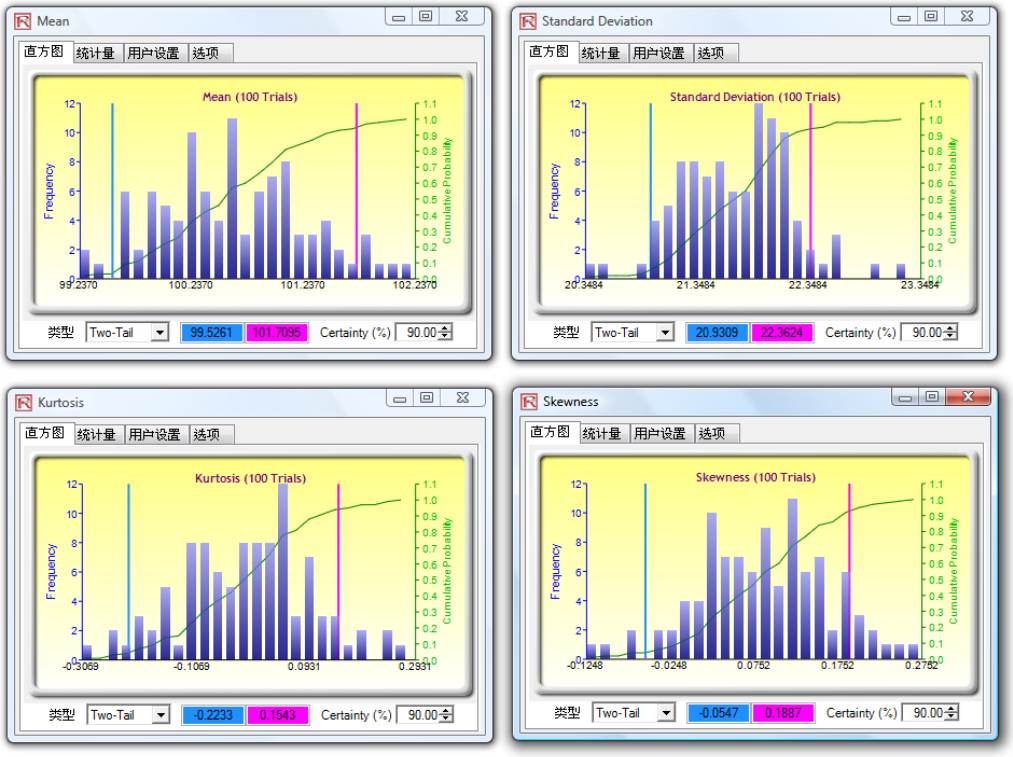


图 5.17——Bootstrap 模拟结果

结果解析：

实际上，非参数 Bootstrap 仿真可以被看作是基于仿真的仿真。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 值落在这个区间，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 值落在这个区间之外，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

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）

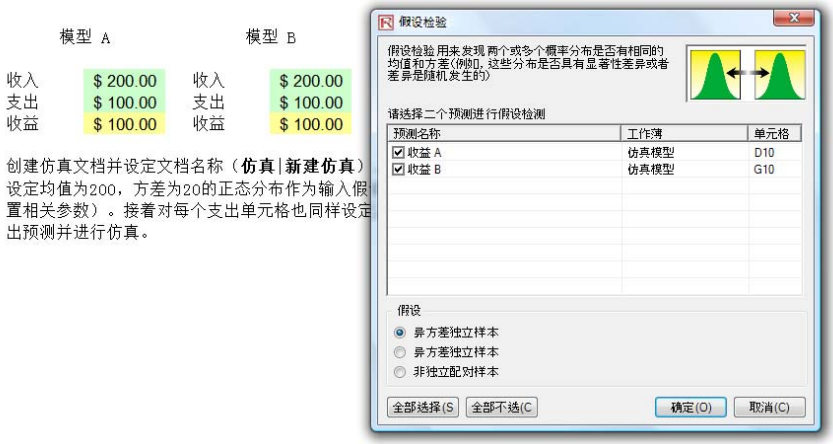
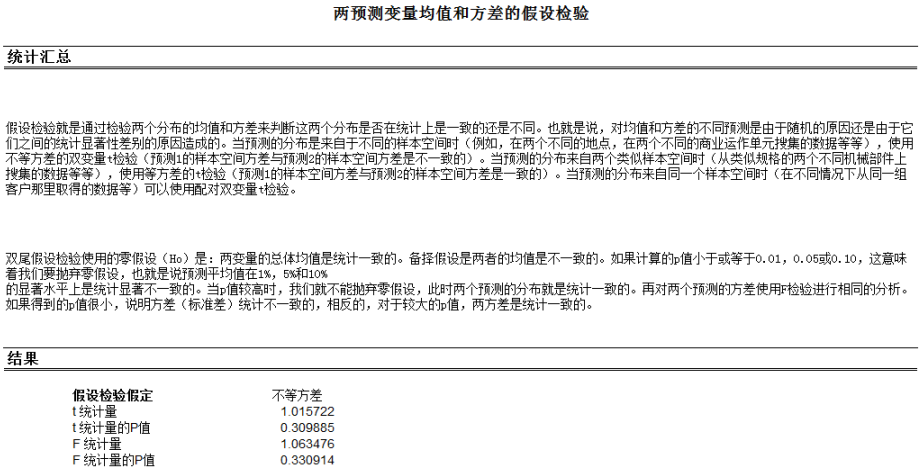


图 5.18——假设检验

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的样本空间方差是不一致的）。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（Ho）是：两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。

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 的样本空间方差是不一致的）。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。

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

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**

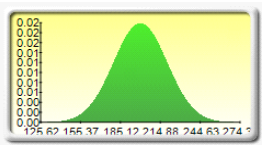
Simulation - Hypothesis Testing

General

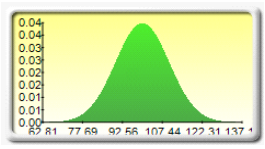
Number of Trials	1000
Stop Simulation on Error	No
Random Seed	123456
Enable Correlations	Yes

Assumptions

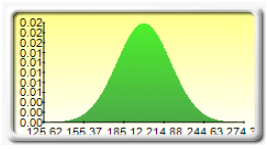
Name	收入
Enabled	Yes
Cell	D8
Dynamic Simulation	No
Range	
Minimum	-Infinity
Maximum	+Infinity
Distribution	Normal
Mean	200
Standard Deviation	20



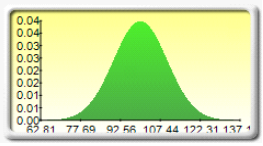
Name	支出
Enabled	Yes
Cell	D9
Dynamic Simulation	No
Range	
Minimum	-Infinity
Maximum	+Infinity
Distribution	Normal
Mean	100
Standard Deviation	10



Name	收入
Enabled	Yes
Cell	G8
Dynamic Simulation	No
Range	
Minimum	-Infinity
Maximum	+Infinity
Distribution	Normal
Mean	200
Standard Deviation	20



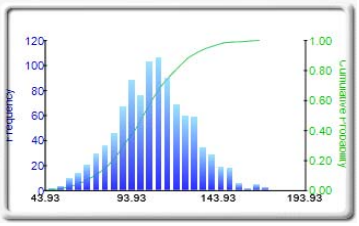
Name	支出
Enabled	Yes
Cell	G9
Dynamic Simulation	No
Range	
Minimum	-Infinity
Maximum	+Infinity
Distribution	Normal
Mean	100
Standard Deviation	10



Forecasts

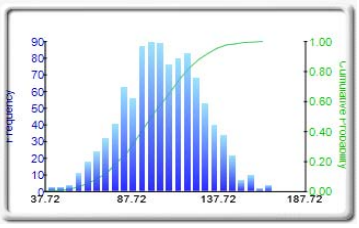
Name	收益 A
Enabled	Yes
Cell	D10
Forecast Precision	
Precision Level	---
Error Level	---

Number of Datapoints	1000
Mean	100.5485
Median	100.4388
Standard Deviation	21.6421
Variance	468.3797
Average Deviation	17.1009
Maximum	167.0852
Minimum	38.8005
Range	128.2848
Skewness	0.0794
Kurtosis	-0.0220
25% Percentile	86.6101
75% Percentile	114.8279
Error Precision at 95%	0.0133



Name	收益 B
Enabled	Yes
Cell	G10
Forecast Precision	
Precision Level	---
Error Level	---

Number of Datapoints	1000
Mean	99.1790
Median	99.1303
Standard Deviation	22.5914
Variance	510.3735
Average Deviation	18.2828
Maximum	162.7509
Minimum	32.5085
Range	130.2424
Skewness	-0.0498
Kurtosis	-0.2711
25% Percentile	84.1171
75% Percentile	114.9416
Error Precision at 95%	0.0141



Correlation Matrix

	收入	支出	收入	支出
收入	1.00			
支出	0.00	1.00		
收入	0.00	0.00	1.00	
支出	0.00	0.00	0.00	1.00

图 5.21——样本仿真报告

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，微数缺测性，平稳和随机性，误差的正态性和球性 ，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	7.73	1042	6.6

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 方）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（微数缺测性），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会产生不正确的结果。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。在使用回归钱，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

	异方差		微数缺测性		异常		非线性	
变量	W检验 p值	假设检验 结果	近似 结果	本性 下届	本性 上界	潜在的 异常数	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	**0.0335**	非线性
变量 X3	0.3649	同方差	没问题	-5.77	15.69	3	**0.0305**	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）。

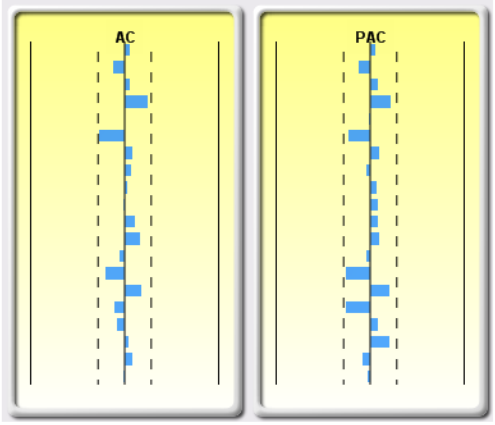
如果忽略这种关系，那得到的预测是有偏的，也是不精确的。对这样的情况，使用自回归模型或自回归求和滑动平均效果将会更好（**Risk Simulator** |预测|ARIMA）。最后，需要指出，对那些非稳态数据的自相关函数趋向于较慢的衰减（参见非稳态报告）。

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。

自相关

时间延迟	AC	PAC	下届	上届	Q统计量	概率
1	0.0580	0.0580	-0.2828	0.2828	0.1786	0.6726
2	-0.1213	-0.1251	-0.2828	0.2828	0.9754	0.6140
3	0.0590	0.0756	-0.2828	0.2828	1.1679	0.7607
4	0.2423	0.2232	-0.2828	0.2828	4.4865	0.3442
5	0.0067	-0.0078	-0.2828	0.2828	4.4890	0.4814
6	-0.2654	-0.2345	-0.2828	0.2828	8.6516	0.1941
7	0.0814	0.0939	-0.2828	0.2828	9.0524	0.2489
8	0.0634	-0.0442	-0.2828	0.2828	9.3012	0.3175
9	0.0204	0.0673	-0.2828	0.2828	9.3276	0.4076
10	-0.0190	0.0865	-0.2828	0.2828	9.3512	0.4991
11	0.1035	0.0790	-0.2828	0.2828	10.0648	0.5246
12	0.1658	0.0978	-0.2828	0.2828	11.9466	0.4500
13	-0.0524	-0.0430	-0.2828	0.2828	12.1394	0.5162
14	-0.2050	-0.2523	-0.2828	0.2828	15.1738	0.3664
15	0.1782	0.2089	-0.2828	0.2828	17.5315	0.2881
16	-0.1022	-0.2591	-0.2828	0.2828	18.3296	0.3050
17	-0.0861	0.0808	-0.2828	0.2828	18.9141	0.3335
18	0.0418	0.1987	-0.2828	0.2828	19.0559	0.3884
19	0.0869	-0.0821	-0.2828	0.2828	19.6894	0.4135
20	-0.0091	-0.0269	-0.2828	0.2828	19.6966	0.4770



分布滞后

自变量的分布滞后期数的p值

变量	1	2	3	4	5	6	7	8	9	10	11	12
X1	0.8467	0.2045	0.3336	0.9105	0.9757	0.1020	0.9205	0.1267	0.5431	0.9110	0.7495	0.4016
X2	0.6077	0.9900	0.8422	0.2851	0.0638	0.0032	0.8007	0.1551	0.4823	0.1126	0.0519	0.4383
X3	0.7394	0.2396	0.2741	0.8372	0.9808	0.0464	0.8355	0.0545	0.6828	0.7354	0.5093	0.3500
X4	0.0061	0.6739	0.7932	0.7719	0.6748	0.8627	0.5586	0.9046	0.5726	0.6304	0.4812	0.5707
X5	0.1591	0.2032	0.4123	0.5599	0.6416	0.3447	0.9190	0.9740	0.5185	0.2856	0.1489	0.7794

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

回归平均误差	0.00
误差的标准差	141.83
D 统计量	0.1036
1%下的D临界值	0.1138
5%下的D临界值	0.1225
10%下的D临界值	0.1458

零假设：误差是正态分布的。

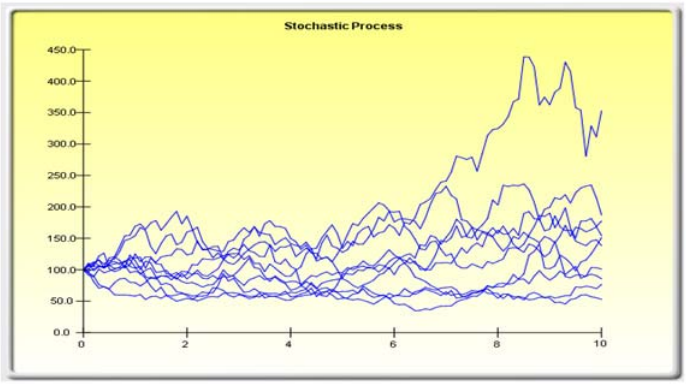
结论：误差是 正态分布的 在 1% 显著性水平下.

误差	相关频率	观察值	预期值	O-E
-219.04	0.02	0.02	0.0612	-0.0412
-202.53	0.02	0.04	0.0766	-0.0366
-186.04	0.02	0.06	0.0948	-0.0348
-174.17	0.02	0.08	0.1097	-0.0297
-162.13	0.02	0.10	0.1265	-0.0265
-161.62	0.02	0.12	0.1272	-0.0072
-160.39	0.02	0.14	0.1291	0.0109
-145.40	0.02	0.16	0.1526	0.0074
-138.92	0.02	0.18	0.1637	0.0163
-133.81	0.02	0.20	0.1727	0.0273
-120.76	0.02	0.22	0.1973	0.0227
-120.12	0.02	0.24	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视觉化效果（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					
漂移率	-1.48%	回复率	283.89%	跳跃率	20.41%
波动率	88.84%	远期值	327.72	跳跃大小	237.89

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）。

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**) 回到数据工作薄，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）

图 5.28 – 运行统计分析工具

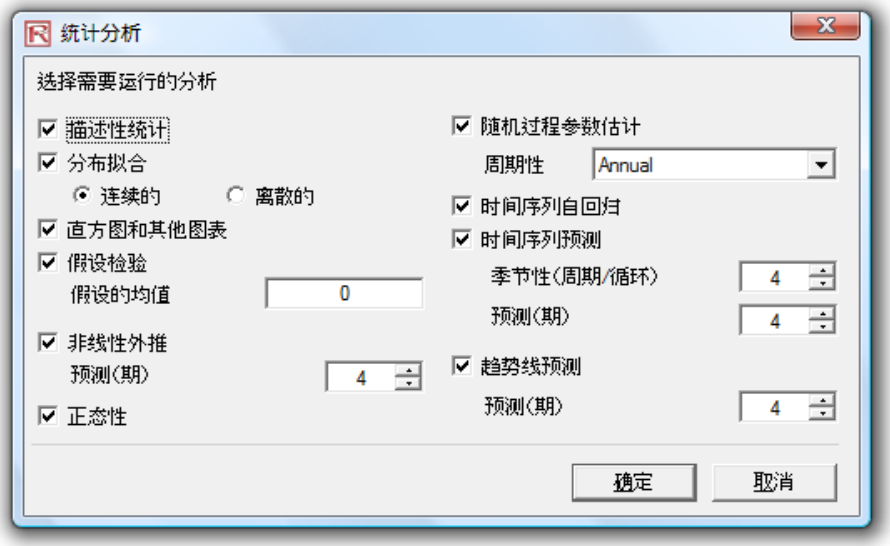


图 5.29 – 统计检验

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数相等。当中位数远离均值的时候就存在偏斜的分布。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离中程度。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。样本标准差不同于总体标准差，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士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）。第一分位差用来度量按照数据从小到大排列的第二十五百分数。第叁分为数是七十五百分数。有时候，分位差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。四分位数间距不同于第一和第三分位差，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。是分部的第四矩。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	四分位数间距	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）

统计汇总

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
		P 值（双尾）	0.0000
使用者使用的统计量			
		零假设（Ho）：	μ = 假设的均值
假设均值	0.00	备择假设（Ha）：	μ < > 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。

假设检验汇总

单变量t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）

双尾假设检验

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。

左尾假设检验

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。

右尾假设检验

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
标准差	172.91
D 统计量	0.0859
1%显著性水平的D值	0.1150
5%显著性水平的D值	0.1237
10%显著性水平的D(	0.1473

零假设：数据的分布服从正态分布

结论：样本数据是 正态 分布的 1% 显著性水平.

数据	相关频率	观察值	估计值	O-E
47.00	0.02	0.02	0.0497	-0.0297
68.00	0.02	0.04	0.0635	-0.0235
87.00	0.02	0.06	0.0783	-0.0183
96.00	0.02	0.08	0.0862	-0.0062
102.00	0.02	0.10	0.0918	0.0082
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	0.3620	0.0180

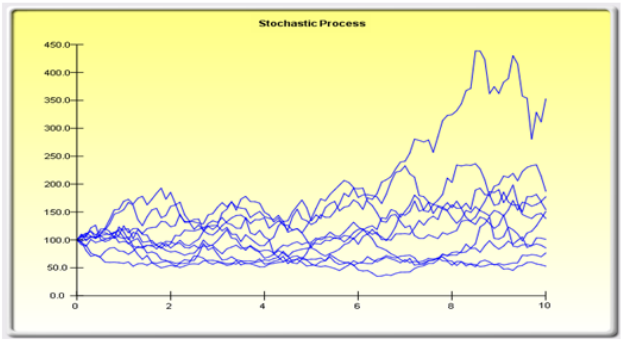
图 5.32 – 示例统计分析报告(正态检验)

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

漂移率	-1.48%	回复比率	283.89%	跳跃率	20.41%
波动性	88.84%	长期值	327.72	跳跃大小	237.89

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)

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，也包括理论的分布四矩。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%。

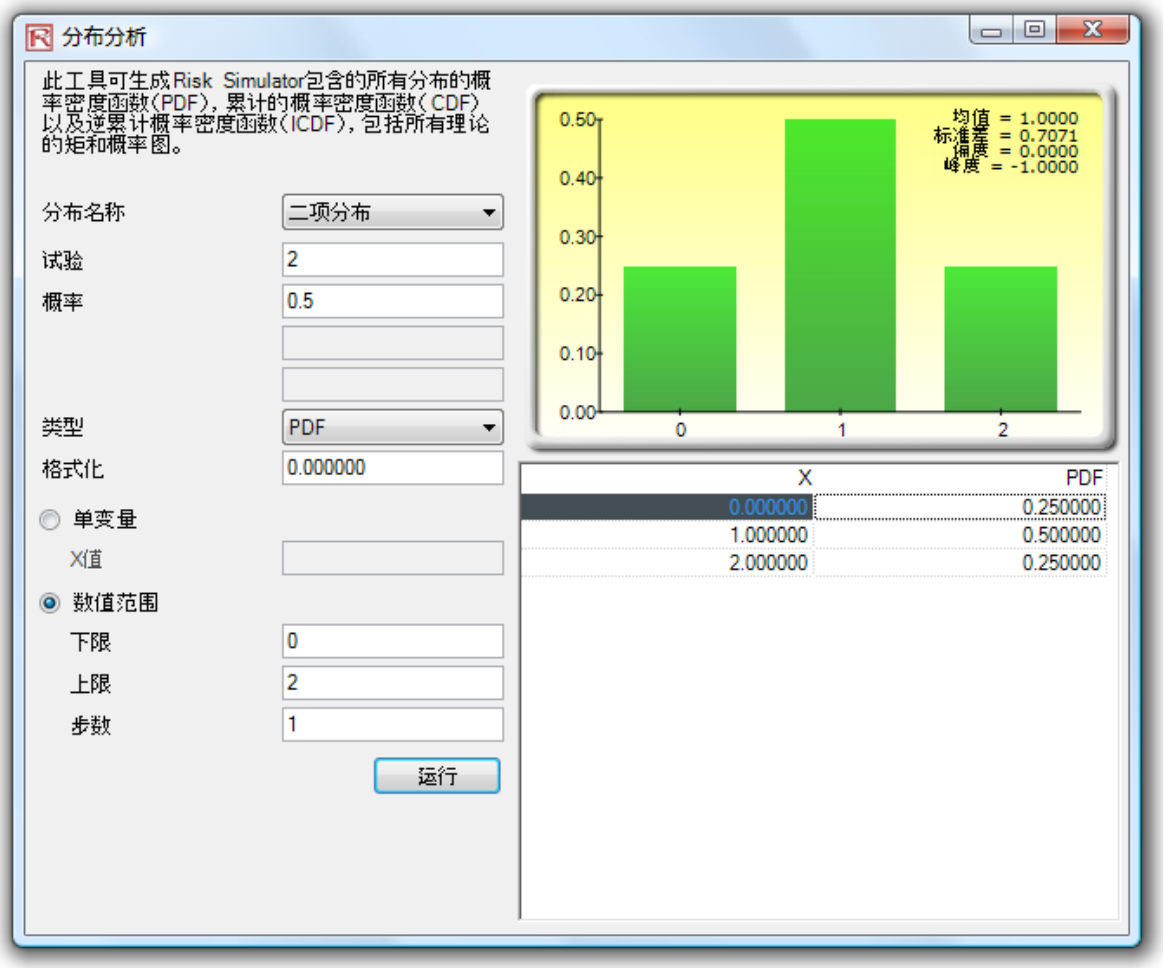


图 5.34 – 示例统计分析报告(2 次试验的二项分布)

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。

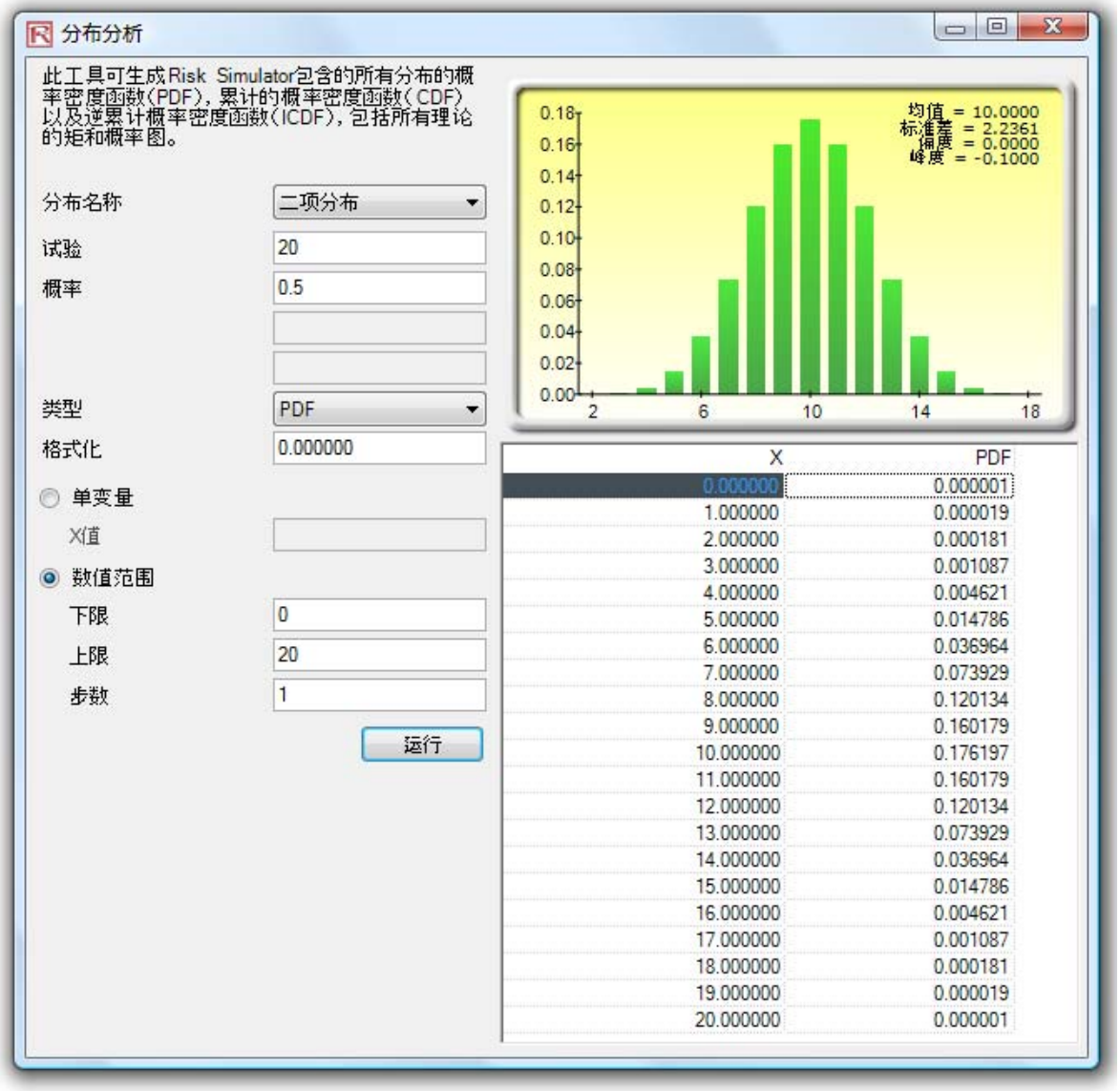


图 5.35 –示例统计分析报告(20 次试验的二项分布)

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。

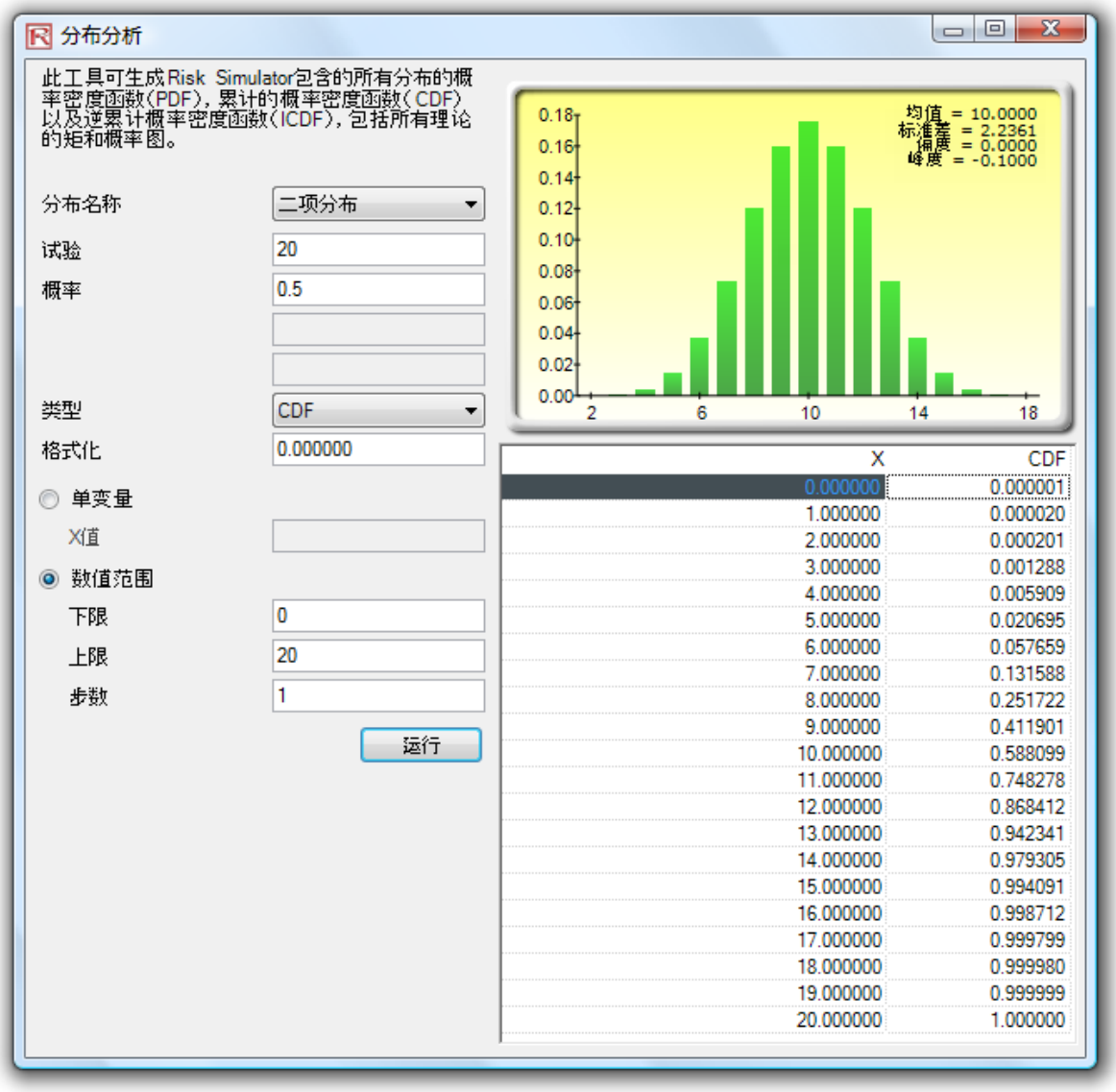


图 5.36 – 示例统计分析报告(20 次试验二项分布的 CDF)

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。

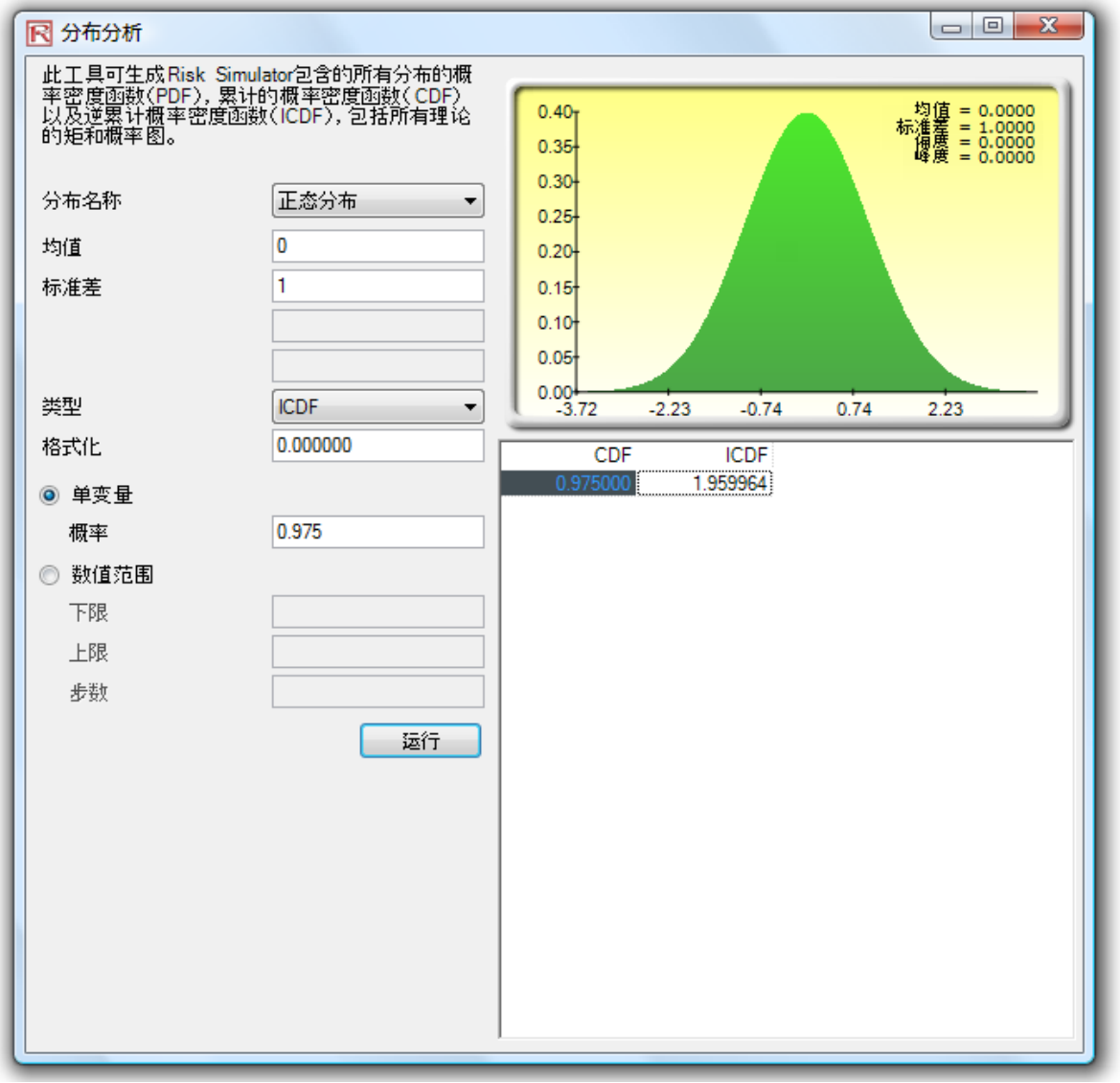


图 5.37 – 示例统计分析报告(正态分布的 ICDF 和 Z 值)