

RISK SIMULATOR

用戶手冊

Johnathan Mun, Ph.D., MBA, MS, BS, CRM, FRM, CFC, MIFC

Real Options Valuation, Inc.

R R I S S K
R R I S S K

REAL OPTIONS VALUATION, INC.



本手冊及其中涉及的軟體由 Real Options Valuation, Inc 公司授權提供，用戶須在遵守《最終用戶許可協定》條款的情況下使用和複製。本文件中所提供的資訊僅供參考，隨時可能更改，不另行通知，不代表 Real Options Valuation, Inc 公司的商業義務。不管出於何種目的，在沒有 Real Options Valuation, Inc 公司書面授權情況下，任何人不得以任何形式、任何方式，包括影印和錄音等方式複製或傳播本手冊的任何章節。本文參考了 Real Options Valuation, Inc 公司創始人及首席執行官 Dr. Johnathan Mun 的出版物。由 Dr. Johnathan Mun 編寫。在美國境內編寫設計出版。如想購買本手冊的補充讀物，請按以下的 e-mail 地址聯繫 Real Options Valuation, Inc 公司：

admin@real-consulting.com or visit www.realoptionsvaluation.com.cn. © 2005 -2012 by Dr. Johnathan Mun. 公司保留一切解釋權利。Microsoft®是微軟公司在美國境內及其它國家的註冊商標。下文中涉及的其他產品可能是屬於相關擁有者的商標和/或註冊商標。

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

Real Options Valuation, Inc.

4101F Dublin Blvd., Ste. 425

Dublin, California 94568 U.S.A.

Phone 925.271.4438 • Fax 925.369.0450

admin@realoptionsvaluation.com

www.risksimulator.com

www.realoptionsvaluation.com



目錄

1. 系統簡介	6
1.1 歡迎使用 Risk Simulator 軟體.....	6
1.2 安裝	7
1.3 授權過程.....	7
1.4 2011/2012 版有哪些新增功能？	10
1.4.1 一般功能	10
1.4.2 模擬模組	11
1.4.3 預測模組	12
1.4.4 優化模組	13
1.4.5 分析模組	13
1.4.6 統計學和 BizStats 模組	14
2. Monte Carlo 模擬	16
2.1 什麼是 Monte Carlo 模擬？	16
2.2 開始學習 Risk Simulator	17
2.2.1 軟體的總體介紹	17
2.2.2 運行 Monte Carlo 模擬.....	17
2.3 相關性和精度控制	27
2.3.1 相關性的基礎知識	27
2.3.2 在 Risk Simulator 中應用相關性.....	28
2.3.3 相關性對 Monte Carlo 模擬的影響	29
2.3.4 精度和誤差控制	31
2.3.5 對預測統計量的理解.....	33
2.3.6 瞭解 Monte Carlo 模擬的概率分佈	36
2.4 離散分佈.....	39
2.5 連續分佈.....	46

3. 預測	66
3.1 不同類型的預測方法.....	66
3.2 運行 Risk Simulator 中的預測工具	68
3.3 時間序列分析	68
3.4 多元回歸.....	71
3.5 隨機預測.....	76
3.6 非線性外推法	78
3.7 Box-Jenkins ARIMA 高級時間序列.....	80
3.8 自回歸求和移動平均模型(Box-Jenkins ARIMA 高級時間序列).....	85
3.9 基本計量經濟學模型.....	86
3.10 J-S 曲線預測.....	87
3.11 GARCH 波動率預測.....	88
3.12 馬爾可夫鏈.....	91
3.13 最大似然估計模型 (MLE)	91
3.14 樣條模型 (三次樣條內插和外插模型).....	93
4. 最佳化	96
4.1 最佳化方法.....	96
4.2 連續型決策變數的最佳化.....	98
4.3 離散整數型最佳化	103
4.4 有效邊際和高級最佳化選項	106
5. 風險模擬分析工具	108
5.1 模擬中的颶風和敏感性分析工具	108
5.2 敏感性分析.....	115
5.3 分佈擬合：單變數和多元變數	118
5.4 Bootstrap 模擬 (拔靴法)	121
5.5 假設檢驗.....	124
5.6 資料輸出和保存模擬結果.....	125
5.7 創建報告.....	126
5.8 回歸和預測診斷工具.....	128
5.9 統計分析工具	135

5.10 分佈分析工具.....	138
5.13 Risk Simulator 2011/2012 新工具	143
5.14 亂數產生，蒙特卡羅與拉丁超立方，以及相關性 Copula 方法	143
5.15 資料季節性因素和趨勢性因素消除	144
5.16 主成分分析.....	145
5.17 結構性中斷分析	146
5.18 趨勢線預測.....	147
5.19 模型檢驗工具.....	148
5.20 百分位分佈擬合工具	149
5.21 分佈圖表：概率分佈工具.....	151
5.22 ROV BizStats	154
5.23 神經網路和組合模糊邏輯預測方法	158
5.24 優化器目標搜尋	160
5.25 單變數優化.....	161
5.26 遺傳演算法優化	162
5.27 ROV 決策樹模組	164
5.27.1 決策樹	164
5.27.2 模擬模型	166
5.27.3 貝葉斯分析.....	166
5.27.4 精確資訊的預計數值，最大最小分析，風險檔，不精確資訊的數值。	167
5.27.5 敏感性分析.....	167
5.27.6 情景表	167
5.27.7 生成效用函數.....	168
6. 幫助貼士和技巧.....	176
貼士：假設（設定輸入假設使用者介面）	176
貼士：複製和粘貼	176
貼士：相關性.....	177
貼士：資料診斷和統計分析	177
貼士：分佈分析，圖表.....	177
貼士：有效邊際.....	177
貼士：預測儲存格	178

貼士：預測圖表.....	178
貼士：預測.....	178
貼士：預測：ARIMA	178
貼士：預測：基本計量經濟學.....	178
貼士：預測：Logit，Probit 以及 Tobit.....	178
貼士：預測：隨機過程.....	178
貼士：預測：趨勢線.....	179
貼士：函式呼叫.....	179
貼士：入門練習和視頻.....	179
貼士：硬體 ID.....	179
貼士：拉丁超立方抽樣（LHS）vs Monte Carlo 模擬.....	179
貼士：線上資源.....	180
貼士：優化.....	180
貼士： 文檔.....	180
貼士：右鍵快速鍵或者其他快速鍵	180
貼士：保存.....	181
貼士： 抽樣和模擬技術.....	181
貼士：軟體卡發工具（SDK）和動態連結程式庫	181
貼士：在 Excel 中使用 Risk Simulator.....	181
貼士：極速模擬.....	181
貼士：颶風圖分析	182
貼士：Troubleshooter	182



1. 系統簡介

1.1 歡迎使用 Risk Simulator 軟體

Risk Simulator 版本是一套集 Monte Carlo 模擬，預測和最佳化為一體的軟體。本軟體使用 microsoft.NET #C 作為開發工具，並附加具備 Excel 功能。此外本軟體還與 Real Option Valuation 公司的 SLS 軟體和 ESOV 軟體相容，可同時使用。注意，儘管我們努力讓本手冊包含所有的內容，但是它不能代替配套的 DVD，培訓課程以及軟體發展者的相關著作（例如：Dr.Johnathan Mun 的 *Real Option Analysis* 《實物期權分析》，第二版，Wiley Finance 2005；*Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《風險建模：應用 Monte Carlo 模擬，實物期權分析，預測及最佳化》，第二版，Wiley 2006/2010；*Valuing Employee Stock Options (2004 FAS 123R)*，Wiley Finance 2004。）更多相關資訊請訪問我們的網站：www.realoptionsvaluation.com.cn。

Risk Simulator 套裝軟體含以下幾個模組：

- Monte Carlo 模擬模組（利用不同的模擬文檔、截距和相關性模擬、常態分佈、控制置信度和誤差的模擬以及其他一些運算法則對 42 種概率分佈進行參數和非參數模擬）
- 預測模組（進行 Box-Jenkins ARIMA、多元回歸、非線性推斷、隨機過程和時間序列分析）
- 不確定情況下的最佳化模組（利用離散整數和連續變數對有模擬和無模擬的有價證券和專案進行最佳化）
- 建模和分析工具模組（進行颶風圖、蜘蛛圖和敏感性分析，同時還有內部抽樣模擬，假設核對總和分佈擬合等）

Real Options SLS 軟體可以用於計算簡單和複雜的期權，同時具備創造自定義期權模型的功能。**Real Options SLS** 軟體包括以下幾個模組：

- 單資產 SLS（用於解決放棄期權，選擇期權，延遲期權，擴展期權以及自定義期權）
- 多重資產和多階段 SLS（用於解決多階段連續期權，包含多重資產和

多個階段，多階段期權與放棄、選擇、收縮、延遲、擴展及轉換等期權的結合，也可以解決自定義的期權問題)

- 多叉 SLS (用於解決三叉均值回復期權，四叉跳躍-擴散期權以及五叉彩虹期權)
- Excel 增益集功能 (在 Excel 環境下解決上述所有期權模型及閉合模型和自定義期權模型)

1.2 安裝

跟隨螢幕上的指示安裝本軟體。本軟體的最低要求如下：

- Pentium IV 處理器及以上
- Windows XP/Vista/Windows 7
- Windows Excel XP, 2003, 2007, 2010 版及以上
- Microsoft .NET Framework 2.0/3.0
- 500M 空間, 推薦 2GB 記憶體
- 安裝軟體的管理員許可權

現在大多數新電腦都已經自帶安裝了 Microsoft .NET Framework 2.0/3.0 元件。但是，如果在安

裝 Risk Simulator 的過程中出現錯誤提示，要求安裝.NET Framework，那就要首先退出安裝程式，然後從 CD 裏（可以自己選擇語言）安裝相關的.NET Framework 元件。完成之後重啓電腦，再重新安裝 Risk Simulator 軟體。

1.3 授權過程

我們的軟體有一個 30 天的試用期。如果想到得到完整的公司授權，請聯繫 Real Option Valuation 公司 郵箱：admin@real-consulting.com or admin@realoptionsvaluation.com 或者撥打電話（925）271-4438，還可以訪問我們的網站：www.realoptionsvaluation.com.cn。

如果您已安裝本軟體，並購買了使用該軟體的授權，您需要將您的硬碟碼用電子郵件的方式發送給我們，我們將會為您生成所需的授權文件。具體可參照如下指導：

Windows XP 作業系統，使用 Excel XP，Excel 2003, Excel 2007 和 Excel 2010：首先，在 Excel 中單擊模擬授權，[並將得到的包括字母和數位的 11-20 位元硬碟碼發送 Email 到 admin@real-consulting.com or admin@realoptionsvaluation.com](#)。一旦收到您的硬碟碼，我們將會為您生成一個永久使用軟體的授權文件，並將該授權文件發送 Email 給您。當您得到該授權文件後，下載存放到您的硬碟上，然後運行 Excel，點擊模擬授權，並

單擊**安裝授權**，載入授權文件。然後重新啟動 Excel 就完成了。整個過程只需短短的幾分鐘，您就可以成為授權用戶來使用該軟體了。

Windows Vista / Windows 7 使用 Excel XP, Excel 2003 或 Excel 2007 和 Excel 2010：首先，在 Windows Vista/Windows 7 中打開 Excel2007/2010，並選中模擬功能表欄，單擊授權圖示或模擬授權，[將得到的包括字母和數位的 11-20 位元硬碟碼發送 Email 到 admin@real-consulting.com](mailto:admin@real-consulting.com) or admin@realoptionsvaluation.com。一旦收到您的硬碟碼，我們將會為您生成一個永久使用軟體的授權文件，並將該授權文件發送 Email 給您。當您得到該授權文件後，下載存放到您的硬碟上。運行 Excel，點擊模擬授權或授權圖示，並單擊**安裝授權**，載入授權文件。

然後重新啟動 Excel 就完成了。整個過程只需短短的幾分鐘，您就可以成為授權用戶來使用該軟體了。

安裝完成之後，開啓 Microsoft Excel，如果安裝成功的話，會看到 Excel 的功能表列中會出現一項額外的“模擬”項，還有一個新的圖示，如下圖 1.1。除此以外，還會出現下圖 1.2 中的程式啟動畫面，表示軟體被啟動並載入到 Excel 中。圖 1.3 是 Risk Simulator 的工具列。如果可以在 Excel 中找到以下專案，那您現在就可以開始使用本軟體了。接下來的章節將會一步步地介紹如何使用本軟體。

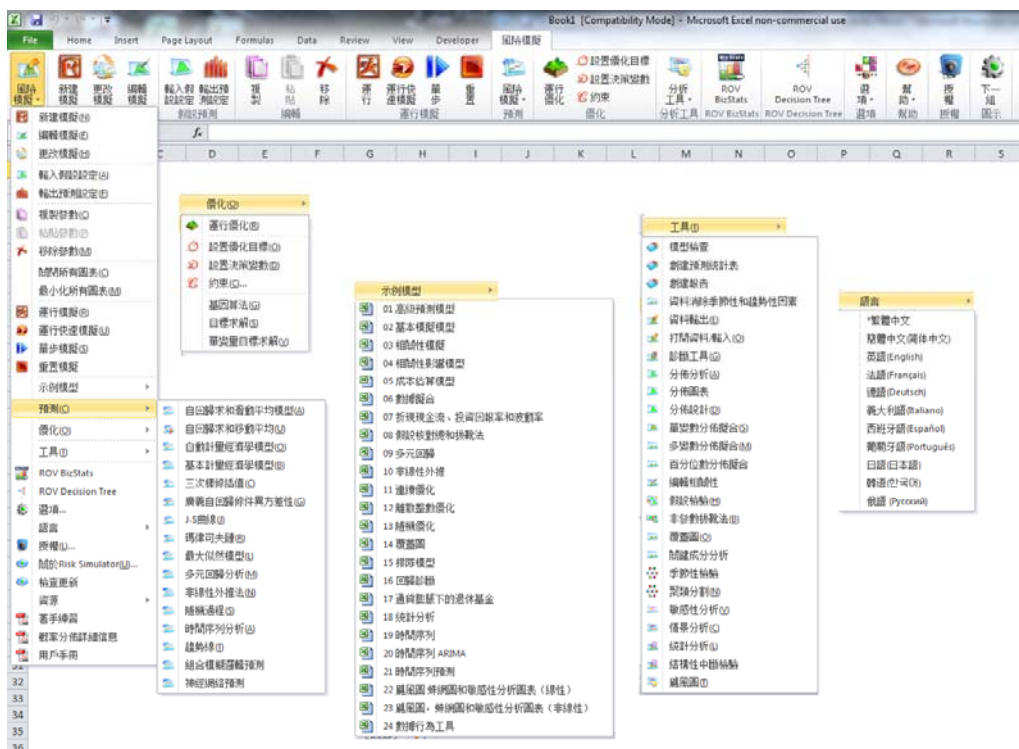


圖 1.1 ——Risk Simulator 功能表和圖示



圖 1.2 ——Risk Simulator 啟動畫面

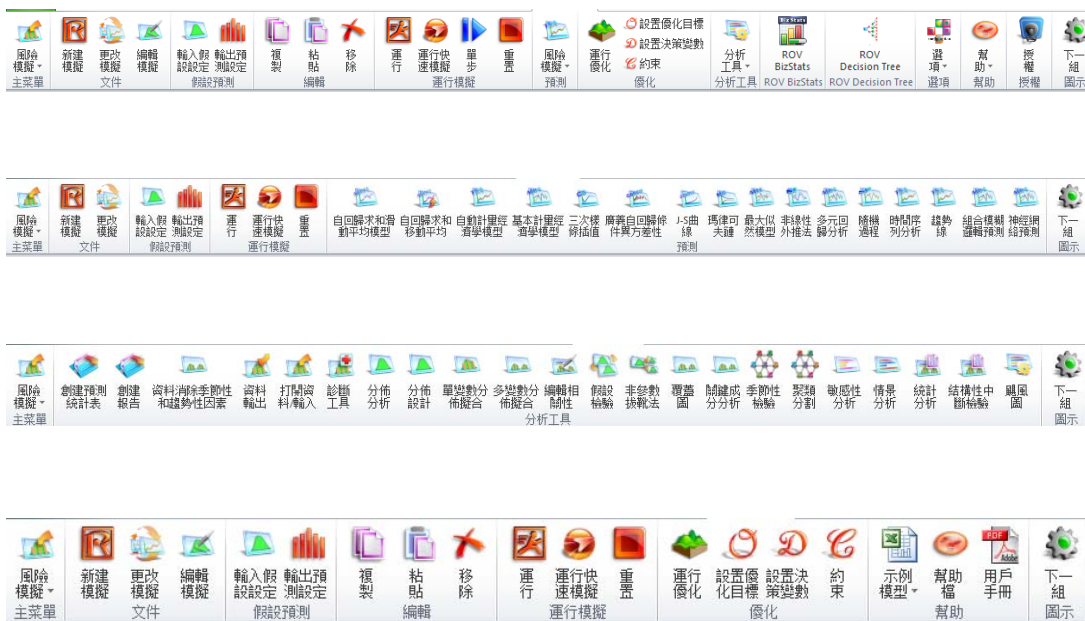


圖 1.3 ——Risk Simulator 工具欄

1.4 2011/2012 版有哪些新增功能？

Risk Simulator 功能完整清單

以下清單列出了 Risk Simulator 的主要功能，強調顯示的部分是 Risk Simulator 2011/2012 版增的最新功能。

1.4.1 一般功能

1. 支援 11 種語言 - 英語，法語，德語，義大利語，日語，**韓語**，葡萄牙語，西班牙語，簡體中文，**俄語**和**繁體中文**。
2. ROV 決策樹軟體用於創建和評估決策樹模型。高級方法和分析包括：
 - 決策樹模型
 - 蒙特拉羅模擬
 - 敏感性分析
 - 情景分析
 - 貝葉斯定理(連接及後可能性更新)
 - 期望數值資訊
 - 極大極小
 - 極大極小
 - 風險文件
3. 書籍 - 由 10 本書籍提供分析理論，應用和案例研究支援。
4. 批註儲存格 - 打開或關閉儲存格的批註和決定你是否想在輸入假設，輸出預測和決策變數的儲存格顯示批註。
5. 詳細的示例模型 - Risk Simulator 裡面的 24 種示例模型和 Modeling Toolkit 裡面的超過 300 個模型。
6. 詳細報告 - 所有的分析包含都詳細報告。
7. 詳細的用戶手冊 - 分步介紹的用戶手冊。
8. 靈活的授權方式 - 你可以打開或關閉的軟體的某些功能，能夠讓你定制你的風險分析經驗。例如，如果你只對 Risk Simulator 的預測工具感興趣，您可以獲取一個特定的授權來啟動預測工具模組而不啟動其它的模組從而節省一些軟體費用。
9. 靈活的軟體需求 - 適用於 Window 7，Vista 和 XP 系統；與 Excel 2010，2007，2003 和 Mac 作業系統上運行的虛擬機器。
10. 完全可定制的颜色和圖表 - 傾斜，立體，顏色，圖表類型，以及更多！

11. 動手練習 - 詳細 Risk Simulator 的分步使用指南，包括解釋結果的指南。
12. 多個儲存格複製和粘貼 - 允許對輸入假設，決策變數，預測進行複製和粘貼。
13. 在一個模型中創建多個模擬文檔 - 允許在一個模型中創建多個模擬文檔（在一個模型中可以創建，複製，編輯和運行模擬文檔）。
14. 在 Excel 2007/2010 中修訂了的圖示 -- 完全重新設計的圖示工具列，更加直觀和用戶友好。將有四套圖示來匹配大部分的螢幕解析度（1280 × 760 及以上）。
15. 右鍵快速鍵 - 可以右擊滑鼠來進入 Risk Simulator 的工具和功能表。
16. ROV 軟體集成 - 與其他 ROV 軟體相容的軟體包括，包括 Real Options SLS， Modeling Toolkit， Basel Toolkit， ROV Compiler， ROV Extractor 和 Evaluator, ROV Modeler， ROV Valuator， ROV Optimizer， ROV Dashboard， ESO Valuation Toolkit， 以及更多！
17. Excel 中的 RS 函數功能 - 在 Excel 中插入 RS 功能來進行輸入假設和預測設置，並支援用滑鼠按右鍵應用。
18. 疑難排解：此工具允許您重新啟用該軟體，檢查您的系統要求，獲取硬體 ID，等等。
19. 高速分析：這種新的功能能夠高速運行預測和其它分析（在版本 5.2 基礎上提升）。分析和結果與就版本仍是相同的，但現在計算和報表生成的速度更快。
20. 網路資源，案例研究和視頻 - 免費下載模式，開始使用視頻，案例研究，白皮書，和我們網站的其它材料。

1.4.2 模擬模組

21. 6 種隨機生成器 - ROV 高級減法生成器，減法隨機打亂生成器，長期打亂生成器，可轉移隨機打亂生成器，快速 IEEE Hex 打亂生成器，基本最小可轉移生成器。
22. 兩種抽樣方法 - 蒙特卡羅和拉丁超立方。
23. 3 種相關性 Copula 函數 - 應用正態度 Copula 函數，T Copula 函數，和相關模擬的擬正態 Copula 函數。
24. 42 種概率分佈 - 反正弦，伯努利，Beta，Beta 3，Beta 4，二項，柯西，卡方，余弦，自訂，離散均勻，雙對數，Erlang，指數，指數 2，F 分佈，伽瑪，幾何，Gumbel 分佈最大，最小 Gumbel 分佈，超幾何，拉普拉斯 (Laplace)，邏輯，對數正態分佈 (算術) 和對數正態分佈 (對數)，對數正態分佈 3 (算術) 和對數正態分佈 3 (邏輯)，負二項，正態，拋物，帕累托，巴斯卡，皮爾遜 5，皮爾森 6，PERT 法，泊松，功率，功率 3，瑞利 (Rayleigh)，T 和 T2，三角，均勻，韋伯，

韋伯 3。

25. 替代參數 - 使用輸入百分位作為輸入參數的替代方法。
26. 自訂非參數分佈 - 定制自己的分佈，運行歷史模擬，和運用德爾菲法。
27. 分佈截斷 - 設定資料的界限。
28. Excel 函數 - 使用 Excel 函數來設定假設和預測。
29. 多維模擬 - 對不確定性輸入參數的模擬。
30. 精密控制 - 確定模擬試驗次數是否足夠。
31. 超高速模擬- 在幾秒鐘內運行 100,000 次模擬。

1.4.3 預測模組

32. ARIMA 模型 - 自回歸整合移動平均模型 ARIMA (P, D, Q)。
33. 自動 ARIMA 模型 - 運行 ARIMA 模型最常見的組合，找到最佳的擬合模型。
34. 自動計量經濟學 - 運行數以千計的模型組合及排列獲取對現存資料擬合最佳的模型（線性，非線性，相互作用，滯後，超前，按速率變化，差異）。
35. 基本計量經濟學 - 計量經濟學和線性/非線性和相互作用回歸模型。
36. 三次樣條 - 非線性插值和外推。
37. GARCH 模型 - 適用廣義自回歸條件異方差模型來預測波動率：GARCH, GARCH-M, TGARCH, TGARCH - M, EGARCH, EGARCH-T, GJR - GARCH, 以及 GJR - TGARCH。
38. J 曲線 - 指數 J 曲線。
39. 有限相依變數 - Logit 模型, Probit, 與 Tobit。
40. 瑪律可夫鏈 - 隨著時間推移的兩競爭因素和市場佔有率預測。
41. 多元回歸 - 常規線性和非線性回歸，與逐步方法（向前，向後，相關性，向前-向後）。
42. 非線性外推 - 非線性時間序列預測。
43. S 曲線 - 邏輯 S 曲線。
44. 時間序列分析 - 8 個時間序列分解模型，預測基數，趨勢和季節性型。
45. 趨勢線 - 使用線性，非線性多項式，指數，對數，指數和移動平均法進行預測和擬合，並有擬合優度。
46. 神經網絡預測（線性，邏輯，雙曲正切，雙曲正切的餘弦）。

1.4.4 優化模組

- 47. 線性優化 - 多相優化和一般線性優化。
- 48. 非線性優化 - 詳細的結果包括 Hessian 矩陣，拉格朗日函數和更多。
- 49. 靜態優化 - 快速運行連續，整數離散和二元制優化。
- 50. 動態優化 - 模擬與優化。
- 51. 隨機優化 - 二次，切，中，向前，收斂標準。
- 52. 有效前沿 - 多元有效前沿的隨機和動態優化組合。
- 53. **遺傳演算法** - 適用於各種優化問題。
- 54. 多相優化 - 局部與全域優化測試，從而更好地控制優化如何運行，提高了結果的準確性和相互關係。
- 55. 百分位和條件均值 - 隨機優化的其它統計資訊，包括百分位元以及條件均值，這對計算在險價值很關鍵。
- 56. **搜索演算法** - 對於單個決策變數和目標搜索應用是一種簡單，快速，有效的演算法。
- 57. 動態和隨機優化中的超高速模擬 - 與優化結合的超高速模擬。

1.4.5 分析模組

- 58. **模型檢查** - 測試你的模型中最常見的錯誤。
- 59. 相關性編輯器 - 允許直接輸入和編輯大量相關性矩陣。
- 60. 創建報告 - 自動生成模型裡面的假設和預測報告。
- 61. 建立統計量報告 - 生成所有預測統計量比較報告。
- 62. 資料診斷 - 運行異方差，微數缺測性，離群，非線性，自相關，正態，球形，非平穩性，多重共線性和相關性測試。
- 63. 資料提取和匯出 - 提取資料到 Excel 或文字檔和 RiskSim 檔，運行統計報告和預測結果報告。
- 64. 資料打開和導入 - 重新檢索以前的模擬運行結果。
- 65. **去季節化和去趨勢化** - 對您的資料去季節化和去趨勢化處理。
- 66. 分佈分析 - 對 42 種分佈計算精確的 PDF，CDF 和 ICDF，並生成概率分佈表。
- 67. 分佈設計器 - 創建自己的定制分佈。
- 68. 分佈擬合（多個變數） - 同時運行多個變數的分佈擬合併考慮相關性和相關性的影響。
- 69. 分佈擬合（單個變數） - 對連續分佈進行柯爾莫哥洛夫，斯米爾諾夫

和卡方檢驗，生成完整的報告和分佈假設。

70. 假設檢驗 - 測試兩個統計量是否統計顯著相同或者有差異。
71. 非參數拔靴法 - 對統計量進行模擬，以獲取結果的精度和準確性。
72. 覆蓋圖 - 對假設和預測生成完全定制的覆蓋圖 (CDF, PDF, 2D/3D 圖表類型)。
73. **主成分分析** - 測試最佳預測變數和方法，以減少資料的陣列。
74. 情景分析 - 可以生成數以萬計的靜態二維情景。
75. 季節性測試 - 測試各種季節性滯後。
76. 分割聚類 - 分割你的資料，將資料集納入不同的統計類別。
77. 敏感性分析 - 動態敏感性測試 (同時進行分析)。
78. **結構突變測試** - 測試你的時間序列資料是否有統計結構突變。
79. 龍捲風分析 - 敏感性的靜態擾動，蜘蛛網分析和龍捲風分析，情景分析表格。

1.4.6 統計學和 BizStats 模組

80. **百分位分佈擬合** - 使用百分位和優化，以找到最佳擬合分佈。
81. **概率分佈圖表和表格** - 運行 45 種概率分佈，計算它們的四個矩，CDF, ICDF, PDF, 圖表，疊加多個分佈圖，並生成概率分佈表。
82. 統計分析 - 描述性統計，分佈擬合，長條圖，圖表，非線性外推，正態性測試，隨機參數估算，時間序列預測，趨勢預測等。
83. **ROV BIZSTATS** - 超過 130 種商業統計和分析模型:

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

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

2

2. Monte Carlo 模擬

以著名的摩納哥賭城命名的 Monte Carlo 模擬是一種非常有效的方法。模擬可以輕而易舉地為從業者解決那些困難複雜的現實問題。Monte Carlo 通過生成成百上千種甚至上百萬種試驗路徑的結果來類比和分析未來可能發生的事件，並觀察它們的普遍共性。作為公司的分析師，僅僅學習大學水平的數學課程是不合乎實際的。一名優秀的分析師會盡可能地利用其所有可利用的工具，用最簡單和最實用的方法來得到相同的答案。在所有案例之中，只要模型正確，Monte Carlo 模擬與其他數學上優美的方法一樣，也可以提供近似的答案。現在我們的問題是什麼是 Monte Carlo 模擬以及它是如何運作的？

2.1 什麼是 Monte Carlo 模擬？

簡單來說 Monte Carlo 模擬就是一個用於預測、估計和風險分析的亂數發生器。模擬就是通過反復地從一個事先確定的變數的概率分佈中挑選出一些數值，並將這些數值用於模型來計算一個模型的無數種情景。由於所有這些情景在模型中產生關聯的結果，每個情景可以產生一種預測。預測通常是包含公式或者函數的事件，它被定義為模型的一個重要輸出量。這些事件通常是諸如總量、淨利潤和總支出等。

可以簡單地將 Monte Carlo 模擬方法想象成從一個大籃子裏取出高爾夫球然後放回的重復過程。籃子的大小和形狀取決於**輸入假設**的分佈（例如，一個均值為 100，標準差為 10 的常態分佈相比於均勻分佈或三角分佈），那些有深度並且比較對稱的籃子會令到某些球被取出的頻數高於其他球，重復取出球的次數取決於模擬試驗的次數。對於一個很大，有著多重相關性的模型，將它想象成一個超級大的籃子，裏面有很多小籃子。每一個籃子裏面都有一堆高爾夫球跳來跳去。有時候這些小籃子會手挽著手（在這些變數之間存在相關性的情況下），這時有些高爾夫球會一前一後跳動，其他一些會彼此獨立跳動。在存在這些交互作用的情況下，將每次取出的球的編號輸入表格，這就提供了一份模擬的預測輸出結果。

2.2 開始學習 Risk Simulator

2.2.1 軟體的總體介紹

Risk Simulator 套裝軟體含幾種不同的應用模組，具體有 Monte Carlo 模擬，預測以及最佳化。

- 模擬應用模組允許您在基於 Excel 的模型中進行模擬，生成模擬預測（結果的分佈），進行分佈擬合（自動尋找最優的統計分佈），計算相關性（變數之間的關係），確定敏感度（製作颶風圖和敏感度圖表），同時還可以運行自定義和非參數模擬（利用歷史資料，但不需指定分佈或者參數的模擬）。
- 預測應用模組可以用於生成 Box-Jenkins ARIMA 計量經濟預測，自動時間序列預測（包含季節性和趨勢），多元回歸（線性和非線性回歸），非線性外推（曲線擬合），隨機過程（隨機遊走，均值回歸以及跳躍擴散過程）。
- 最最佳化應用模組用於多元離散整數變數、連續變數和混合決定變數，在某個約束條件下，最大化或最小化某個目標。並且可以和 Monte Carlo 模擬一起用於靜態最最佳化或動態最最佳化，以及在不確定情況下的隨機最最佳化，還可以用於解決線性和非線性最最佳化問題。
- Real Options SLS 是與 Risk Simulator 軟體相獨立又相互補充的一款軟體，可用於解決多種簡單或複雜的實物期權問題。

2.2.2 運行 Monte Carlo 模擬

一般情況下，爲了在已有 Excel 模型下運行模擬，需要遵循以下步驟：

1. 新建一個模擬文檔或打開一個現有文檔
2. 在相關的單格定義輸入假設
3. 在相關的單格定義輸出假設
4. 運行模擬
5. 結果解析

如果想要試驗一下，可以打開名叫“基本模擬模型”的示例文件，然後跟隨下面的例子

來進行模擬。這個示例文件可以在開始功能表的目錄下找到：開始|Real Options Valuation|Risk Simulator|示例。

新建一個模擬
文檔

爲了開始一次新的模擬，您必須首先創建一個模擬文檔。這個模擬文檔包含一整套關於

如何進行一次模擬的指導說明，也就是所有的假設、預測、運行選項等等。這個文件簡化了創建多個模擬的過程。也就是說利用一個相同的模型可以創建幾個文檔，每個都對應自己的模擬特性和要求。同一個人可以利用不同的分佈假設和輸入量來創建不同的測試情景，多個人也可以在同一個模型下根據他們自己的假設量和輸入量進行測試。

- 打開 Excel 表格，創建一個新的或打開一個現有的模型（您可以使用基本模擬模型的例子來繼續）
- 點擊**模擬**圖示，選擇**新建模擬**
- 為模擬設定一個標題，並填入其他相關資訊（圖 2.1）

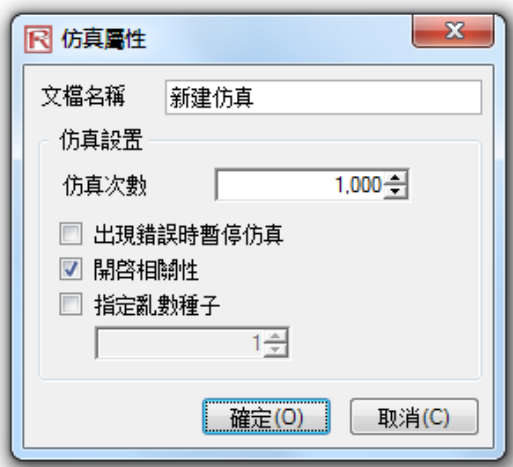


圖 2.1——新建模擬文檔

- 標題：** 對模擬文檔進行命名，可以在同一個 Excel 模型中創建多個模擬文檔。這意味著可以在不取消現存假設和不因新模擬情景的需要而改變假設的情況下，在同一模型中保存不同的模擬情景的文檔。之後您也可以任意改變文檔名（**模擬編輯模擬**）。
- 試驗次數：** 這裏要求輸入的是需要模擬試驗的次數。進行 1000 次試驗意味著會產生基於輸入假設的 1000 個不同的結果。可以根據需要改變試驗次數，但是鍵入的數值必須是正整數，系統默認的運行次數是 1000 次。同時可以利用置信度和誤差控制來自動幫助決定需要進行模擬的次數（更多詳情請參考**置信度和誤差控制**章節）。
- 模擬錯誤時停止：** 如果選中此項，那麼在 Excel 模型中每次出現錯誤時模擬就會停止。也就是說，如果您的模型遇到一個計算錯誤（例如，在某次模擬試驗中可能會有某張電子資料出現輸入值除以零的錯誤），模擬就會停止。這點對於審核您的模型以確保在 Excel 模型中沒有任何計算錯誤很有幫助。當然，如果您對自

己的模型有足夠的信心，那麼也就沒有必要選這一項了。

開啓相關性：

如選中這一選項，就會考慮輸入假設之間的相關性。否則，相關性就會被設置為零，模擬會在輸入假設間不存在相關性的假設下進行。舉例來說，如果存在正相關的話應用相關性會得到更精確的結果，但如果存在負相關性的話，得到的預測可信度就會下降。如果在此處開啓相關性的話，稍後您可以在生成的每個假設處設置相應的相關係數（更多詳情請參見*相關性*章節）。

設置亂數種子：

每次模擬所得到的結果都會有細微的差別。這是由 Monte Carlo 模擬中亂數據生成方法的特點決定的，這也是所有亂數據生成器的理論事實。但是在作報告的時候，可能需要相同的結果（尤其是當報告要展示一系列結果時，在報告過程中想要展示相同的生成結果；或是當與別人採用的是相同的模型時，希望每次都得到相同的結果），那麼就選中這一選項，然後設置一個種子值。這個種子值可以是任意正整數。使用相同的初始亂數種子，相同的試驗次數，相同的輸入假設，進行模擬後，通常會生成相同序列的亂數，以保證最終結果的相同。

注意一旦創建一個新的模擬文檔，您可以在任何時候返回去修改這些設置。但是要保證當前文件是您想要修改的文檔。否則，點擊**模擬更改模擬**，選擇您想要修改的那個文檔，點擊**確定**（下圖 2.2 有一個例子顯示當前有幾個文檔，如何啟動想要修改的文檔）。然後點擊**模擬編輯模擬**，作相應的修改。同時還可以複製或重命名現存的文檔。

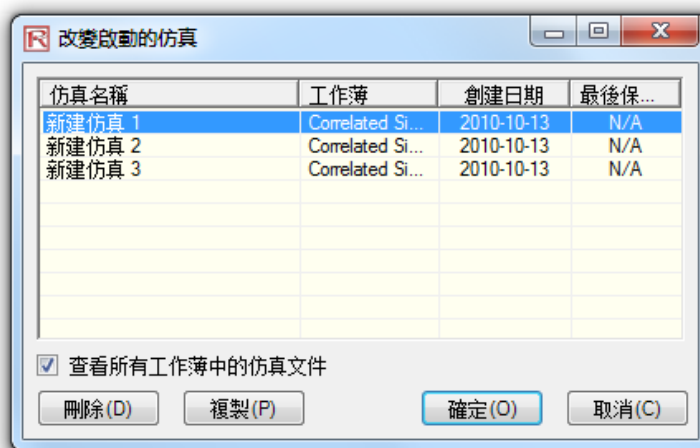


圖 2.2 ——修改當前模擬

定義輸入假設

接下來這步是在您的模型中設置輸入假設，注意假設不能以任何等式或函數的形式分配到單格中。例如，在模型中只能對數值單格設定輸入假設，反之只能對含有等式或函數的單格設定輸出預測。記住只有在模擬文件已經存在的條件下才可以設置假設和預測。按照下列步驟來設置模型中新的輸入假設：

- 確保模擬文檔已經存在，您可以打開一個現有文檔，或是新建一個文檔（**模擬|新建模擬**）
- 選擇您想要設置輸入假設的單格（例如，在**基本模擬模型**例子中的 G8 單格）
- 點擊**模擬輸入假設設定**或是點擊 Risk Simulator 工具欄上的第三個圖表
- 選擇想要的相關分佈，並鍵入相關的分佈參數，點擊**確定**將這些輸入假設插入到模型中（圖 2.3）

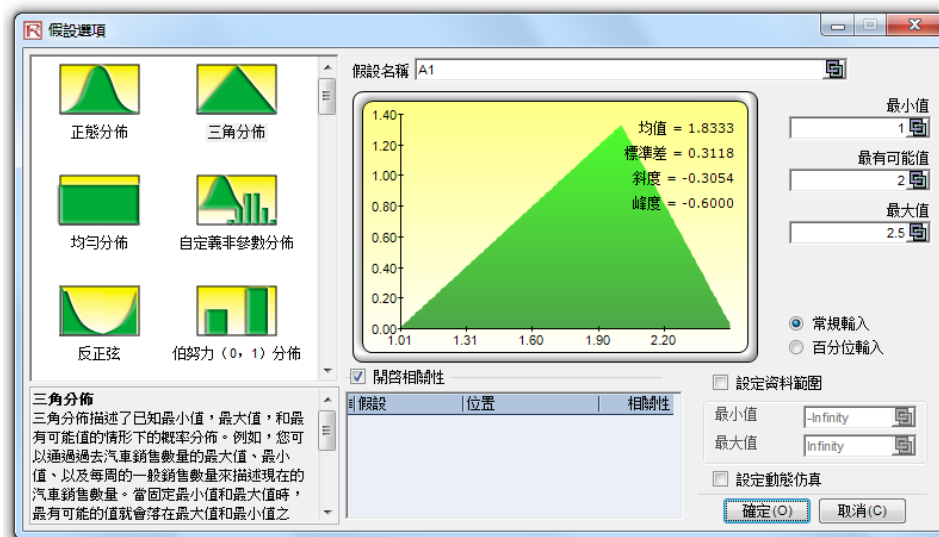


圖 2.3 設定輸入假設

注意，在假設屬性中有幾個需要留意的地方。下圖 2.4 指出了這些地方：

- **假設名稱**：這塊可選擇的功能允許您為假設輸入一個獨特的名稱，以便區分不同的假設，好的模型通常使用精簡的假設名稱。
- **分佈圖庫**：左邊這塊區域列示了軟體中所有可用的分佈。可以通過右擊圖庫的任何地方，選擇大圖表，小圖表或列表來變換視圖，有超過 45 種的分佈可以使用。
- **輸入參數**：基於選定的分佈來確定所需的參數。您既可以直接輸入參數，也可以鏈結到具體的工作表中。在輸入參數需要可見或是允許改變（點擊鏈結圖示將輸入參數與工作表單格鏈結）的情況下，鏈結到工作表單格比較適用。
- **分佈邊界**：一般情況下分析師是不會使用的，它們只是用於縮小分佈假設的範圍。例如，如果選擇常態分佈，那麼它的理論邊界就是正無窮和負無窮之間。然而在實際情況中，模擬變數只在一個小範圍記憶體在，這個範圍可以用來適當縮減分佈範圍。
- **相關性**：可以將輸入假設設置為一對相關量。如果假設需要的話，記住點擊**模擬編輯模擬**來開啓相關性選擇。更多詳情請參考本章末的關於相關性設置和其對模型

影響的討論。

- **簡短描述：**圖庫裏的每個分佈都有一個簡短描述。它說明了何時使用該分佈，以及輸入參數的要求。參見瞭解 Monte Carlo 模擬中的概率分佈章節關於軟體中可用的分佈類型的詳細介紹。

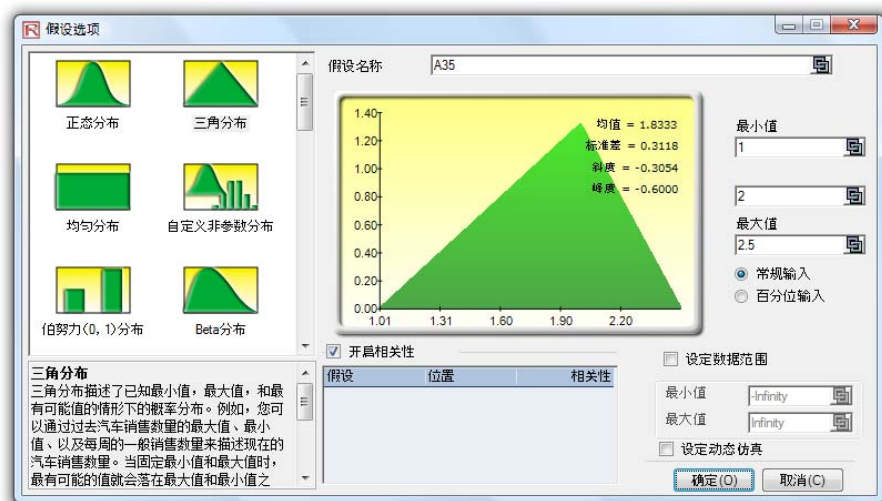


圖 2.4 假設屬性

注意：如果繼續本例的話，在 G9 單格中設定其他的假設。使用最小值為 0.9，最大值為 1.1 的均勻分佈。然後，在下個步驟中定義輸出預測。

定義輸出 預測

下一步是在模型中定義輸出預測，輸出預測只能設置在含有等式或函數的單格上。以下是設定預測的程式：

- 選擇想要設置假設的單格（例如，基本模擬模型例中的單格 G10）
- 點擊**模擬**，選擇**輸出預測設定**或是點擊 Risk Simulator 工具欄上的第四個圖示（圖 1.3）
- 鍵入相關的資訊按**確認**鍵。
- 下圖 2.5 說明了預測屬性的設置
- 預測名稱：定義預測單格，這點非常重要，因為當您的模型很大並且包含很多預測單格時，單獨命名預測單格有助於您迅速找到正確的結果。不要小瞧這一步驟的重要性，最好是使用精確簡短的假設名稱。
- 預測準確性：您可以通過設置置信水平和誤差控制而不是根據推測來

估計所需的試驗次數。當模擬達到所需的誤差-置信水平結合的要求之後，模擬就會停止並告知您所達到的置信水平，通過這樣的方式可以自動地獲得試驗次數，不需要我們去估計它的數值。更多詳情請參見 *誤差和置信度控制* 章節。

- 顯示預測視窗：允許使用者顯示或不顯示具體的預測視窗，系統的默認選項是顯示預測視窗。

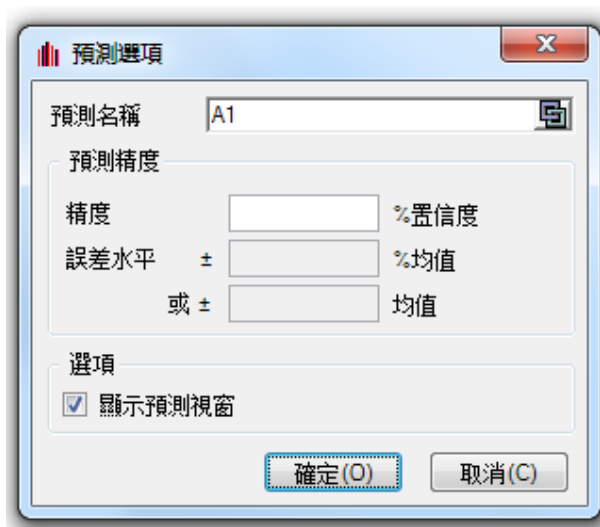


圖 2.5 輸出預測設定

運行模擬

如果一切準備就緒，就可以點擊**模擬|運行模擬**或是點擊**運行**圖示（Risk Simulator 工具欄上的第八個圖示）開始模擬。還可以在運行之後重新設置模擬再運行一次（**模擬|重置模擬**或是工具欄上的第十個圖示），或是在運行過程中中止它。單步功能（**模擬|單步模擬**或是工具欄上的第九個圖示）可以允許您進行一次模擬試驗，一次一個，這對於初學者來說很有用（例如，可以在每次試驗中，將所有假設單格的值替換掉，並重新計算整個模型）

預測結果解析

Monte Carlo 模擬的最後一個步驟就是對預測結果圖進行解釋說明。圖 2.6 到圖 2.13 是運行模擬之後生成的預測圖和統計表，以下是在解釋模擬結果時需要說明的重要因素：

- 預測圖：圖 2.6 中的預測圖顯示的是整個模擬試驗中數值出現頻數的概率直方圖。豎條代表的是某個具體的 x 值在總試驗次數中出現的次數，累積頻數（一條光滑的曲線）代表的是不大於 x 值的所有數值在預測中出現的總概率。
- 預測統計量表：圖 2.7 中的預測統計表根據分佈的四矩對預測值的分佈進行了總結。更多關於這些統計指標意義的解釋請參見 *瞭解預測統計量* 章節。通過按空白鍵您可以在直條圖和統計量表之間進行切換。

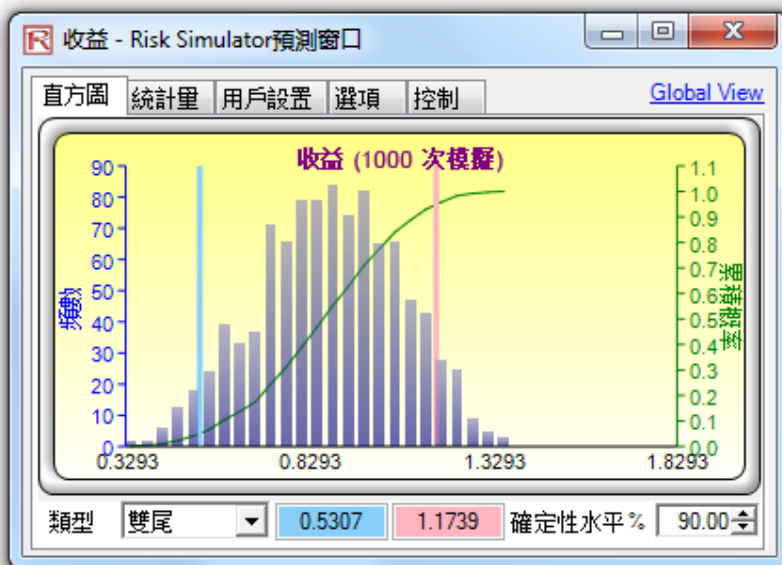


圖 2.6 預測圖

統計表	結果
模擬次數	1000
均值	0.8626
中位數	0.8674
標準差	0.1933
方差	0.0374
變異係數	0.2241
極大值	1.3570
極小值	0.3019
極差	1.0551
偏度	-0.1157
峰度	-0.4480
25分位數	0.7269
75分位數	1.0068
95%置信度	1.3888%

圖 2.7 預測統計表

用戶設置:

預測統計表裏的用戶設置標籤可以讓您改變圖表的外觀。例如，如果選擇**總在最前端顯示**，那麼不管您的電腦裏有什麼別的軟體在運行，預測圖表都是可見的。直方圖清晰度允許您改變直方圖的豎條數量，從 5 到 100 之間的任意數均可。此外，**資料更新**板塊允許您控制模擬的運行速度(相對於預測圖表的更新頻數)。也就是說，如果您想要每次試驗時都更新預測圖表，那就要犧牲模擬的速度，因為用於模擬的記憶體將被分配用於更新圖表。這僅僅是一種用戶選擇，對模擬的結果不會有任何影響，只是影響了模擬的完成速

度。爲了進一步增加模擬的運行速度，您可以在模擬過程中將 Excel 表格最小化，這樣減少了更新 Excel 電子資料表所需的記憶體，並釋放了運行模擬的記憶體。**關閉所有視窗及最小化**能控制所有打開的預測圖表。

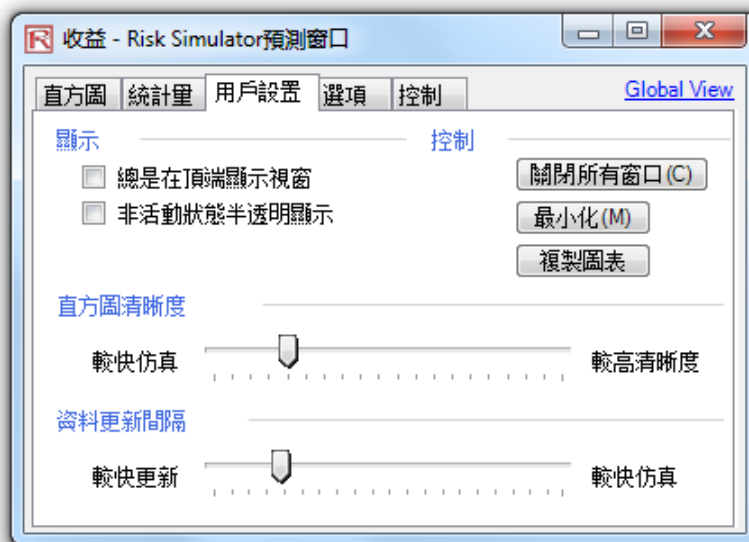
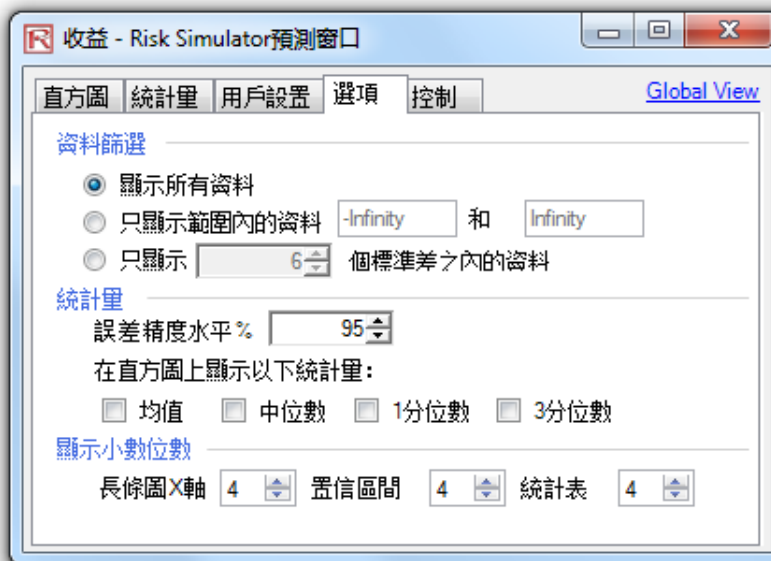


圖 2.8 預測圖參數選擇

選項：

預測圖表裏的選項功能允許您顯示所有的預測資料或過濾在某一區間或某一標準差間隔內的輸入/輸出資料。此外，這裏還可以爲此預測設置置信度，以便從統計學的觀點顯示誤差水平。更多詳情參見**誤差和置信度控制**章節，如果想要在預測圖表中列出均值，中位數，第一分位線以及第四分位線（第 25 百分點和第 75 百分點），那麼顯示下面這些統計量是一個很好的選擇。



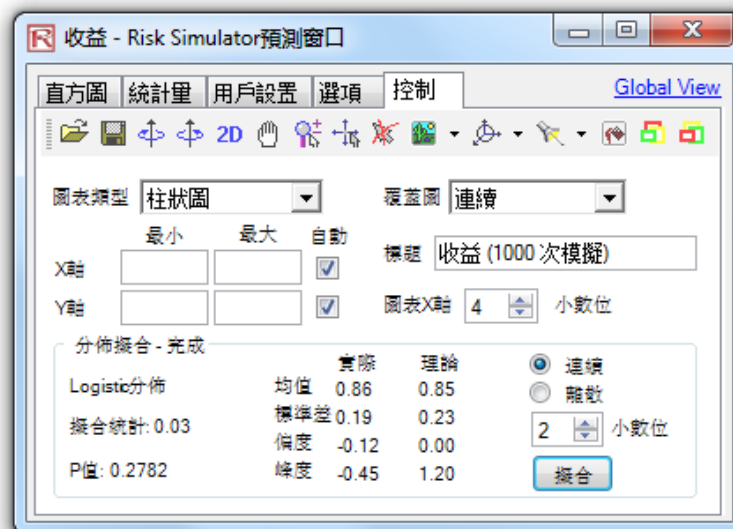


圖 2.9 預測圖表選項

使用預測圖表和置信區間

在預測圖表中，您可以確定事件發生的概率也就是通常所稱的置信區間。給定兩個值，模擬結果會落在這兩個值中間區域的概率是多少？圖 2.10 中可以看出最終結果（在本例中是收入水平）會落在\$0.5307 和\$1.1739 之間的概率是 90%。置信區間可以這樣計算：首先類型選擇雙尾，然後鍵入所需的置信度（例如 90），按下鍵盤上的 TAB 健，對應此置信度的兩個數值就被計算出來了。在本例中，收入低於\$0.5307 的概率是 5%，高於\$1.1739 的概率也是 5%。這意味著，雙尾分佈的置信區間是以中值或五十分位點數值為中心的一個對稱區域。因此，雙尾具有相同的概率分佈。

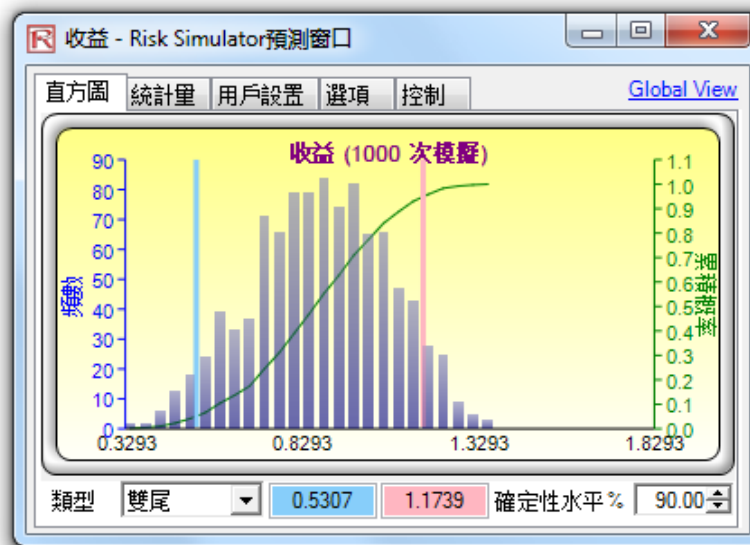


圖 2.10 雙尾置信區間預測圖

作為選擇，也可以使用單尾概率。圖 2.11 是一個置信度 95% 的左尾概率分佈（例如類型選擇左尾，在置信度處鍵入 95，按下鍵盤上的 TAB 鍵）。這意味著收入低於 \$1.1739 的概率是 95%，而高於 \$1.1739 的概率是 5%，與圖 2.10 中的情況很吻合。

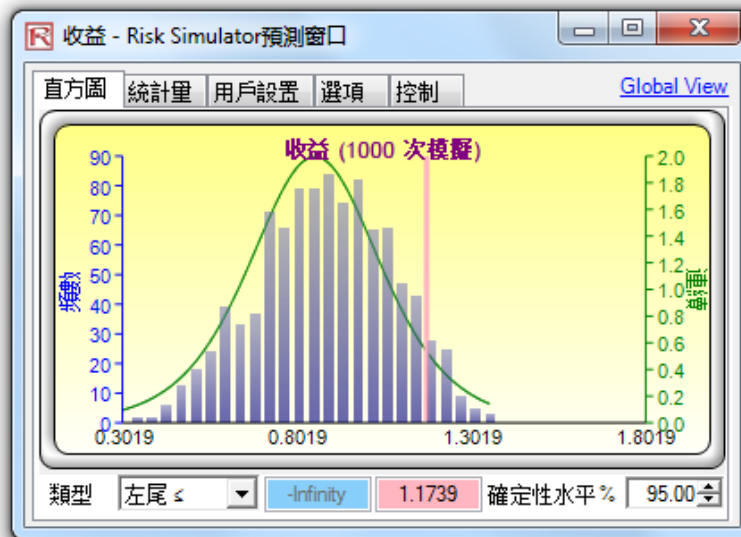


圖 2.11 單尾置信區間預測圖

除了可以估計預測區間（例如給出一個概率水平，確定相應的收入值），您還可以確定獲得某個收益範圍的概率。例如，收入小於 \$1 的概率是多少？為了計算這個數值，您可以在類型中選擇左尾，在數值框中鍵入 1 然後敲 TAB 鍵。相應的置信度就會被計算出來（在本例中，收入低於 \$1 的概率為 74.30%）。

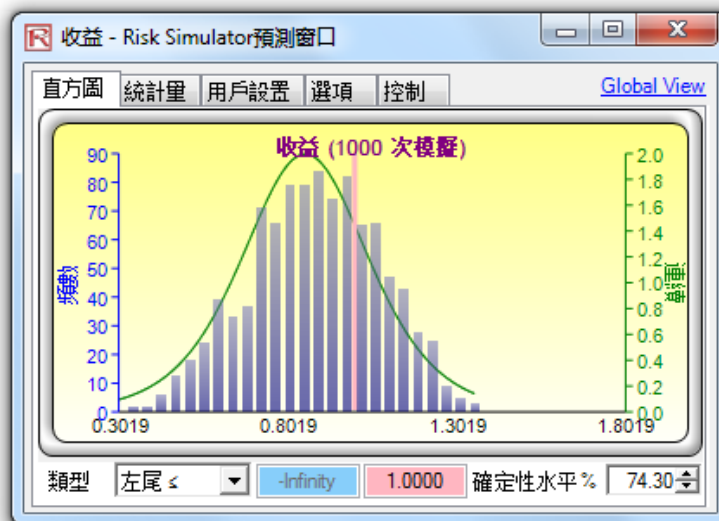


圖 2.12 預測圖概率估計

爲了完整起見，您可以在概率類型中選擇右尾，在數值框中鍵入 1 然後敲 TAB 鍵。得出的結果是大於 1 的右尾概率，即收入大於\$1 的概率（在本例中，我們可以看出此概率是 25.70%）。

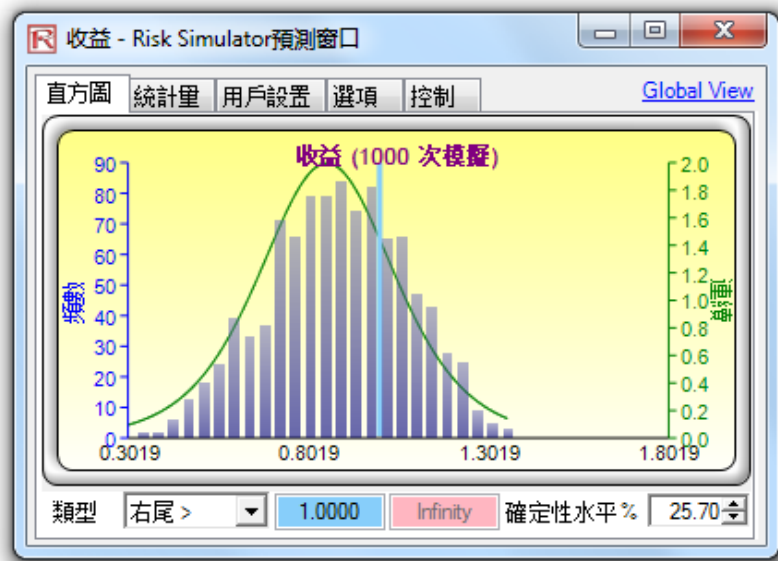


圖 2.13 預測圖概率估計

小貼士：

可以通過拖動預測窗口的右下角來調整預測窗口的大小。通常比較明智的做法是在再次運行模擬之前，重置現有的模擬（**模擬重置模擬**）。記住當需輸入數值或左右尾數值時，敲擊鍵盤上的 TAB 鍵來更新資料和結果。

2.3 相關性和精度控制

2.3.1 相關性的基礎知識

相關係數測量的是兩個變數之間相關性的強弱和方向，它可以是從-1.0 到 +1.0 之間的任何數值。相關係數可以分解爲符號（兩變數是正相關還是負相關）和相關性的大小或強弱（相關係數的絕對值越大，兩者之間的相關性越強）。

相關係數的計算方法有幾種。第一種方法是利用以下的公式來手動計算變數 x 和 y 之間的相關係數 r：

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

第二種方法是使用 Excel 的 CORREL 函數。例如，如果 x 和 y 各有 10 個資料，分佈在單格 A1 : B10 區域，可以直接使用函數 CORREL (A1 : A10 , B1 : B10)。

第三種方法是運行 Risk Simulator 多元擬合工具來計算。

重點注意：相關並不意味著存在因果關係。兩個完全無關的隨機變數也可能呈現出一定的相關性，但是這並不意味著它們之間有任何的因果關係（例如，太陽黑子與股市有一定的相關性，但是它們之間並不存在因果關係）。

存在兩種類型的相關：參數相關和非參數相關。Pearson 相關係數是最常用的相關量度標準，通常簡稱為相關係數。但是 Pearson 相關是一種參數測量方法，要求兩個相關的變數都屬於常態分佈，並且變數之間的關係是線性的。但是在 Monte Carlo 模擬中，不符合這種條件的例子經常出現，此時非參數相關就顯示出其重要性了。非參數相關主要有 Spearman Rank 相關和 Kendall' s Tau 相關。Spearman Rank 相關使用最普遍並且很適合應用於 Monte Carlo 模擬中——因為它不要求常態分佈和線性關係，所以對不同分佈變數的相關也適用。為了計算 Spearman Rank 相關，首先將變數 x 和 y 的值排序，然後再利用 Pearson 相關性來計算。在 Risk Simulator 的例子中用到的相關大部分是非參數 Spearman Rank 相關。但是為了簡化模擬過程並與 Excel 表中的相關函數保持一致，軟體所要求的相關輸入是 Pearson 相關係數。Risk Simulator 然後會利用自身的運算法則將它們轉換成 Spearman Rank 相關，這樣就簡化了整個過程。但是，為了簡化用戶介面，我們允許使用者鍵入更常見的 Pearson 積差相關係數（例如，利用 Excel 表的相關函數計算），但在數學編碼中，我們會將這些簡單的相關轉換成 Spearman Rank 相關再進行分佈模擬。

2.3.2 在 Risk Simulator 中應用相關性

相關可以通過幾種方式應用到 Risk Simulator 中：

- 在定義假設（**模擬設置輸入假設**）時，只需將相關係數填入分佈圖的相關距陣網格中。
- 利用現有資料運行多元擬合工具（**模擬工具|分佈擬合|多元變數**）來進行分佈擬合，得到成對變數之間的相關係數矩陣。如果模擬文件已經存在，那麼對應的假設會自動包含這些相關係數。
- 存在假設之後，您可以在用戶介面上直接點擊**模擬編輯相關係數**來鍵入所有假設的相關係數。

注意，相關係數矩陣必須是正定矩陣。也就是相關係數在數學上必須是有效的。例如，假設您想要將以下三個變數相關聯：畢業生某年的成績，他們每

周的啤酒消費量，他們每周學習的小時數。則必須假設存在以下的相關關係：

成績和啤酒：負相關 喝的越多，成績越低（不會在考試中表現出來）

成績和學習時間：正相關 花在學習上的時間越多，成績越高

啤酒和學習時間：負相關 喝的越多，花在學習上的時間越少（一直醉酒和聚會）

如果您將成績和學習之間的關係定義為負相關，並且假設相關係數的數值很高，那麼相關矩陣就是非正定的了。它不符合邏輯、相關性要求和數學上矩陣的一些常識。但是，即使邏輯性很差，小的相關係數有時還是存在的。當您鍵入非正定或錯誤的相關矩陣時，Risk Simulator 會自動提示您，並在維持原有相關關係整體結構的基礎上為您提供調整後的半正定相關係數矩陣（同樣的符號和大小）。

2.3.3 相關性對 Monte Carlo 模擬的影響

儘管在模擬中對相關係數變數進行的計算很複雜，但是結果卻是很清晰的。圖 2.14 顯示的是一個簡單相關性模型（示例文件夾中的相關性影響模型）。收入的計算很簡單，用價格乘以數量。同樣的模型將會同時應用於價格和數量之間不存在相關性，存在正相關（+0.8），存在負相關（-0.8）這三種情況進行比較。

相關模型	相關模型		
	不相關	正相關	負相關
價格	\$2.00	\$2.00	\$2.00
數量	1.00	1.00	1.00
收益	\$2.00	\$2.00	\$2.00

圖 2.14 簡單相關模型

圖 2.15 是運行結果的統計資料。注意不存在相關性模型的標準差是 0.23，相比之下，正相關的標準差是 0.30，負相關的是 0.12。在簡單模型中，負相關傾向於縮小分佈的波動範圍，生成的預測分佈更加緊湊集中，相比較而言，正相關的波動範圍更大些。但是均值大致是保持穩定的。這意味著相關性對專案的預期值影響不大，但是會降低或增加專案的風險。

收入負相關性 - Risk Simulator 預測窗口

直方圖 統計量 用戶設置 選項 控制 [全視圖](#)

統計表	結果
模擬次數	1000
均值	1.9976
中位數	1.9961
標準差	0.0717
方差	0.0051
變異係數	0.0359
極大值	2.2148
極小值	1.8197
極差	0.3951
偏度	0.1040
峰度	-0.3191
25分位數	1.9435
75分位數	2.0486
95%置信度的百分誤	0.2224%

收入無相關性 - Risk Simulator 預測窗口

直方圖 統計量 用戶設置 選項 控制 [全視圖](#)

統計表	結果
模擬次數	1000
均值	2.0036
中位數	1.9995
標準差	0.1450
方差	0.0210
變異係數	0.0724
極大值	2.3907
極小值	1.6844
極差	0.7063
偏度	0.0304
峰度	-0.7316
25分位數	1.8944
75分位數	2.1127
95%置信度的百分誤	0.4486%

收入正相關性 - Risk Simulator 預測窗口

直方圖 統計量 用戶設置 選項 控制 [全視圖](#)

統計表	結果
模擬次數	1000
均值	2.0020
中位數	1.9992
標準差	0.1886
方差	0.0356
變異係數	0.0942
極大值	2.4147
極小值	1.6278
極差	0.7869
偏度	0.0788
峰度	-0.9641
25分位數	1.8468
75分位數	2.1475
95%置信度的百分誤	0.5839%

圖 2.15 相關結果

圖 2.16 所示的是運行模擬之後的結果，這裏展示了提取的假設的原始資料，計算了變數之間的相關係數。模擬結果中的資料隱含了輸入假設。也就是說，如果您鍵入+0.9 和-0.9 的相關係數，模擬出的結果會具有相同的相關係數。

以下是從模擬中提取出來的原始資料。然後，對它們進行相關性檢驗，檢驗它們之間的相關性是否等同於我們實際輸入模型的相關性係數。Pearson 相關係數是線性參數相關，結果顯示，我們輸入的相關係數（+0.80和-0.80）的確是變數之間的相關係數。詳情請參見Dr. Johnathan Mun 的” Modeling Risk “ (Wiley 2006) 。

價格正相關	品質正相關		價格負相關	品質負相關	
1.95	0.91		1.89	1.06	
1.92	0.95		1.98	1.05	
2.02	1.04	Pearson相關	1.89	1.09	Pearson相關
2.04	1.03		1.88	1.04	
1.89	0.91	0.80	1.96	0.93	-0.80
1.98	1.05		2.02	0.93	
2.05	1.03		2.00	1.02	
1.87	0.91		1.86	1.04	

圖 2.16 相關係數恢復

2.3.4 精度和誤差控制

Monte Carlo 模擬中比較有效的工具之一是精度控制。例如，運行一個複雜的模型需要多少次試驗才足夠呢？精度控制會自動確定所需的試驗次數，當達到預先設定的精度水平時模擬就會停止。

精度控制功能允許您自己設置想要的精度。一般來說，試驗的次數越多，置信區間就越窄，統計資料也越精確。Risk Simulator 裏的精度控制功能利用置信區間的特徵來確定是否達到了某統計量的精度水平。對於每個預測，您都可以指定具體的相應精度的置信區間。

不要混淆這三個不同的概念：誤差，精度和置信度。儘管它們聽起來差不多，但是三者有很大的區別。下面是對它們的一個簡單解釋。假設您是一位玉米麵豆卷包裝製造商，想要找出平均每箱中 100 個包裝裏的損壞量是多少。一種方法是收集事先包裝好的一箱 100 個玉米麵豆卷包裝，打開箱子計算有多少是破損的。您每天可以生產一百萬箱（這是您的總體），您只隨機打開 10 箱（這是您的樣本大小，就是我們所說的模擬中的試驗次數）進行檢查。每箱裏破損的包裝數量統計如下：24，22，4，15，33，32，4，1，45，2。計算出來的平均破損量是 18.2。基於這十個樣本的試驗，平均值是 18.2，80% 的置信區間是從 2 到 33（也就是說在這個樣本空間或試驗次數條件下，破損量在 2 到 33 之間的可能性是 80%）。但是您能在多大程度上確定 18.2 就是正確的平均數呢？10 次試驗足以證明這一點嗎？從 2 到 33 的置信區間範圍太廣，不確定性太大。假設您想要一個更加精確的平均值，90% 的置信度和誤差控制在±2 個玉米麵豆卷包裝以內——意思是如果您打開一天生產的所有一百萬箱產品，九十萬箱產品的平均破損量都在某個平均值的±2 幅度波動。要達到這種精度水平，您需要選擇多少箱玉米麵豆卷包裝作為樣本（或試驗次數）呢？這裏的 2 代表誤差水平，90% 是精度水平。如果試驗次數足夠多，由於對均值的估計更加精確，那麼 90% 的置信區間的估計也就更加精確。舉

一個例子，假設平均值是 20，那麼 90%的置信區間就是 18 到 22，這個區間的精度是 90%，也就是在所有打開的一百萬個箱子中有九十萬箱的破損數量在 18 到 22 個之間。要達到這個精度所需的試驗次數可以利用樣本的誤差公式 $\bar{x} \pm z \frac{s}{\sqrt{n}}$ 計算： $z \frac{s}{\sqrt{n}}$ 代表的是誤差 2， \bar{x} 是樣本均值，代表由 90%的精度水平得到的標準正態 Z 值，s 是樣本標準差，n 是滿足特定誤差和精度水平所需的試驗次數。圖 2.17 和圖 2.18 解釋了精度控制是如何運用於 Risk Simulator 的多元模擬預測中的，這個功能使用戶不需要根據自己猜測來決定試驗的次數。

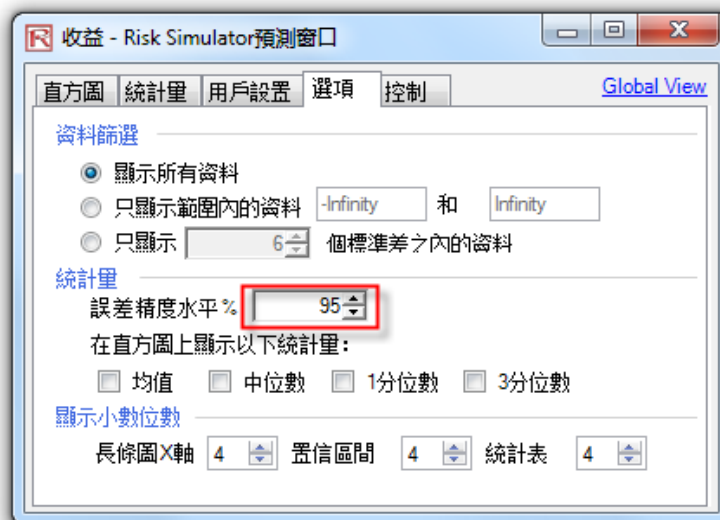


圖 2.17 設置預測的精度水平

統計表	結果
模擬次數	1000
均值	0.8626
中位數	0.8674
標準差	0.1933
方差	0.0374
變異係數	0.2241
極大值	1.3570
極小值	0.3019
極差	1.0551
偏度	-0.1157
峰度	-0.4480
25分位數	0.7269
75分位數	1.0068
95%置信度的百分誤	1.3888%

圖 2.18 誤差的計算

2.3.5 對預測統計量的理解

大多數分佈都可以通過四個矩的計算來描述。第一矩描述分佈的位置或集中趨勢（期望值），第二矩描述分佈的寬度或幅度（風險），第三矩是分佈的偏度方向（最可能事件），第四矩是分佈的尖峰性或尾部的厚度（突發損失或收益）。在實際專案的分析過程中，所有這四個部分都要被考慮到以便提供充分全面的觀點。Risk Simulator 從統計學觀點出發，在預測圖表中提供了所有這四部分的結果。

測量分佈的中心值——第一矩

分佈的第一矩描述的是某個專案的期望回報率。它測量了專案的定位和可能的平均回報。第一部分常用的統計量包括均值（平均值），中位數（分佈的中心）和眾數（最可能出現的值）。圖 2.19 顯示了第一矩——在本例中，是以均值（ μ ）或平均數來測量的。

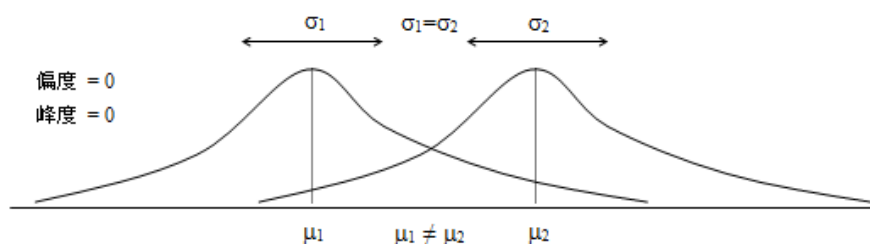


圖 2.19——第一矩

測量分佈的範圍——第二矩

分佈的第二矩描述的是分佈的幅度，是風險的測量。分佈的幅度或寬度描述了變數的可變性，即是變數會落在分佈不同區域的可能性——也就是結果的可能範圍。圖 2.20 顯示的是兩個分部，它們有著相同的第一矩（同樣的均值），但是第二矩或風險是不同的。圖 2.21 中可以看得更加清楚。舉例來說，假設有兩隻股票，第一隻股票（黑色線）的波動很小，比較而言，第二隻股票（虛線）的價格波動很大。因為風險大的股票與風險小的股票相比其結果不可預知性將加大，很明顯投資者會認為波動更大，股票風險更大。圖 2.21 的縱軸代表股票價格，因此大風險股票的波動範圍更寬。在圖 2.20 中這個範圍是以分佈的寬度（橫軸）表示的，寬的分佈代表風險大的資產。因此，分佈的寬度或幅度代表變數的風險。

注意到在圖 2.20 中，兩個分佈有相同的第一矩或集中趨勢，但是很明顯兩個分佈是不同的。分佈寬度的差別是可以測量的。從數學和統計學的角度來說，一個變數的寬度或風險可以通過幾個不同的統計變數來計算，它們包括範圍，標準差，方差，變異係數和百分位數。

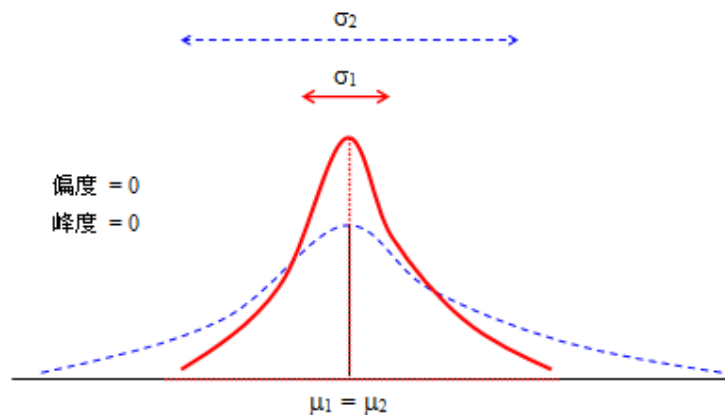


圖 2.20——第二矩

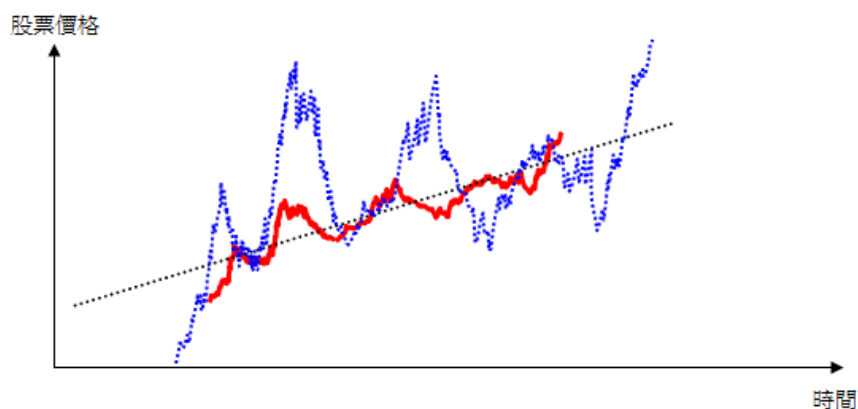


圖 2.21 股票價格波動

測量分佈的偏度——第三矩

分佈的第三矩測量了分佈的偏度，就是分佈偏方和幅度。圖 2.22 顯示的是負偏或左偏（分佈的尾部偏向左邊），圖 2.23 顯示的是正偏或右偏（分佈的尾部偏向右邊）。均值一般偏向分佈的尾部，而中位數是保持不變的。從圖中也可以看出均值是不同的，但中位數和方差可以是相同的。如果不考慮第三矩，只考慮期望值（中位數或均值）和風險（標準差），可能會錯誤地選擇一個正偏的專案！例如如果以橫軸表示專案的淨收益，那麼我們會更偏好左偏分佈，因為與較小值出現的可能性較高的分佈（圖 2.23）相比，這種分佈中大值出現的可能性更高（圖 2.22）。因此，在偏態分佈中，中位數是一個很好的測量回報的工具，由於圖 2.22 和 2.23 的中位數是相同的，風險也是相同的，那麼此時左偏分佈的淨收益就是一個較好的選擇。如果不考慮專案的分佈偏度的話，可能會導致您選擇錯誤的專案（例如兩個專案可能有相同的第一個矩和第二個矩，也就是兩個專案有相同的回報和風險，但是它們的分佈偏度是非常不一樣的）。

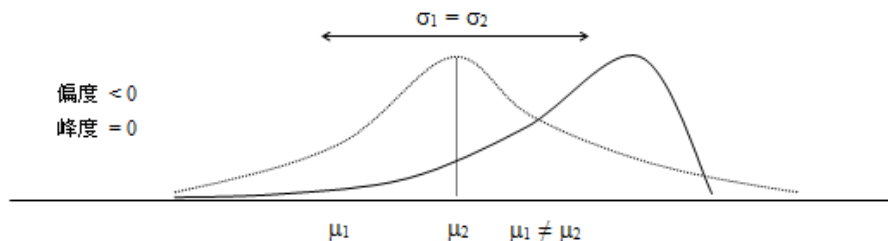


圖 2.22 第三個矩 (左偏)

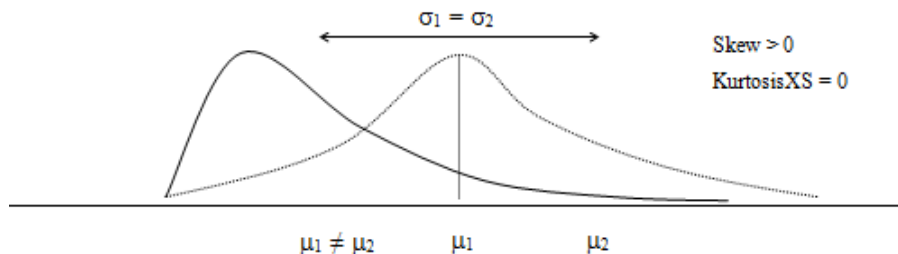


圖 2.23 第三矩 (右偏)

測量分佈中的
尾部突發事件
——第四矩

第四矩測量的是分佈的峰度，圖 2.24 顯示了這一效果。圖中（虛線）是一個峰度為 3.0 的或超峰度為 0.0 的常態分佈。Risk Simulator 的結果顯示了超峰度的值，用 0 來作為峰度的正常水平，這意味著負的超峰度代表比較扁平的尾部（與均勻分佈類似的低峰態分佈），正值代表比較肥胖的尾部（與 Student's T 分佈和對數常態分佈類似的尖峰態分佈）。粗線代表的分佈有著更大的超峰度，因此曲線尾部所覆蓋的區域比中心區域更多。這個因素對圖 2.24 中兩個分佈的風險分析有重大的影響，前三個矩（均值，標準差和偏度）是相同的，但是第四個矩（峰度）是不同的。這種情況表明儘管收益和風險相同，但對於高峰度的分佈來說極端和突發事件（可能出現的大的損失或收益）發生的可能性比較大（股票市場的收益就是尖峰度分佈，峰度值很大）。忽視專案的峰度可能是非常危險的。一般來說，高超峰度值暗示著分佈兩端的風險比較大（例如專案的在險價值（VaR）可能是非常重要的）。

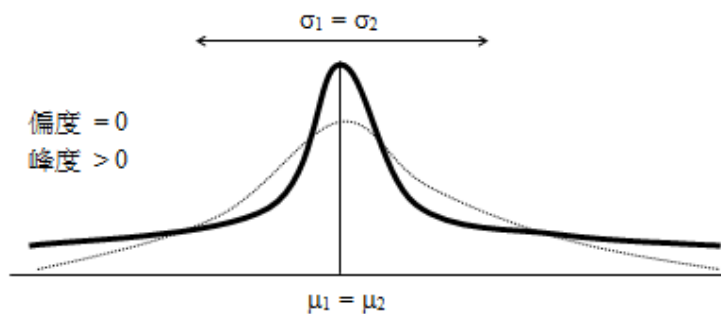


圖 2.24 第四矩

2.3.6 瞭解 Monte Carlo 模擬的概率分佈

本章節介紹了 Monte Carlo 模擬的作用，但是為了開始模擬，首先要瞭解概率分佈中的一些概念。在開始介紹概率之前，先來看看這個例子：您想要查看一家大公司某個部門裏未被免職人員的工資分佈。第一步您要搜集原始資料——在本例中是指此部門每個未被免職員工的工資。第二步將資料組織好繪成頻數分佈圖。為了畫頻數分佈圖，首先要將工資按照區間進行分組，並在圖的橫軸上列出這些區間。然後在縱軸上每個對應的區間標出員工的個數或頻數。現在您可以清晰地看到此部門的未免職員工的工資分佈圖。

圖 2.25 中顯示出大部分員工（大約 180 人裏有 60 人）每小時的工資是位於 \$7.00 到 \$9.00 之間。

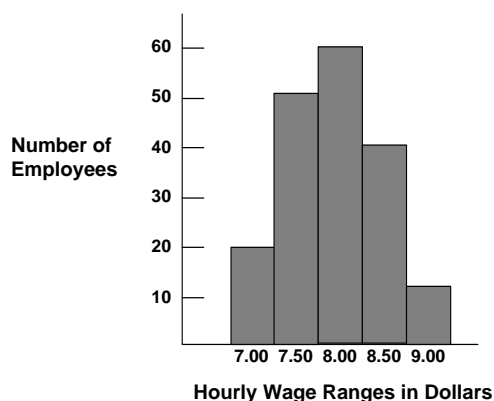


圖 2.25 頻數方柱圖 I

您還可以將這些資料繪成概率分佈圖。概率分佈代表的是每個區間的員工人數占總員工人數比重。為了製作這個概率分佈圖，您可以用每個區間的員工人數除以總員工人數，將得到的結果標在圖表的縱軸上。

圖 2.26 顯示的就是每組員工人數占總員工人數的比例；您可以估計從整個集合中隨機抽取一位員工，他的工資位於某一區間的可能性或概率。例如，假設在同樣的條件下抽取樣本，那麼從整個集合中隨機抽取一位員工其小時收入落在 \$8.00 到 \$8.50 之間的概率是 0.33（三個之中有一個）。

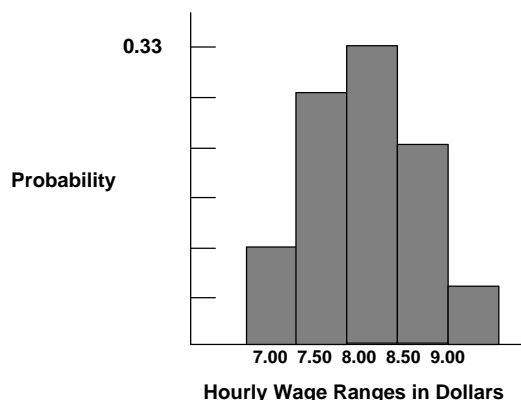


圖 2.26 頻數方柱圖 II

概率分佈可以是離散的，也可以是連續的。*離散概率分佈*描述的是獨立的值，通常是整數，沒有中間值，用一系列直方柱來表示。例如離散分佈可以用來描述四次拋硬幣過程中出現頭像的次數 0，1，2，3 或 4。*連續分佈*是一個比較抽象的數學概念，因為它假設在兩個數位之間存在中間值。也就是說連續分佈假設在分佈的任意兩點之間存在無數的數值。但是在很多情況下您可以用連續分佈來近似描述一個離散分佈，即使連續模型不能確切地描述這種情況。

選擇正確的概 率分佈

將資料繪製成點圖是選擇概率分佈的方法之一。以下的步驟提供了另外一種在試算表中為不確定變數選擇最佳概率分佈的方法。

為了選擇合適的概率分佈，請參照以下步驟：

- 查看需要討論的變數,列出已知的關於此變數的所有相關情況。可以從此未知變數的歷史資料中搜集資訊。如果得不到歷史資料，就基於以往的經驗做出判斷。
- 回顧關於概率分佈的描述。
- 選擇可以表現變數特徵的分佈。當分佈的特徵與變數的特徵相符時，就說這個分佈可以表現該變數的特徵。

Monte Carlo 模擬

簡單地來說 Monte Carlo 模擬就是一個亂數生成器，用來預測，估計和進行風險分析。模擬就是通過重複地從一個事先定義的不確定性變數的概率分佈中挑選出一些數值，並將這些數值用於模型來計算一個模型的眾多情景。由於所有這些情景在模型中都會產生關聯性的結果，每個情景可以有一種預測結果。預測通常包含公式和函數，是模型的一個重要輸出量。這些事件通常是總量、淨利潤和總成本等等。

簡單說來，蒙地卡羅模擬就像是從一個大籃子中不斷的抽取和放回高爾夫球。籃子的形狀和大小取決於分佈的假設（例如，均值為 100 標準差為 10

的常態分佈，均勻分佈或者三角分佈)。由於有的籃子較深，有的籃子較為對稱，讓某些球被更為頻繁地取到。重復取出球的次數取決於模擬的次數。對於包含相關性假設的更大的模型，就像是一個非常巨大的籃子，裏面還包含很多的小籃子。每個小籃子都包含一系列的高爾夫球。有的時候小籃子和小籃子之間會纏在一起（如果變數之間包含相關性），當前一個球被取到的時候，後一個球也有機會被取到。這些小球每次被取出，它們之間的交互作用被記錄和計算下來，作為模擬的預測結果。

利用 Monte Carlo 模擬，Risk Simulator 為每個概率分佈假設所生成的亂數都是完全獨立的。換句話說，某次試驗所選擇的亂數對下一組亂數的生成沒有任何影響。也可以使用蒙地卡羅抽樣技術在試算表中對現實世界進行情景類比。

2.4 離散分佈

以下是一份 Monte Carlo 模擬中可以使用的不同類型概率分佈的詳細介紹列表。為了便於用戶參考我們在附錄中也包含了此份列表。

伯努力分佈

伯努力分佈是一種離散的概率分佈，滿足該分佈的事件只包含兩種結果（例如，正面或反面，成功或失敗），這就是它也被稱為 $(0, 1)$ 分佈的原因。只進行一次二項分佈實驗的結果滿足伯努力分佈。它往往是任何複雜概率分佈的基礎。

例如：

- 二項分佈：多次試驗的伯努力分佈，計算的是總試驗次數中成功次數 x 的概率。
- 幾何分佈：多次試驗的伯努力分佈，計算的是第一次成功出現之前總的失敗次數。
- 負二項分佈：試驗次數較多的伯努力分佈，計算的是第 x 次成功出現之前總的失敗次數。

伯努力分佈的數學運算式如下：

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

或者

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

$$\text{均值} = p$$

$$\text{標準差} = \sqrt{p(1-p)}$$

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

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

成功的概率 p 是唯一的分佈參數。要注意伯努力分佈只有一次試驗，模擬值可能是 0 或 1 中的一個。

參數輸入要求：

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

二項分佈

二項分佈描述的是在固定次數的試驗中某個事件發生的次數，例如在十次拋硬幣的過程中頭像出現的次數或是選擇的 50 個物品中次品的數目。

條件

二項分佈的三個先決條件是：

- 每次試驗只可能出現兩種結果，並且它們之間是互斥的。
- 試驗之間是相互獨立的——第一次試驗的結果不會影響下一次試驗。
- 每次試驗中事件發生的概率是相同的。

二項分佈的數學運算式如下：

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

$$\text{均值} = np$$

$$\text{標準差} = \sqrt{np(1-p)}$$

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

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

分佈的參數有成功的概率 (p) 和總試驗次數 (n)。試驗的成功次數用 x 表示。注意成功概率值為 0 和 1 的情況價值不高，不需要進行模擬，所以在軟體中不允許這兩種情況。

輸入假設：

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

試驗次數 ≥ 1 且 ≤ 1000 ，必須為正整數 (對於大型試驗，則採用常態分佈，計算出的二項均值和標準差做為常態分佈的參數)。

離散均勻分佈

離散型均勻分佈是一個描述等可能發生事件的概率分佈，如描述 N 個事件，每個事件發生的概率相同。分佈類似於均勻分佈，但是是離散的而非連續的。

離散均勻分佈的數學運算式如下：

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

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

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

偏度 = 0 (分佈是完全對稱的)

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

參數輸入要求：

最小值 < 最大值，並且兩者必須均是整數（包含負整數和 0）

幾何分佈

幾何分佈描述了直到第一次成功出現試驗進行的次數，例如，在贏得轉盤遊戲之前所轉轉盤的次數。試驗的次數不是固定的，試驗不斷進行直到成功事件第一次出現，每次試驗的成功概率都是一定的。

條件

幾何分佈的三個先決條件是：

- 試驗次數是不固定的。
- 試驗要一直進行直到第一次成功出現。
- 在每次試驗中事件發生的概率是相同的。

幾何分佈的數學運算式如下：

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

$$\text{均值} = \frac{1}{p} - 1$$

$$\text{標準差} = \sqrt{\frac{1-p}{p^2}}$$

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

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

成功概率 p 是唯一的分佈參數。試驗的成功次數用 x 表示，只能取正整數。

參數輸入要求：

成功的概率 >0 並且 <1 （也就是 $0.0001 \leq p \leq 0.9999$ ）。注意成功概率值為 0 和 1 的情況意義不大，不需要進行模擬，所以在軟體中不允許這兩種情況。

超幾何分佈

超幾何分佈與二項分佈類似，都是描述在一個固定數目的試驗次數中，某一特定事件發

生的次數。區別在於二項分佈中的試驗是相互獨立的，而超幾何分佈的試驗會改變每一個後發試驗的概率，被稱為“不放回的試驗”。例如，假設已知一個裝有機器零部件的箱子裏有一些次品。您從箱子裏挑選了一件後，發現是次品，就將它從箱子裏移走。如果再從箱子裏取出另一個零部件，與第一次相比它是次品的概率就降低了，因為已經取走了一件次品。如果放回那件次品的話，概率還是相同的，整個過程就滿足二項分佈的條件了。

條件

超幾何分佈的先決條件有：

- 總的元素或專案（樣本空間）的數目是固定的，一個有限的總體。總體空間必須小於等於 1750。
- 樣本空間（或試驗次數）是總體的一部分。
- 在每次試驗後總體中最初已知的成功概率會發生變化。

超幾何分佈的數學運算式如下：

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

$$\text{均值} = \frac{N_x n}{N}$$

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

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

超峰度 = 峰度 - 3（具體函數形式比較複雜，在此不作具體描述）

參數輸入要求：

總體空間 ≥ 2 ，並且是整數

樣本空間 > 0 ，並且是整數

總體成功數 > 0 ，並且是整數

總體空間 $>$ 總體成功數

樣本空間 $<$ 總體成功數

總體空間 < 1750

負二項分佈

負二項分佈被用於類比基於既定成功次數 (R) 的額外實驗次數的概率分佈。例如，為了獲得一個 10 份訂單的機會，在給定每個電話的成功幾率的前提下，在打過 10 個電話之後，還需要打多少個電話？x 軸展示了需要打的額外的電話次數或者失敗的電話次數。試驗的次數並不固定，試驗繼續進行直到第 R 次成功之後，每次成功的概率一定。

條件

負二項分佈的三個先決條件是：

- 試驗次數是不固定的。
- 試驗要一直進行直到第 R 次成功出現。
- 每次試驗某類結果發生的概率是一個常數。

負二項分佈的數學運算式如下：

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

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

$$\text{均值} = \frac{r(1-p)}{p}$$

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

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

分佈參數有成功的概率 (p) 和成功試驗次數 (R)。

參數輸入要求：

需要的成功次數必須是正整數，且 >0 ， <8000 。

成功的概率 >0 ， <1 （也就是， $0.0001 \leq P \leq 0.9999$ ），這點要注意

成功概率值為0和1的情況意義不大，不需要進行模擬，所以在軟體中不允許這兩種情況。

帕斯卡分布

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

帕斯卡分布的数学表达式如下：

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

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

$$\text{均值} = \frac{s}{p}$$

$$\text{标准差} = \sqrt{s(1-p)p^2}$$

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

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

输入要求

成功数 >0 必须是整数

$0 \leq \text{概率} \leq 1$

泊松分佈

泊松分佈描述的是在某個指定區間內事件發生的次數，例如每分鐘的電話個數或是文件中每頁的錯誤數量。

條件

泊松分佈的三個先決條件是：

- 在任何範圍內事件出現的次數可以是無限的。
- 事件之間是相互獨立的。某一區間的事件數量不會影響其他區間的事件數量。
- 區間之間的平均事件發生概率要保持一致。

泊松分佈的數學運算式如下：

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

$$\text{均值} = \lambda$$

$$\text{標準差} = \sqrt{\lambda}$$

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

$$\text{超峰度} = \frac{1}{\lambda}$$

唯一的分佈參數是比率或（ λ ）。

參數輸入要求：

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

2.5 連續分佈

反正弦分布

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

反正弦分布的数学表达式如下：

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

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

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

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

偏度=0

超峰度= 1.5

输入要求：

最小值<最大值

Beta 分佈

Beta 分佈非常的靈活，通常用來表現某個固定區域內的可變性。當描述的歷史資料和預測隨機行為為百分比形式時，也就是說數值的取值在 0 和 1 之間，您往往可以使用 Beta 分佈。改變 Beta 分佈的形狀和大小只要改變該分佈的兩個重要參數：alpha 和 beta（只能取正數）。並且如果這兩個參數相等，分佈就是對稱的；如果任意一個參數為 1，另外一個大於 1，分佈就是三角形的或者 J 形的；如果 alpha 小於 beta，分佈是正偏度的（大部分的值都靠近最小值）。如果 alpha 大於 beta，分佈是負偏度的（大部分值都靠近最大值）。

Beta 分佈的數學運算式如下：

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

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

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

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

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

分佈包含兩個參數 α 和 β ， Γ 是 γ 函數

條件

Beta 分佈的兩個先決條件是：

- 不定變數是 0 到 1 之間的隨機正值。
- 利用兩個正值可以確定分佈的形狀。

參數輸入要求：

α 和 $\beta > 0$ ，可以取任意正值

Beta 乘数转换分布

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

输入要求

Alpha > 0

Beta > 0

位置参数可以是任何正数或者负数包括 0

参数 > 0

柯西分佈、
Lorentzian 分佈或
Breit-Wigner
分佈

柯西分佈，也被稱為 **Lorentzian** 分佈或 **Breit-Wigner** 分佈，是用於描述共振行為的連續分佈。它還可以用來描述與 x 軸成任意角度相交的某條斜線的水平距離的分佈。

柯西分佈或洛倫茲分佈的數學運算式如下：

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

柯西分佈是一個特例，它不存在任何理論部分（均值，標準差，偏度和峰度），因為它們都是不明確的。

模的位置（ α ）和尺度（ β ）是本分佈的兩個參數。位置參數描述了分佈的高峰和眾數的位置，而尺度參數描述了分佈的寬帶。此外，柯西分佈或洛倫茲分佈的均值和方差都是不確定的。

另外，柯西分佈是自由度為 1 的學生分佈。

此分佈還可以由兩個標準常態分佈的比率構成（均值為 0，方差為 1 的常態分佈），且兩者相互獨立。

參數輸入要求：

位置參數 α 可以取任意值

尺度參數 β 可以是大大於 0 的任何正值

卡方分佈

卡方分佈是用來進行假設檢驗的概率分佈，和 Gamma 分佈和標準常態分佈類似。例如，獨立的標準常態分佈平方和是滿足自由度為 k 的卡方分佈：

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

卡方分佈的數學運算式如下：

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

均值 = k

標準差 = $\sqrt{2k}$

偏度 = $2\sqrt{\frac{2}{k}}$

超峰度 = $\frac{12}{k}$

Γ 代表 γ 函數。自由度 k 是唯一的分佈參數。

γ 分佈可以通過一定的設置變為卡方分佈：

形狀參數 = $\frac{k}{2}$ ，尺度 = $2S^2$ ，S 表示尺度。

參數輸入要求：

自由度為大於 1 小於 300 的整數

余弦分布

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

余弦分布的数学表达式如下：

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

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

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

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

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

偏度=0

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

输入要求：

最小值 < 最大值

双对数分布

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

双对数分布的数学表达式如下：

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

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

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

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

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

$$\text{偏度} = 0$$

输入要求：

最小值 < 最大值

Erlang 分布

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

Erlang 分布的数学表达式如下：

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

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

$$\text{均值} = \alpha\beta$$

$$\text{标准差} = \sqrt{\alpha\beta^2}$$

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

$$\text{超峰度} = \frac{6}{\alpha} - 3$$

输入要求：

Alpha（形状参数）>0 必须是整数

Beta（形状参数）>0

指数转换分布

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

输入变量的要求：

Lambda >0

位置参数可以是任何正数或者负数包括0

指數分佈

指數分佈被廣泛地用於描述獨立隨機事件發生的時間間隔的分佈，例如電子產品的壽命分佈或者到達服務攤點的時間間隔。它與泊松分佈相關（泊松分佈用來描述在某個時間間隔內事件發生的次數）。指數分佈的重要特性就是它的無記憶性，這意味著未來產品的壽命也滿足相同的概率分佈，不必考慮時間的存在。換句話說，時間對未來事件的發生沒有影響。

指數分佈的數學運算式如下：

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

$$\text{均值} = \frac{1}{\lambda}$$

$$\text{標準差} = \frac{1}{\lambda}$$

偏度=2（這個值適用於所有成功率 λ ）

超峰度=6（這個值適用於所有成功率 λ ）

成功率（ λ ）是唯一的分佈參數。成功的試驗次數用 x 表示。

指數分佈的先決條件是：

- 指數分佈描述的是事件之間的時間間隔。

參數輸入要求：

成功率 > 0

極值分佈或 Gumbel 分佈

極值分佈（類型 1）通常用於描述一段時間內，極大值回應的大小（例如，洪水，降雨，地震等等）。其他一些應用包括材料的斷裂強度，建築設計，飛機負荷量等。極值分佈也被稱為 **Gumbel** 分佈。

極值分佈的數學運算式如下：

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

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

$$\text{標準差} = \sqrt{\frac{1}{6} \pi^2 \beta^2}$$

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

超峰度=5.4（適用於所有的模和尺度值）

偏度=（適用於所有的模和尺度值）

超峰度=（適用於所有的模和尺度值）

模（ α ）和尺度（ β ）是分佈參數。

參數的計算：

極值分佈裏存在兩個標準參數：模和尺度。模參數是變數的極大值（概率分佈的最高點）。在您選擇模參數之後，您就可以對尺度參數進行估計。尺度參數的值大於 0。尺度參數越大，變數越大。

參數輸入要求：

模 α 可以取任意值

尺度 $\beta > 0$

F 分佈

F 分佈，也稱作 **Fisher-Snedecor** 分佈，也是一種連續的概率分佈，常常用來進行假設檢驗。特別是，在方差分析中用來檢驗兩個方差的統計性差異。分子自由度為 n ，分母自由度為 m 的 F 分佈與卡方分佈之間的關係如下：

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

$$\text{均值} = \frac{m}{m-2}$$

$$\text{標準差} = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)} \text{ for all } m > 4$$

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

超峰度 =

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

分子自由度 n 和分母自由度 m 是唯一的分佈參數。

參數輸入要求：

分子和分母自由度均為大於 0 的整數

Gamma 分佈 (Erlang 分佈)

可以應用 Gamma 分佈的物理量很多，並且其與其他一些分佈都有關聯：如對數常態分佈，指數分佈，Pascal 分佈，Erlang 分佈，泊松分佈和卡方分佈等。它可以被用於表示氣象過程中的污染濃度和降水量，或是在當事件過程並不是完全隨機情況下，度量事件發生的間隔，其他一些應用包括存貨控制，經濟理論和保險事故理論等。

條件

Gamma 分佈最常見的用途是用於泊松過程中某事件第 r 次發生時的時間分

佈。應用於此種情況必須的三個條件有：

- 任一計量單位裏事件可能發生的次數不限定於某個固定的數目。
- 每次發生的過程是相互獨立的。在某個計量單位發生的次數不影響另一計量單位的發生次數。
- 單位之間發生次數的平均值保持不變。

Gamma 分佈的數學運算式如下：

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

$$\text{均值} = \alpha\beta$$

$$\text{標準差} = \sqrt{\alpha\beta^2}$$

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

$$\text{超峰度} = \frac{6}{\alpha}$$

形狀參數 (α) 和尺度參數 (β) 是分佈參數， Γ 是 Gamma 函數。

參數 α 為正值的 Gamma 分佈被稱為 Erlang 分佈，被用於預測排隊系統中的等待時間，Erlang 分佈是一些獨立同一的且服從無記憶指數分佈的隨機分佈變數的總和。假定這些隨機變數的個數為 n ，那麼 Erlang 分佈的數學運算式如下：

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

參數輸入要求：

尺度參數 β 可以為大於 0 的任意數

形狀參數 α 可以為大於等於 0.5 任意數

位置數可以取任意值

拉普拉斯分布

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

拉普拉斯分布的数学表达式如下：

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

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

均值 = α

标准差 = 1.4142β

偏度 = 0

超峰度 = 3

输入要求：

Alpha（位置参数）可以取任何整数或者负数，包括0

Beta（尺度参数） > 0

Logistic 分佈

Logistic 分佈通常用來描述增長現象，例如，人口數量隨著時間變化的函數。也用來描述化學反應和群體或者個體的增長過程。

Logistic 分佈的數學運算式如下：

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

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

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

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

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

對所有均值和尺度均適用

均值（ α ）和尺度（ β ）是分佈參數。

參數的計算：

Logistic 分佈的兩個標準參數為：均值和尺度。均值參數就是平均值，在這個分佈裏也就是指眾數，因為它屬於對稱分佈。在選擇均值參數之後，您就可以估計出尺度參數。尺度參數為大於 0 的任意值。尺度參數越大，方差越大。

參數輸入要求：

尺度 β 為大於 0 的任意值

均值 α 可以取任意值

對數常態分佈

對數常態分佈被廣泛地用於數值偏度為正的情形。例如，在證券估價（如股價等）的財務分析中，標的數值大小都不可能小於 0，且通常都是正偏而非常態分佈的（對稱的）。同樣地，房地產的價格也表現出正偏特性。這些含有不確定性的變數不可能小於零，但大部分值都接近於下限。

對數常態分佈的三個先決條件是：

- 不確定變數的增長沒有上限，但是不能低於 0。
- 不確定變數的分佈是正偏的，也就是大部分的值都集中在低端。
- 不確定變數的自然對數服從常態分佈。

一般來說，如果變異係數大於 30%，就使用對數常態分佈。否則，就使用常態分佈。

對數常態分佈的數學運算式如下：

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

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

均值

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

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

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

均值（ μ ）和標準差（ σ ）是分佈參數。

參數輸入要求：

均值和標準差均可以為大於 0 的任意正值

設定對數常態分佈的參數

在通常情況下，對數常態分佈使用的是算術平均值和標準差。在可以得到歷史數

據的情況下，我們使用對數平均值和標準差，或是幾何平均數和標準差更合適。

常態分佈

常態分佈是概率裏最重要的分佈，因為它可以描述很多自然現象，如人們的IQ 值或高度。決策者們也可以利用常態分佈來描述諸如通貨膨脹率或未來汽油價格等不定變數。

條件

常態分佈的三個先決條件是：

- 不確定變數的某些值是最可能值（分佈的均值）。
- 不確定變數可能高於均值，也可能低於均值（對稱的分佈在均值周圍）。
- 不確定變數更可能分佈在均值的附近而不是較遠處。

常態分佈的數學運算式如下：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

x 和 σ 可取任何值; $\sigma > 0$

均值 = μ

標準差 = σ

偏度=0（對於所有的均值和標準差均適用）

超峰度=0（對於所有的均值和標準差均適用）

均值（ μ ）和標準差（ σ ）是分佈參數。

參數輸入要求：

標準差可以為大於 0 的任意正值

均值可以取任意值

拋物線的分配

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

拋物線的分配的数学表达式如下：

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

Alpha = Beta = 2

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

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

偏度=0

超峰度= -0.8571

输入要求：

最小值<最大值

Pareto 分佈

Pareto 分佈被廣泛地用於這樣的現象，例如城市的人口數量，自然資源的數量，公司大小，個人收入，股票價格的波動，通信電路的錯誤集中率。

Pareto 分佈的數學運算式如下：

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

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

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

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

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

形狀 (α) 和位置 (β) 是分佈參數。

參數計算：

Pareto 分佈存在兩個標準參數：位置和形狀。位置參數是變數的下界。在您選擇了位置參數之後，您可以估計出形狀參數。形狀參數的值為大於 0 的

某個數，通常是大於 1 的。形狀參數越大，方差越小，分佈的右尾就越粗。

參數輸入要求：

位置可以為大於 0 的任意正值

形狀參數 ≥ 0.05

佩尔森 V 分布

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

佩尔森 V 分布的数学表达式如下：

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

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

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

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

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

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

输入要求：

Alpha (形状参数) > 0

Beta(尺度参数) > 0

佩尔森 VI 分布

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

的数学表达式如下：

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

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

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

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

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

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

输入要求：

Alpha1(形状1)>0

Alpha2(形状2)>0

Beta(尺度参数)>0

PERT 分布

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

PERT 分布的数学表达式如下：

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

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

and B is the Beta function

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

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

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

输入要求：

最小值 ≤ 最有可能值 ≤ 最大值 可以是正数，负数，或者 0

Power 分布

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

的数学表达式如下：

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

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

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

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

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

输入要求：

Alpha（形状）> 0

Power 乘数转换分布

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

输入要求:

Alpha (形状) >0

位置参数可以是任何正数或者负数包括 0

参数>0

學生 t 分佈

t 分佈 (學生分佈) 是在假設檢驗中使用的最廣泛的一種分佈。它被用於在樣本空間很小的情況下, 估計某常態分佈總體的均值, 還被用於檢驗兩樣本均值的統計顯著性區別或是小樣本空間的置信區間。

t 分佈的數學運算式如下:

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

均值=0 (這適用於自由度 r 取所有值的情況, 除非分佈轉移到一個非 0 為中心的位置)

$$\text{標準差} = \sqrt{\frac{r}{r-2}}$$

偏度 = 0 (適用於自由度為 r 的所有值)

$$\text{超峰度} = \frac{6}{r-4}, r > 4$$

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

自由度 r 是唯一的分佈參數。

t 分佈與 F 分佈的關係如下: 自由度為 r 的 t 分佈的值的平方, 服從自由度為 1 和 r 的 F 分佈。除了它更扁幅度更廣或是它的尖頂峰度 (更肥胖的尾部和更尖的中部) 以外, t 分佈的概率密度函數的形狀與均值為 0, 方差為 1 的常態分佈變數的形狀類似。當自由度增加時 (例如超過 30), t 分佈就趨近於均值為 0, 方差為 1 的常態分佈。

參數輸入要求:

自由度必須為大於 1 的整數

三角分佈

三角分佈描述的是已知最小值, 最大值以及最大似然值時的一種情形。例如, 您可以通過過去汽車銷售數量的最大值、最小值、以及每周的一般銷售數量來描述現在的汽車銷售數量。

三角分佈的三個先決條件是：

- 專案的最小值是固定的。
- 專案的最大值是固定的。
- 專案的最大似然值落在最小值和最大值之間，形成一個三角形的分佈，顯示在最大值和最小值附近發生的概率小於在最大似然值附近發生的概率。

三角分佈的數學運算式如下：

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

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

標準差

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

偏度=

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

超峰度= -0.6（適用於所有的最大，最小，最有可能值）

最小值（Min），最有可能值（Likely）和最大值（Max）是分佈的三個參數。

參數輸入要求：

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

均勻分佈

在均勻分佈情況下，所有值都落在最大值和最小值之間的範圍內且發生的概率相同。

條件：

- 最小值是固定的。
- 最大值是固定的。
- 在最大值和最小值範圍內的值的發生概率是相同的。

均勻分佈的數學運算式如下：

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

$$\text{均值} = \frac{Min + Max}{2}$$

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

偏度 = 0 (此值對於任意最大值和最小值均成立)

超峰度 = - 1.2 (此值對於任意最大值和最小值均成立)

此值對於任意最大值和最小值均成立

最小值 (Min) 和最大值 (Max) 是分佈的兩個參數。

參數輸入要求：

最小值 < 最大值，且可以取任意值

韋伯分佈 (Rayleigh 分佈)

Weibull 分佈用於描述來自壽命和衰變測試的資料滿足的概率分佈。通常被用於描述可靠性分析中的出錯次數，包括材料和質量控制中的強度測試。

Weibull 分佈也用於表現不同的物理量，比如音速。**Weibull** 分佈是一族分佈的統稱，可以用於表現很多事物的特性。例如，通過選取形狀參數，**Weibull** 分佈可以用於類比指數分佈 (形狀參數=1.0) 和 **Rayleigh** 分佈 (形狀參數=2.0) 等等。當形狀參數=1.0 時，通過設置 **Weibull** 分佈中心位置尺度參數 (beta)，可以建立起始點不在原點 (0,0) 的指數分佈。當形狀參數小於 1.0 時，**Weibull** 分佈成削尖的曲線。製造業可以使用這個分佈描述零件在燒制過程中的出錯率。

韋伯分佈的數學運算式如下：

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

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

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

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

超峰度

=

$$\frac{-6\Gamma^4(1+\beta^{-1})+12\Gamma^2(1+\beta^{-1})\Gamma(1+2\beta^{-1})-3\Gamma^2(1+2\beta^{-1})-4\Gamma(1+\beta^{-1})\Gamma(1+3\beta^{-1})+\Gamma(1+4\beta^{-1})}{[\Gamma(1+2\beta^{-1})-\Gamma^2(1+\beta^{-1})]^2}$$

形狀 (α) 和中心位置尺度 (β) 是分佈的兩個參數， Γ 是 γ 函數。

參數輸入要求：

形狀參數 $\alpha \geq 0.05$

尺度參數 β 可以為大於 0 的任意正值

Weibull 乘數 轉換分布

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），可以建立起起始点不在原点（0, 0）的指数分布。当形状参数小于1.0时，Weibull分布成削尖的曲线。制造业可以使用这个分布描述零件在烧制过程中的出错率。形状参数（ α ）和尺度参数（ β ）是分布的重要参数。

输入变量要求：

形状参数 ≥ 0.05

中心位置控制尺度 $\text{Beta} > 0$ 可以是任何正值

位置参数可以是任何正数或者负数包括0

参数 > 0

3

3 · 預測

預測是基於歷史資料對未來做出的合理估計，或是在不存在歷史資料的情況下對未來進行的揣摩。當存在歷史資料時，最好使用定量方法（或統計學方法），但是當不存在歷史資料時，定性法（或判斷法）就是唯一的選擇了。下圖 3.1 列出了最常見的預測方法。

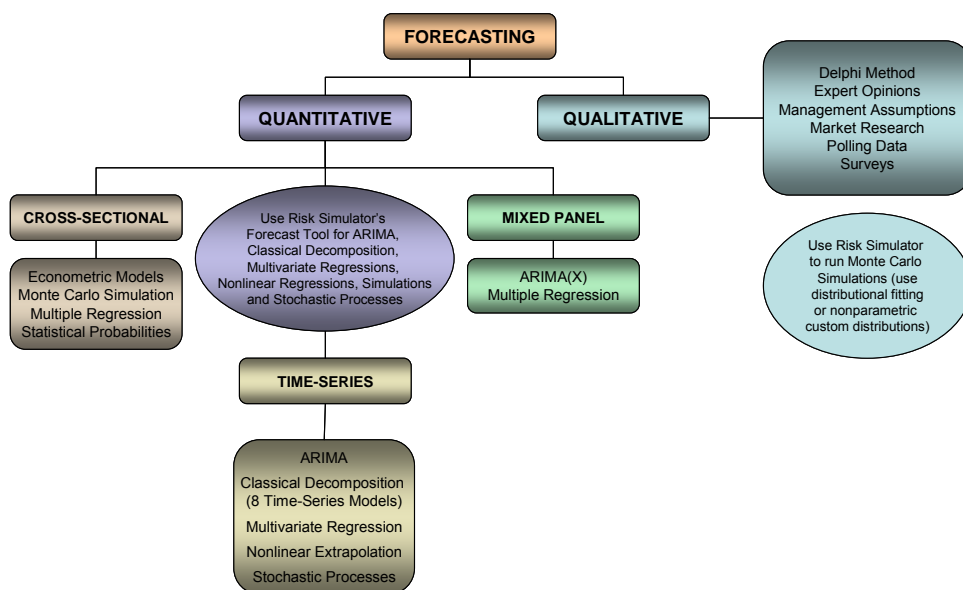


圖 3.1 預測方法

3.1 不同類型的預測方法

一般來說，預測的方法可以分為定性和定量兩種。當幾乎不存在可靠的歷史資料，同期資料或可比資料時，就採用定性法。定性法包含有德爾菲法或專家意見法（行業專家、市場專家，或內部成員達成共識的預測）、管理假設法（由高級管理層設置的目標增長率），還有市場研究、外部資料、投票和調查（從第三方如行業或地區指標獲得的資料，或是積極的市場調研）等。這些估計可以是單點估計值（一般意見）或是一系列預測值（預測分佈）。後者可以在 Risk Simulator 中輸入一個自定義分佈，並對預測結果進行模擬。意思就是可以通過預測的數值本身得到一個分佈來進行一次非參數模擬。

定量法中的可用資料或是需要預測的資料可以分為時間序列資料（涉及

時間因素的變數，如不同年份的收益、通脹率、利率、市場份額、失敗率等），截面資料（與時間無關的變數，如某一年份不同地區大二學生的平均績點，每個學生的 SAT 分數水平，每周酒精飲料的消費量等），和混合面板資料（時間序列資料和面板資料的混合體，例如，在給定市場成本預算及市場份額的前提下，預測未來十年的銷售量，這意味著銷售資料是一個時間序列的外生變數，將銷售成本及市場份額作為模型參數有助於模擬預測）。

風險類比軟體為用戶提供以下的各個預測方法：

1. ARIMA（自回歸求和移動平均）
2. 自動 ARIMA
3. 計量經濟學自動分析功能
4. 計量經濟學基本功能
5. 神經網路
6. 三次樣條插值法
7. GARCH（廣義自回歸條件異方差）
8. J-曲線
9. S-曲線
10. 組合模糊邏輯
11. 馬爾可夫鏈
12. 最大似然法
13. 非線形外推
14. 回歸分析
15. 隨機過程
16. 時間序列分析
17. 趨勢線

對於每個不包含在用戶手冊中的預測方法的分析細節，讀者如果想瞭解更多細節，請參看由 Johnathan Mun 博士撰寫的《*風險建模：應用蒙地卡羅類比，實物期權分析，預測與最佳化技術*》(Wiley Finance 2006)。他還是**風險類比**軟體的開發者。然而，以下還闡述一些更加常用的方法。所有其他的預測方法可以在**風險類比**軟體中相當容易地進行應用。

3.2 運行 Risk Simulator 中的預測工具

一般說來，在開始預測之前先進行以下幾個步驟：

- 打開 Excel 表格輸入歷史資料或是打開已有的歷史資料表格
- 選擇資料，點擊**模擬**選擇**預測**
- 選擇相關項（ARIMA、多元回歸、非線性外推法、隨機預測、時間序列分析），然後輸入相關的參數

圖 3.2 顯示了預測工具及多種預測方法



圖 3.2 Risk Simulator 的預測方法

接下來的部分將對每種方法作一個簡短的介紹，並提供了一些軟體使用的例子。例子使用的資料檔案可以通過以下路徑獲得：**開始|程式|Real Option Valuation|Risk Simulator|示例**。

3.3 時間序列分析

理論：

圖 3.3 列出了八種最常見的時間序列模型，通過季節性和趨勢性來分類。例如，如果資料沒有季節性和趨勢性，那麼用單滑動平均模型或單光滑指數模型就足夠了。但是如果存在季節性但是沒有表現明顯的趨勢性，那麼最好使用季節附加模型或是季節乘積模型。

	無季節性	有季節性
無趨勢性	單滑動平均模型	季節附加模型
	單光滑指數模型	季節乘積模型
有趨勢性	雙滑動平均模型	Holt-Winter 附加模型
	雙光滑指數模型	Holt-Winter 乘積模型

表 3.3 八種最常見的時間序列方法

步驟:

- 打開 Excel，如果需要的話，打開您的歷史資料（下面的例子中使用的是示例文件夾中的時間序列預測文件）
- 選擇歷史資料（資料必須列在同一列中）
- 選擇模擬預測時間序列分析
- 選擇要應用的模型，輸入相關的參數，點擊**確定**

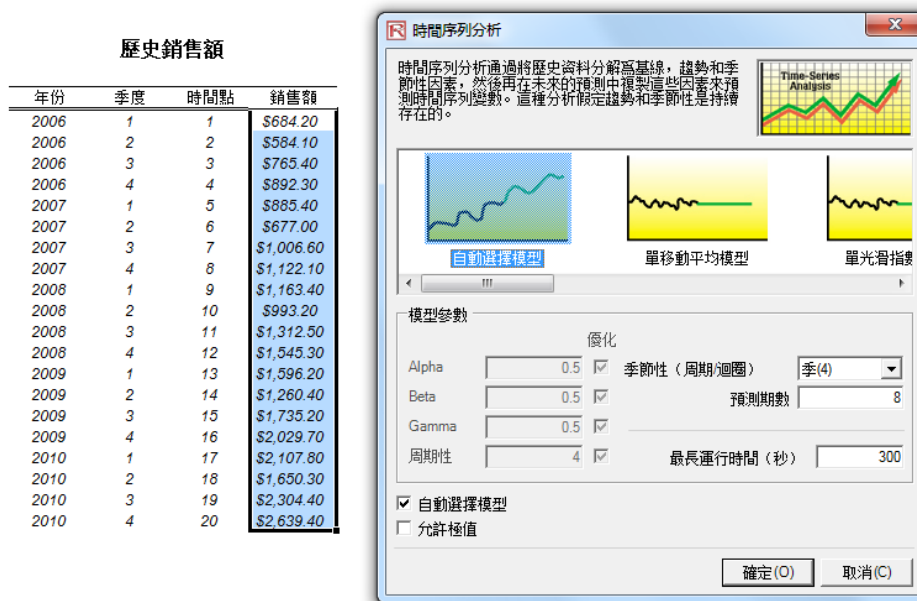


圖 3.4 時間序列分析

結果解析:

圖 3.5 中所示是使用預測工具生成的結果樣本。使用的是 Holt-Winter 乘積模型。注意到圖 3.5 中的模型擬合和預測圖顯示出 Holt-Winter 乘積模型很好地表現了趨勢性和季節性。時間序列分析報告會提供相關的參數 α , β , γ 的

最優值，誤差度量，擬合數據，預測值，以及擬合-預測圖等。參數僅供參考。 α 針對的是基準水平隨時間變化的記憶影響， β 是趨勢參數，衡量了趨勢的強度， γ 衡量了歷史資料的季節性強度。分析首先將歷史資料分解為三個因素，然後又將這三個因素重組起來預測未來的資料。擬合數據利用重組模型分析了歷史資料和擬合數據，表現了預測和過去的接近程度（這種技術叫倒推）。預測值可能是單點估計或假設（如果選擇了自動生成假設選項，並存在一個模擬文檔）。圖中顯示了歷史值，擬合值以及預測值。圖表是一種有效的交流途徑，並在視覺上為我們顯示了預測模型的效果。

注意：

如圖 3.3 所示時間序列分析模組包含八個時間序列模型。您可以基於趨勢和季節性來選擇特定的模型運行，或是選擇自動選擇模型選項，它會自動在八種方法中重復匹配，最最佳化參數，最終找出這些資料的最佳擬合模型。如果您選擇了八個模型的其中之一，您也可以不選擇最最佳化這一檢驗欄，而是輸入您自己的 α ， β ， γ 參數。想要瞭解更多關於這些參數的技術細節，請參考 Dr. Johnathan Mun 的 *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《風險建模：應用蒙特卡洛類比，實物期權分析，預測及最最佳化》，第二版，Wiley 2010。如果您選擇自動選擇模型或是其他任何季節性模型，您需要輸入相關的季節性周期。輸入的必須是一個正整數（例如，如果資料是季節性的，就輸入 4 作為一年的季節周期，如果是月度資料就輸入 12）。接下來輸入預測的期數。同樣這個數也必須是正整數。設置的最長運行時間是 300 秒。一般來說，這個不需要修改。但是如果運行大批量歷史資料時，分析需要的時間可能會比較長，如果運行時間超過了這個最長時間，過程就會中止。您也可以選擇預測自動生成假設。也就是說，預測可以通過假設的分佈進行而不是單點估計。最後，選擇極值參數令您可以在最佳化 α ， β ， γ 參數時，包含 0 值和 1 值。某些預測軟體允許這些極值參數，但是有些不允許。Risk Simulator 允許您在這兩者之間進行選擇。一般情況下不必選擇允許極值參數。

Holt-Winter乘積模型

統計匯總

Alpha, Beta, Gamma	均方根誤差	Alpha, Beta, Gamma	均方根誤差
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		

模型參數 alpha = 0.2429, beta = 1.0000, gamma = 0.7797, and seasonality = 4

時間序列分析匯總

當原始資料中體現出季節性和趨勢性時，更高級的模型將資料分解成三部分：資料的基本水平其權重由參數alpha表示，趨勢性部分其權重由參數beta表示，季節性部分其權重由參數gamma表示。誠然，在這方面有很多方法，但普遍使用的是Holt-Winters季節附加模型和Holt-Winters乘積模型。對Holt-Winter附加模型而言，資料基本水平，季節性部分和趨勢性部分是加在一塊去擬合進行預測的。

滑動平均預測的最佳擬合測試利用均方根誤差(RMSE)。RMSE計算的是擬合值與真實資料之間的均方差的平方根。

均方誤差 (MSE統計量) 是使用誤差 (真實的歷史數值和模型擬合的數值之差) 平方的方法來剔除誤差符號的影響，得到的一種絕對誤差測量。MSE統計量使用平方，傾向於使誤差大的點在整個統計量中所占權重更大而誤差小的點所占權重更小，這可用來比較不同的時間序列模型的優劣。均方根誤差 (RMSE) 也叫做二次損失函數是由MSE統計量開根號得到，是最常用的一種誤差測量。RMSE統計量可以被看成預測誤差的均值的絕對值。當預測的誤差所帶來的損失和預測的誤差絕對值大小成比例時，這是非常有用的。RMSE統計量被看成判斷最優時間序列擬合的一種標準。

平均絕對百分比誤差 (MAPE統計量) 計算相對歷史資料的平均絕對百分比，是一種相對誤差統計學測量。當預測誤差所造成的損失和誤差百分比更有關而不是太小時，該統計量是很有用的。最後我們提到的一種有用的誤差度量是Theil's U統計量，該統計量是對模型直觀的檢測。當Theil's U統計量小於1.0時，表示該模型預測提供的估計從統計上來講比主觀預測要好。

時間點	真實值	擬合值	誤差度量
1	684.20		均方根誤差 71.8132
2	584.10		均方誤差 5157.1348
3	765.40		平均絕對離差 53.4071
4	892.30		平均絕對誤差率 4.50%
5	885.40	684.20	Theil's U 0.3054
6	677.00	667.55	
7	1006.60	935.45	
8	1122.10	1198.09	
9	1163.40	1112.48	
10	993.20	857.95	
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	
預測變數21		2713.69	
預測變數22		2114.79	
預測變數23		2900.42	
預測變數24		3293.81	
預測變數25		3346.55	
預測變數26		2580.81	
預測變數27		3506.19	
預測變數28		3947.61	

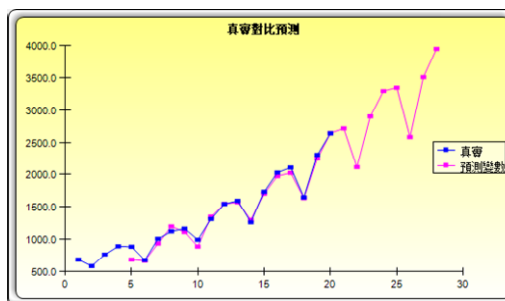


圖 3.5 Holt-Winter 預測報告示例

3.4 多元回歸

理論：

假設使用者對回歸分析的基礎知識已有充分的瞭解。一般的雙變數線性回歸的形式是： $Y = \beta_0 + \beta_1 X + \epsilon$ ，其中 β_0 代表截距， β_1 代表斜率， ϵ 是誤差項。之所以稱其為雙變數是因為它只存在兩個變數，因變數Y，引數X，同時X也被稱為回歸量（有時候雙變數回歸也被稱為單變數回歸，因為只存在一個引數X）。之所以取名為因變數是因為它受引數影響，例如，銷售收益與產品廣告和促銷帶來的市場成本量有關，那麼銷售收益就是因變數，市場成本就是引數。雙變數回歸的一個例子可以看二維空間中如圖 3.6 中的左圖，簡單地在一系列資料點中插入一條最佳擬合直線。還有些例子是多元回歸，就是存在多個或是n個引數，此時的回歸運算式為 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \epsilon$ 。在這個例子中，最佳擬合線就位於一個n+1維的空間中了。

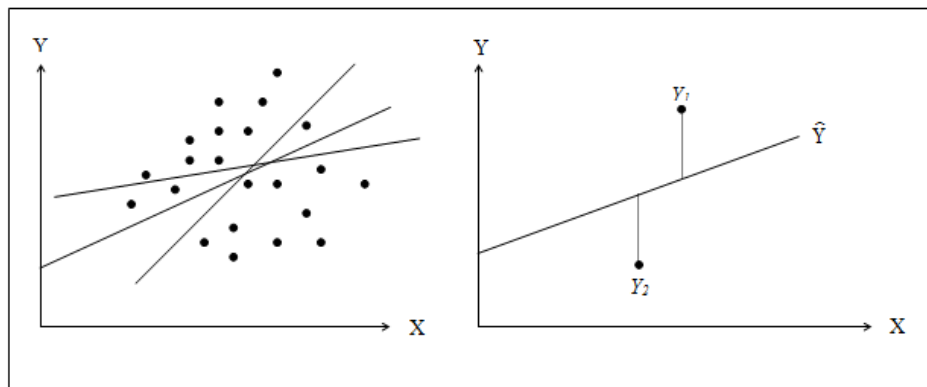


圖 3.6 雙變數回歸

但是類似於圖 3.6，對點狀圖中的一系列資料用直線進行擬合，可能會有許多結果。總體垂直誤差（也就是圖 3.6 右圖中顯示的真實資料點（ Y_i ）和估計直線（ \hat{Y} ）之間的距離絕對值之和）最小的那條線就被稱為最佳擬合直線。為了找出讓誤差最小的最佳擬合直線，我們需要一種更加嫺熟的方法，那就是回歸分析。回歸分析通過最小化總體誤差來找出唯一的最佳擬合直線。這是通過計算如下方程式

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

來實現的。

只有唯一一條直線可以使誤差平方和最小。我們將誤差（真實資料點和預測線之間的垂直距離）平方以避免正負誤差之間相互抵消。為了解決這個涉及截距和斜率的最小化問題，我們需要通過它的一階導數值為 0 的性質，來得到方程：

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

這樣就可以得到雙變數回歸的最小二乘等式：

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

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

對於多元回歸，可以依此類推計算多個因變數，例如，

$Y_i = \beta_1 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i$ ，斜率的估計可以通過如下式子來計算：

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

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

在運行多元回歸時，對求解過程和解釋結果需要更加注意。例如，在建立一個合適的模型之前需要對計量建模的知識有很好的掌握（識別回歸的缺陷如結構中斷點，多元共線性、異方差性、自相關、規格測試和非線性等）。關於多元回歸以及如何識別回歸缺陷的詳情分析及討論請參考 *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《風險建模：應用蒙特卡洛類比、實物期權分析、預測及最最佳化》，第二版（Wiley 2010）。

步驟：

- 打開 Excel，如果需要的話打開您的歷史資料（下面的圖示使用的是示例文件夾中的多元回歸文件）
- 檢查以確定資料都排在同一列，選擇包括變數名稱的整個資料區域，然後選擇**模擬預測多元回歸**
- 選擇引數並檢查相關選項（滯後，逐步回歸，非線性回歸等），然後點擊**確定**

結果解析：

圖 3.8 是一份多元回歸的結果樣本報告。這份報告很完整，包括所有的回歸結果，變數結果的分析，擬合圖以及假設檢驗結果。對於這些結果技術細節的解釋不包括在本手冊範圍內。更多關於多元回歸及回歸報告的解釋分析及討論請參考 Dr. Johnathan Mun 的 *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《風險建模：應用蒙特卡洛類比，實物期權分析，預測及最最佳化》，第二版（Wiley 2010）。

多元回歸

Y	X1	X2	X3	X4	X5
521	18308	185	4.041	79.6	7.2
367	1148	600	0.55	1	8.5
443	18068	372	3.665	32.3	5.7
365	7729	142	2.351	45.1	7.3
614	100484	432	29.76	190.8	7.5
385	16728	290	3.294	31.8	5
286	14630	346	3.287	678.4	6.7
397	4008	328	0.666	340.8	6.2
764	38927	354	12.938	239.6	7.3
427	22322	266	6.478	111.9	5
153	3711	320	1.108	172.5	2.8
231	3136	197	1.007	12.2	6.1
524	50508	266	11.431	205.6	7.1
328	28886	173	5.544	154.6	5.9
240	16996	190	2.777	49.7	4.6
286	13035	239	2.478	30.3	4.4
285	12973	190	3.685	92.8	7.4
569	16309	241	4.22	96.9	7.1
96	5227	189	1.228	39.8	7.5
498	19235	358	4.781	489.2	5.9
481	44487	315	6.016	767.6	9
468	44213	303	9.295	163.6	9.2
177	23619	228	4.375	55	5.1
198	9106	134	2.573	54.9	8.6
458	24917	189	5.117	74.3	6.6
108	3872	196	0.799	5.5	6.9
246	8945	183	1.578	20.5	2.7
291	2373	417	1.202	10.9	5.5
68	7128	233	1.109	123.7	7.2
311	23624	349	7.73	104.2	6.6
606	5242	284	1.515	12.5	6.9
512	92629	499	17.99	381	7.2
426	28795	231	6.629	136.1	5.8
47	4487	143	0.639	9.3	4.1
265	48799	249	10.847	264.9	6.4
370	14067	195	3.146	45.8	6.7
312	12693	288	2.842	29.6	6



1. 選擇資料範圍包括資料的標題(B5:G65)
2. 點擊Risk Simulator|預測多元回歸
3. 選擇因變數(在本例中就是變數Y)然後進行一系列對模型的優化(帶後回歸, 非線性回歸, 逐步回歸)然後點擊確定。查看分析報告作為分析的結果。



圖 3.7 運行一個多元回歸

回歸分析報告

回歸統計量	
R-方 (決定係數)	0.3272
調整後的R平方	0.2508
多元R (多元相關係數)	0.5720
估計的標準差 (SEy)	149.6720
樣本數	50

R方或可決係數表明在此次回歸分析中引數對因變數的解釋程度。但是，在多元回歸中，調整的R方還考慮到了某些額外的引數或回歸量的存在，將R方調整到一個更加精確的水平，以增加回歸的解釋說明能力。因此，只有因變數才能被回歸量解釋。

多元相關係數 (多元R) 度量的是真實因變數 (已知的歷史結果) 與基於回歸方程式的估計或擬合變數 (擬合的結果) 的相關性。它是可決係數 (R方) 的平方根。

估計的標準差描述的是資料點向上和向下偏離回歸線或平面的程度。這個值稍後可用於計算和估計置信區間。

回歸結果						
	截距	X1	X2	X3	X4	X5
係數	57.9555	-0.0035	0.4644	25.2377	-0.0086	16.5579
標準差	108.7901	0.0035	0.2535	14.1172	0.1016	14.7996
t統計量	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	假設檢驗	
回歸自由度	5	臨界t值 (99% 的致信度, 自由度為 44)
殘差自由度	44	臨界t值 (95% 的致信度, 自由度為 44)
總自由度	49	臨界t值 (90% 的致信度, 自由度為 44)
		2.6923
		2.0154
		1.6802

係數提供了回歸模型的截距和斜率。舉例來說，這些係數提供了對真實值的一種估計：因變數 (人口數b) 的值可通過回歸方程 $Y=b_0+b_1X_1+b_2X_2+...+b_nX_n$ 來估計。標準差反映了預測的係數的精確程度，t統計量是預測的係數和它的標準差的比率。

t統計量被用在假設檢驗中。這裏我們設置零假設：係數的均值為0。備擇假設：係數的均值不為0。運用t檢驗並將計算的t統計量和殘差自由度的臨界值進行比較。t檢驗是非常重要的，它的計數值反映了當存在額外的回歸量時，每個係數是否在統計上是有效的。也就是說，t檢驗在統計上說明了一個回歸量或獨立變數是否應該保留在回歸模型中還是從模型中剔除。

如果係數的t統計量超過了相關自由度t統計量的估計臨界值，這些係數在統計上就是有效的。這裏，三種常用的置信水平是90%，95%和99%。如果係數的t統計量超過了臨界值，就認為該係數是統計上有效的。作為選擇，P值計算了取到t統計量的概率，也就是說P值越小，所對應的係數就越有效。P值對應90%，95%，99%三個置信水平最常用的有效臨界值為0.01，0.05和0.10。

那些P值用藍色標出的係數表明它們在95%的置信水平或0.05的α水平上是統計顯著的，那些用紅色標出的表明其在任何α水平上都是統計不顯著的。

變數分析

	平方和	算術平方	F統計量	P值	假設檢驗	
回歸	479388.49	95877.70	4.28	0.0029	F統計量臨界值 (99% 的致信度, 自由度為 5 and 44)	3.4651
殘差	985675.19	22401.71			F統計量臨界值 (95% 的致信度, 自由度為 5 and 44)	2.4270
總和	1465063.68				F統計量臨界值 (90% 的致信度, 自由度為 5 and 44)	1.9828

方差分析表 (ANOVA) 提供了對回歸模型總體統計有效性的F檢驗。與t檢驗關注單個回歸量不同的是，F檢驗關注所有估計係數的統計屬性。F檢驗計算的是回歸的均值的平方和殘差的均值的平方的比例。分子表明回歸模型對預測值的解釋程度，分母表明有多少程度模型不能解釋的。因此，F-Statistic越大表明我們的模型就越有效。相應的P值也被計算用來假設檢驗，判斷回歸模型的整體有效性。這裏零假設為：所有的係數都為0。備擇假設：所有係數不同時為0。如果P值在0.05或0.10的α水平有效水平下，比0.01小，則回歸模型是有效的。同樣的方法在對比計算的F-Statistic和在各種有效水平下的臨界的F值時也能應用。

預測

時間點	真實值(Y)	預測值(F)	誤差 (E)
1	521.0000	299.5124	221.4876
2	367.0000	487.1243	(120.1243)
3	443.0000	353.2789	89.7211
4	365.0000	276.3296	88.6704
5	614.0000	776.1336	(162.1336)
6	385.0000	298.9993	86.0007
7	286.0000	354.8718	(68.8718)
8	397.0000	312.6155	84.3845
9	764.0000	529.7550	234.2450
10	427.0000	347.7034	79.2966
11	153.0000	266.2526	(113.2526)
12	231.0000	264.6375	(33.6375)
13	524.0000	406.8009	117.1991
14	328.0000	272.2226	55.7774
15	240.0000	231.7882	8.2118
16	286.0000	257.8862	28.1138
17	285.0000	314.9521	(29.9521)
18	569.0000	335.3140	233.6860
19	96.0000	282.0356	(186.0356)
20	498.0000	370.2062	127.7938
21	481.0000	340.8742	140.1258
22	468.0000	427.5118	40.4882
23	177.0000	274.5298	(97.5298)

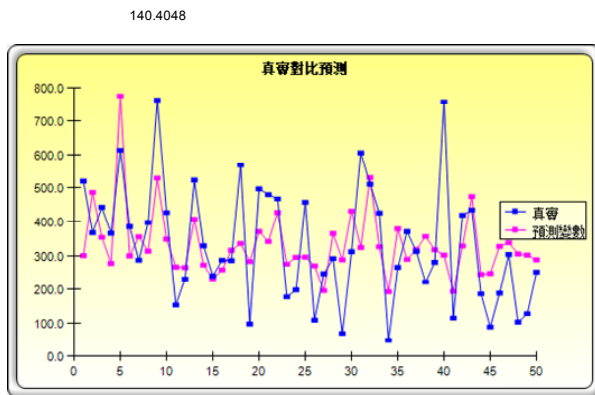


圖 3.8 多元回歸的結果

3.5 隨機預測

理論：

隨機過程其實是一個隨時間的變化而產生一系列結果的數學方程式，從本質上說，它的結果是不確定的。這個過程或方程式不遵循任何簡單的可辨規則，例如價格每年上升 X 個百分點，或由於這個因素收益每年上升 X 乘以 Y 個百分點，這裏的 X, Y 都可以看成隨機過程。隨機過程在定義上是不確定的，您可以在隨機過程方程式中隨意輸入參數並且每次得到不同的結果。例如，股票價格的路徑（股票的走勢）在本質上是隨機的，沒有人可以確定預測股票價格的路徑。但是隨時間變化的價格卻是來源於某個過程。這個過程是事先確定的，但是產生的結果不確定。因此通過隨機模擬，我們可以創造出多條價格路徑，得到這些模擬的一個統計取樣，然後在給定生成時間序列的隨機過程的參數和特性的情況下，推斷出真實價格可能遵循的潛在路徑。Risk Simulator 的預測工具包含三個最基本的隨機過程：布朗運動或隨機遊走，均值回復過程，跳躍擴散過程。布朗運動由於其簡單性和廣泛應用性成爲最流行的隨機過程。

在隨機過程模擬中比較有趣的是歷史資料並不是必需的。也就是使用這個模型並不要求對歷史資料進行擬合。只需要計算出期望值和歷史資料的波動率，或通過可比較的外部資料來得到估計值，以及對變數進行假定後，就能使用該模型。參見 Dr. Johnathan Mun 的 *Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* 《風險建模：應用蒙特卡洛類比，實物期權分析，預測及最最佳化》，第二版，Wiley 2010 來查看關於如何計算每個輸入量的詳細說明（例如均值回復率，跳躍概率，波動率等等）。

步驟：

- 通過**模擬預測隨機過程**按鈕開啓模型
- 選擇想要的過程，輸入參數，點擊更新圖表來保證過程與您的預期一致，然後點擊**確定**（見圖 3.9）

結果解析：

圖 3.10 中顯示的是一個隨機過程示例結果。圖表顯示了疊代結果，報告中闡述了隨機過程的基本知識。此外，還提供了每階段的預測值（均值和標準差）。通過這些值您可以決定哪些時間段與您的分析是相關的，並用這些常態分佈的均值和標準差設定輸入假設。然後可以使用這個輸入假設在您自定義的模型中進行模擬。



圖 3.9 隨機過程預測

隨機過程預測

統計匯總

所謂隨機過程是指服從某一概率分佈的一系列事件和軌迹的集合。就是說，隨機事件雖然隨著時間發生但這些事件都符合某一特殊的統計和概率法則。我們主要碰到的隨機過程包括隨機遊走（布朗運動），均值回復，跳躍—擴散這些。這些過程能用來預測大多數體現隨機傾向但又服從概率分佈的變數的變化。

隨機遊走（布朗運動）能用來預測股票的價格，商品的價格和任何沿著漂移路徑有著漂移或增長率和波動率的隨機時間序列資料。均值回復可以通過逾期目標水平來減小隨機遊走的波動。該過程可以用來預測如利率，通脹這些有長期目標水平（這些長期目標水平由權威機構和市場提供）的時間序列變數。跳躍—擴散過程可以用來預測如石油價格，電力價格（個別外部事件的發生能使價格往上跳躍或下降）這些偶爾伴有隨機跳躍的時間序列資料。最後，這三類隨機過程可根據需要相互配合使用。

右邊的結果展示的是在每次時間步長疊代後的均值和標準差。如果您選中了顯示所有疊代按鈕，那麼每一個疊代產生的路徑將會被顯示在一個單獨的工作簿中。在下面的圖表裏給出了幾個生成的疊代路徑。

隨機過程：帶漂移的布朗運動（隨機行走）

初始值	100	步長	100.00	跳躍率	N/A
漂移率	5.00%	反覆運算次數	10.00	跳躍大小	N/A
波動率	25.00%	回復率	N/A	亂數種子	326768989
水準	10	逾期值	N/A		

時間	均值	標準差
0.0000	100.00	0.00
0.1000	101.62	6.72
0.2000	107.95	10.71
0.3000	111.49	13.48
0.4000	109.27	15.79
0.5000	111.96	19.02
0.6000	112.74	23.79
0.7000	112.77	26.40
0.8000	115.23	27.47
0.9000	121.34	29.12
1.0000	123.65	31.14
1.1000	123.25	34.81
1.2000	122.90	38.86
1.3000	121.58	37.17
1.4000	121.43	36.97
1.5000	124.81	41.49
1.6000	125.68	39.85
1.7000	122.98	44.10
1.8000	129.48	47.98
1.9000	124.57	44.08
2.0000	125.79	45.68
2.1000	127.18	45.62
2.2000	129.11	48.57
2.3000	137.87	58.77
2.4000	145.91	65.66
2.5000	145.77	65.23
2.6000	137.92	61.54
2.7000	140.18	59.79
2.8000	142.31	66.07
2.9000	150.38	76.78
3.0000	151.01	90.66
3.1000	153.98	86.32
3.2000	153.18	86.69
3.3000	152.89	86.12
3.4000	150.29	82.58
3.5000	151.66	88.77
3.6000	155.18	93.29
3.7000	154.23	96.01
3.8000	152.75	93.10
3.9000	159.90	99.41
4.0000	159.62	102.75
4.1000	165.41	110.32
4.2000	165.54	102.83
4.3000	165.10	98.32
4.4000	167.07	96.27
4.5000	170.19	98.38
4.6000	171.51	95.56
4.7000	166.90	90.79
4.8000	166.19	94.17
4.9000	168.13	93.17
5.0000	175.48	97.80
5.1000	175.22	94.75
5.2000	171.10	94.26
5.3000	163.79	82.22

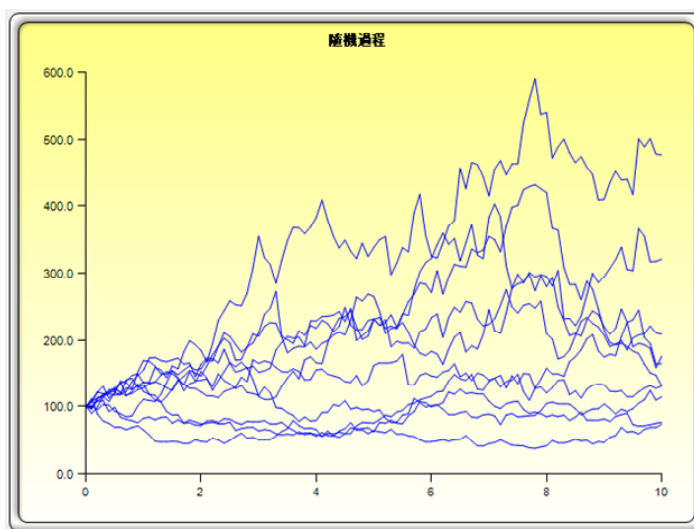


圖 3.10 隨機預測結果

3.6 非線性外推法

理論：

外推法是通過對歷史資料的趨勢用統計學投影的方法來預測未來。它只能被用於時間序列預測中。對於截面資料或混合平板資料（時間序列和截面資料）來說，多元回歸更加適用。這種方法在不期望發生較大的變化，即期望因果關係保持不變時，或某一情景下的因果關係沒有明確被確定時比較有用。它還可以防止將個人偏見引入過程之中。外推法相對比較可靠，簡單，所付的代價也相對較小。但是由於外推法是基於近期的發展趨勢將和歷史趨勢保持一致的假設之上的，所以如果在投影期間內發生了中斷，那麼將會產生大量的預測誤差。也就是說，時間序列的純粹外推假設我們所需要知道的所有資訊都包含在被預測序列的歷史資料之中。如果我們假設過去的行為是未來行為的一個好的預報器，那麼外推法就是適用的。故當所需要知道的所有資訊是一些短期預測時，外推法不失為一種有效的方法。

外推法估計了任意 x 值的 $f(x)$ 函數，通過在所有 x 值間插入一條光滑的非線性曲線，利用這條曲線，外推出一個基於此歷史資料集之外的未來的 x 處 $f(x)$ 的值。外推法可以採用多項式函數的形式或是有理數函數的形式（兩多項式之比）。一般情況下，對於性質好的資料多項式函數形式就足夠了，但是有時候使用有理數函數形式會更加精確（尤其是對於極值函數來說，例如，分母趨近於 0 的函數）。

步驟：

- 打開 Excel 表格，如果需要的話，打開您的歷史資料（下面的例子使用的是示例文件夾
- 中的**非線性外推**的例子）
- 選擇時間序列資料，並選擇**模擬|預測|非線性外推**
- 選擇外推類型（自動選擇，多項式函數或有理數函數），並輸入所需要預測時間段（圖 3.11），點擊**確定**

結果解析：

圖 3.12 中的結果報告顯示了外推預測值，測量誤差和外推結果的圖示。誤差度量法應當被用於檢測預測的有效性，在比較外推法和時間序列分析的預測質量和精度時尤為重要。

注意：

當歷史資料很光滑，並遵循某種非線性模式和曲線時，外推法比時間序列分析法更有效。但是，當資料模式顯示出季節性周期和趨勢的時候，時間序列分析則可以提供更好的結果。

非線性外推法

外推法是通過對歷史資料的趨勢用統計學投影的方法來預測未來。它只能被用於時間序列預測中。這種方法在不期望發生較大的變化，即期望因果關係保持不變時，或某一情景下的因果關係沒有明確被確定時比較有用。它還可以防止將個人偏見引入過程之中。外推法相對比較可靠，簡單，所付的代價也相對較小。但是由於外推法是基於近期的發展趨勢將和歷史趨勢保持一致的假設之上的，所以如果在投影期間內發生了中斷，那麼將會產生大量的預測誤差。

1. 輸入歷史資料然後選擇資料範圍(E13:E24)
2. 點擊Risk Simulator|預測|非線性外推
3. 選擇函數的類型和外推的期數然後選擇確定

歷史銷售額
多項式增長率

年份	月份	期數	銷售額
2010	1	1	\$1.00
2010	2	2	\$6.73
2010	3	3	\$20.52
2010	4	4	\$45.25
2010	5	5	\$83.59
2010	6	6	\$138.01
2010	7	7	\$210.87
2010	8	8	\$304.44
2010	9	9	\$420.89
2010	10	10	\$562.34
2010	11	11	\$730.85
2010	12	12	\$928.43

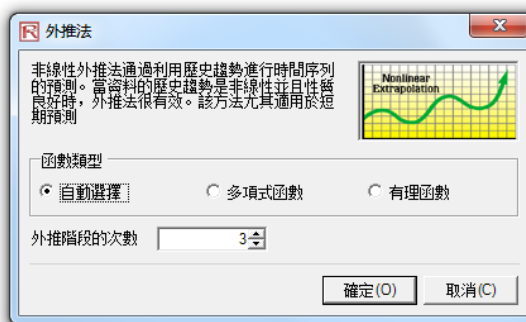


圖 3.11 運行非線性外推

統計匯總

外推法是通過對歷史資料的趨勢用統計學投影的方法來預測未來。它只能被用於時間序列預測中。對於截面資料或混合平板資料（時間序列和截面資料）來說，多元回歸更加適用。這種方法在不期望發生較大的變化，即期望因果關係保持不變時，或某一情景下的因果關係沒有明確確定時比較有用。它還可以防止將個人偏見引入過程之中。外推法相對比較可靠，簡單，所付的代價也相對較小。但是由於外推法是基於近期的發展趨勢將和歷史趨勢保持一致的假設之上的，所以如果在投影期間內發生了中斷，那麼將會產生大量的預測誤差。也就是說，時間序列的純粹外推假設我們所需要知道的所有資訊都包含在被預測序列的歷史資料之中。如果我們假設過去的行爲是未來行爲的一個好的預報器，那麼外推法就是適用的。故當所需要知道的所有資訊是一些短期預測時，外推法不失為一種有效的方法。

外推法估計了任意 x 值的 $f(x)$ 函數，通過在所有 x 值間插入一條光滑的非線性曲線，利用這條曲線，外推出一個基於此歷史資料集之外的未來的 x 處 $f(x)$ 的值。外推法可以採用多項式函數的形式或是有理數函數的形式（兩多項式之比）。一般情況下，對於性質好的資料多項式函數形式就足夠了，但是有時候使用有理數函數形式會更加精確（尤其是對於極值函數來說，例如，分母趨近於0的函數）。

時間點	真實值	預測擬合值	誤差度量
1	1.00		均方根誤差 19.6799
2	6.73	1.00	均方誤差 387.2974
3	20.52	-1.42	平均絕對離差 10.2095
4	45.25	99.82	平均絕對誤差率 31.56%
5	83.59	55.92	Theil's U 1.1210
6	138.01	136.71	
7	210.87	211.96	
8	304.44	304.43	
9	420.89	420.89	
10	562.34	562.34	
11	730.85	730.85	
12	928.43	928.43	
預測13		1157.03	
預測14		1418.57	
預測15		1714.95	

函數類型：有理

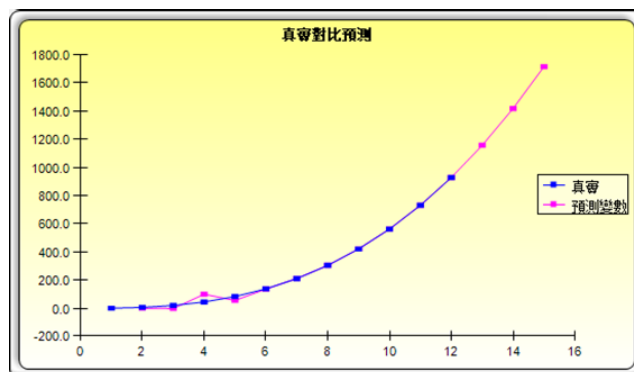


圖 3.12 非線性外推的結果

3.7 Box-Jenkins ARIMA 高級時間序列

理論：

一種非常有效的高級時間序列預測的工具是 ARIMA 或自回歸求和滑動平均過程。ARIMA 預測是由三種獨立的工具集成一種綜合性模型。第一個工具是自回歸或“AR”項，它對應的是無條件預測模型中殘差的滯後值。本質上，這個模型抓住了預測模型真實資料的歷史方差，然後利用這個方差或殘差來生成一個更好的預測模型。第二個工具是求和法或“I”項。這裏求和法指對時間序列資料差分的總和。這個因素可以解釋存在於資料中的任何非線性增長。第三個工具是滑動平均或“MA”項，它實際上就是滯後預測誤差的滑動平均。通過引入這個滯後預測誤差，模型可以通過滑動平均的計算來修正預測誤差或錯誤。ARIMA 模型遵循 Box-Jenkins 方法，其中的每一項都應對於模型構建中所採用的步驟，直到只剩下隨機因素。同時 ARIMA 模型在生成預測時還使用了相關性技術。ARIMA 模型可用於類比那些在資料繪圖中並不顯著的樣品。除此以外，ARIMA 模型還可與外生變數混合，但是要確保外生變數有足夠多的資料點來覆蓋需要預測的額外期數。最後需要注意鑒於模型的複雜性，可能所需的運行時間也比較長。

關於 ARIMA 模型優於普通時間序列分析和多元回歸的原因有很多。時間序列分析和多元回歸的普遍發現是殘差項與它們自身的滯後值是相關的。這種序列相關違背了回歸理論的標準假設，即擾動項之間是互不相關的。由序列相關所引發的主要問題有：

- 對各種不同的線性估計量，回歸分析和基本的時間序列分析並不有效。但是，由於殘差項可用于預測現有殘差，我們可以利用這一資訊優勢，使用 ARIMA 模型來預測因變數。
- 利用回歸和時間序列公式計算的標準差是不正確的，通常都被低估了，如果有滯後因變數被設置為回歸量，回歸估計就是有偏和不一致的，但是使用 ARIMA 就可以將它們固定。

自回歸求和滑動平均過程或 ARIMA (p,d,q) 模型是 AR 模型的延伸，使用 3 個成分在

時間序列資料中類比序列相關性。第一個組成成分是自回歸 (AR) 項。AR (p) 模型在等式中使用了時間序列的 p 階滯後。AR (p) 模型採用如下這種形式： $y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t$ 。第二個組成成分是求和法 (I)。每一種求和法對應於時間序列資料的不同差分的求和。如，I (1) 意味著資料的一次差分的求和。I (d) 意味著差分數據 d 次差分的求和。第三個成分是滑動平均 (MA) 項。MA (q) 模型利用預測誤差的 q 階滯後來改進預測結果。MA (q) 模型的形式： $y_t = e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}$ ，最後 ARIMA (p,d,q) 模型的綜合形式是：

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

步驟：

- 打開 Excel 表格，輸入資料或是打開工作表中現有的用於預測的歷史資料（下圖中使用的是示例文件夾中的*時間序列 ARIMA* 例子）
- 選擇時間序列資料，選擇**模擬預測ARIMA**
- 輸入相關的參數 P, D 和 Q（只能為正整數），然後輸入需要預測的期數，點擊**確定**

結果解析：

對於 ARIMA 模型結果解釋的大部分說明與多元回歸分析結果的解釋一致（請參見《類比風險》，Dr.Johnathan Mun 第二版來獲得更多關於多元回歸分析及 ARIMA 模型結果解釋的技巧）。但是如圖 3.14 中，有幾條額外的關於 ARIMA 分析的解釋。第一條是赤池資訊準則 (AIC) 和史瓦茲資訊準則 (SC)，它們通常被用於 ARIMA 模型選擇和識別。也就是說 AIC 和 SC 被用於確定

某個含 p 、 d 和 q 參數值的具體模型是否是一個好的統計擬合。SC 提供了對是否添加額外變數更強有力的判斷，通常應該選擇那些 AIC 值和 SC 值最低的模型。最後，ARIMA 報告裏還提供了一份額外的自相關 (AC) 和部分自相關 (PAC) 的統計結果。

例如，如果自相關 AC (1) 為非零值，意味著序列是一階序列相關的。如果 AC 隨著滯後的增加呈幾何下降趨勢，這意味著序列遵循一個低階自回歸過程。如果經過幾次滯後之後 AC 值趨於 0，這意味著序列遵循一個低階滑動平均過程。相反的，PAC 衡量了在移除了滯後干擾後的 k 階相關值。如果自相關模式可以用小於 k 階的自回歸描述，那麼 k 階滯後的部分自相關值趨近於 0。報告裏同時還提供了 Ljung-Box 的 k 階滯後 Q 統計值和 p 值，此時被檢驗的零假設是 k 階時不存在自相關。自相關圖裏的虛線近似表示了兩個標準差的區間界限。如果自相關值在此範圍之內，那麼在 5% 的顯著性水平內它不顯著區別於 0。尋找到合適的 ARIMA 模型需要嘗試和經驗。AC、PAC、SC 和 AIC 都是識別正確模型的有效診斷工具。

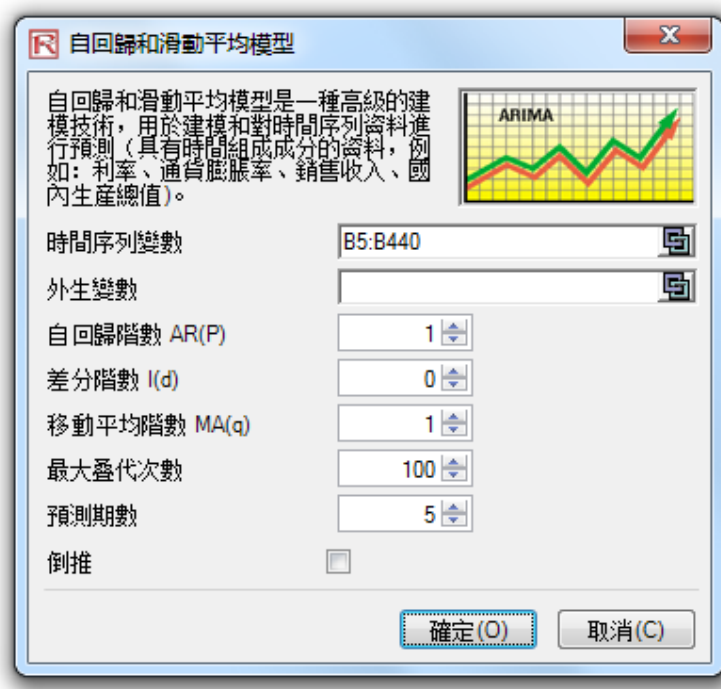


圖 3.13 Box-Jenkins ARIMA 預測工具

ARIMA (自回歸和滑動平均)

回歸統計量			
R方 (可決係數)	0.9999	赤池資訊準則 (AIC)	4.6213
調整後的R方	0.9999	史瓦茲資訊準則 (SC)	4.6632
多元R方 (多元相關係數)	1.0000	最大似然值	-1005.13
估計的標準差	297.52	杜賓 (DW) 統計量	1.8588
樣本個數	435	反覆運算次數	5

自回歸和滑動平均過程簡稱ARIMA (p,d,q) 模型擴展了AR模型，使用了三個部分來對描述時間序列資料的關係。第一部分是自回歸 (AR) 項。在AR (p) 模型中使用p來延遲時間序列。AR (p) 的形式為： $y(t)=a(1)y(t-1)+...+a(p)y(t-p)+e(t)$ 。第二部分是求和 (d) 階數項。每個求和階數對應時間序列的差分階數。I(1)為使用一階差分。I(q)為使用q階差分。第三部分是移動平均 (MA) 項。MA(q)模型使用加上q項預測誤差的延遲來改進預測。MA(q)模型的形式為： $y(t)=e(t)+b(1)e(t-1)+...+b(q)e(t-q)$ 。最後，ARMA(p,q)的形式為： $y(t)=a(1)y(t-1)+...+a(p)y(t-p)+e(t)+b(1)e(t-1)+...+b(q)e(t-q)$ 。

可決係數R方描述了本回歸中引數對因變數的解釋程度。但是在多元回歸中，調整的R方也考慮了其他引數和回歸量的存在，使得調整後的R方對回歸的解釋能力增強。但是在某些ARIMA仿真情況中 (例如，非收斂模型)，R方值就不可靠了。

多元相關係數 (多元R) 度量了真實因變數 (Y) 和基於回歸等式的估計或擬合值 (Y) 的相關程度。它是決定係數 (R方) 的平方根。估計標準差描述的是資料在回歸線或平面周圍上下偏移的離差。這個值稍後可用於計算和估計置信區間。

AIC和SC通常用於模型選擇中。SC提供了對是否添加額外變數更強有力的判斷。通常使用者應該選擇那些AIC值和SC值最低的模型。

Durbin-Watson統計量用來判斷回歸分析中的殘差項是否存在自相關性。總體上來說，DW小於2表明含有正自相關性。

回歸結果

	截距	AR(1)	MA(1)	
係數	-0.0626	1.0055	0.4936	
標準差	0.3108	0.0006	0.0420	
t統計量	-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	假設測試	
回歸自由度	2	臨界F值 (99%置信度, 自由度為432)
殘差自由度	432	臨界F值 (95%置信度, 自由度為432)
總體自由度	434	臨界F值 (90%置信度, 自由度為432)

係數提供了回歸模型的截距和斜率。舉例來說，這些係數提供了對真實值的一種估計；變數 (人口b) 的值可通過回歸方程 $Y=b_0+b_1X_1+b_2X_2+...+b_nX_n$ 來估計。標準差反映了預測的係數的精確程度，t統計量是預測的係數和它的標準差的比率。

t統計量被用在假設檢驗中。這裏我們設置零假設：係數的均值為0，備擇假設：係數的均值不為0。運用t檢驗並將計算的t統計量和殘差自由度的臨界值進行比較。t檢驗是非常重要的，它的計數值反映了當存在額外的回歸量時，每個係數是否在統計上是有效的。也就是說，t檢驗在統計上說明了一個回歸量或獨立變數是否應該保留在回歸模型中還是從模型中剔除。

如果係數的t統計量超過了相關自由度t統計量的估計臨界值，這些係數在統計上就是有效的。這裏，三種常用的置信水平是90%，95%和99%。如果係數的t統計量超過了臨界值，就認為該係數是統計上有效的。作為選擇，p值計算了取到t統計量的概率，也就是說p值越小，所對應的係數就越有效。P值對應90%，95%，99%三個置信水平最常用的有效臨界值選為0.01, 0.05和0.10。

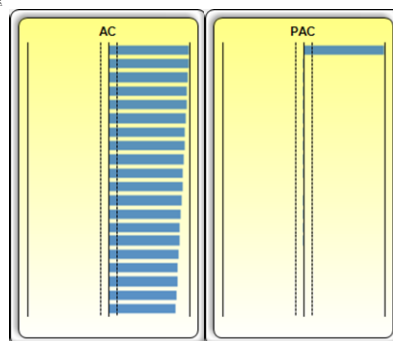
藍色的p值表明在90%的置信水平或0.10 alpha水平下，其對應的係數是統計上有效的。而紅色的p值表明其對應的係數在任何alpha水平下都不是統計上有效的。

方差分析

	平方和	均方	F統計量	P值	假設測試	
回歸	38415447.53	19207723.76	3171851.1	0.0000	臨界F值 (99%置信度, 自由度為2和432)	4.6546
殘差	2616.05	6.06			臨界F值 (95%置信度, 自由度為2和432)	3.0166
總和	38418063.58				臨界F值 (90%置信度, 自由度為2和432)	2.3149

方差分析表 (ANOVA) 提供了對回歸模型總體統計有效性的F檢驗。與t檢驗關注單個回歸量不同的是，F檢驗關注所有估計係數的統計屬性。F檢驗計算的是回歸的均值的平方和殘差的均值的平方的比例。分子表明回歸模型對預測值的解釋程度，分母表明有多少程度模型不能解釋的。因此，F—Statistic越大表明我們的模型就越有效。相應的p值也被計算用來假設檢測，判斷回歸模型的整體有效性。這裏零假設為：所有的係數都為0，備擇假設：所有係數不同時為0。如果p值在0.05或0.10的alpha有效水平下，比0.01小，則回歸模型是有效的。同樣的方法在對比計算的F—Statistic值和在各種有效水平下的臨界的F值時也能應用。

自相關						
時間延遲	AC	PAC	下屆	上屆	Q統計量	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	-



如果自相關項AC(1)不為0,則意味著該資料序列是一階自相關的。如果AC(K)隨著延遲的增加越來越小,則表明該資料序列服從一個低階自回歸過程。如果當延遲較小時,AC(k)高0,則意味著該序列資料服從一個低階移動平均過程。偏相關PAC(k)度量的是帶有k個時間點延遲的資料序列和不帶延遲的資料序列之間的相關性。如果它們的自相關性能被階數小於K的自回歸過程捕捉,那麼在延遲K上的自相關性將會趨向於0。Ljung-Box statistics統計量和它們的p值有零假設:當階數大於k時,不存在自相關性。左圖的兩條虛線是關於自相關性的邊界(通過近似標準差的值來確定邊界)。如果自相關性在這個邊界裏,則說明在5%有效性水平下,自相關性可近似視為0。

預測			
時間點	真實值(Y)	預測值(F)	誤差(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	(0.5232)
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)

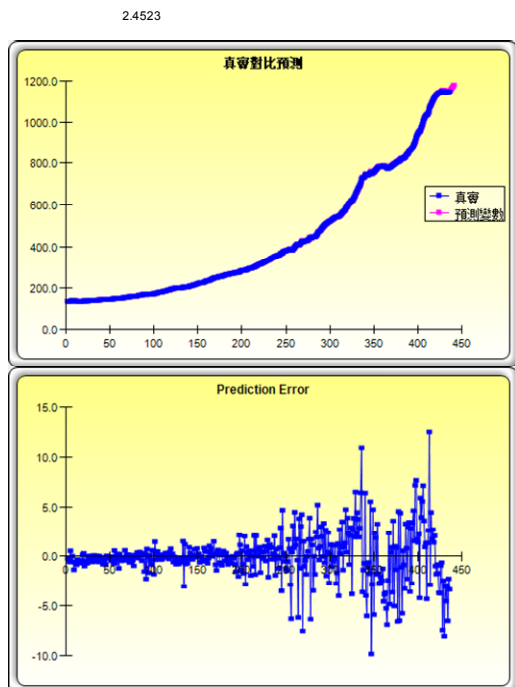


圖 3.13 Box Jenkins ARIMA 預測報告

3.8 自回歸求和移動平均模型(Box-Jenkins ARIMA 高級時間序列)

理論：

該工具和 ARIMA 模組一樣提供了相同的模型，不同點就是自回歸求和移動平均模型可以通過自動檢驗傳統模型中多個參數的排列方式，從而找到最優擬合的模型。運行自回歸求和移動平均模型的方式和普通的 ARIMA 預測方式相同。不同點就是不再需要輸入模型的參數 P,D,Q，而且對這些參數的組合與比較都是自動完成的。

操作過程描述：

- 運行 Excel 輸入資料或者打開包含歷史資料需要預測的工作簿（如圖 3.14 顯示的示例文件）。[高級預測模型](#)在 Risk Simulator 的[示例文件](#)中
- 在 ARIMA 預測 工作簿中,選擇 Risk Simulator | 預測 | 自回歸求和移動平均
- 點擊鏈結圖示選擇現有的時間序列資料登錄需要預測的期數點擊 **確定**

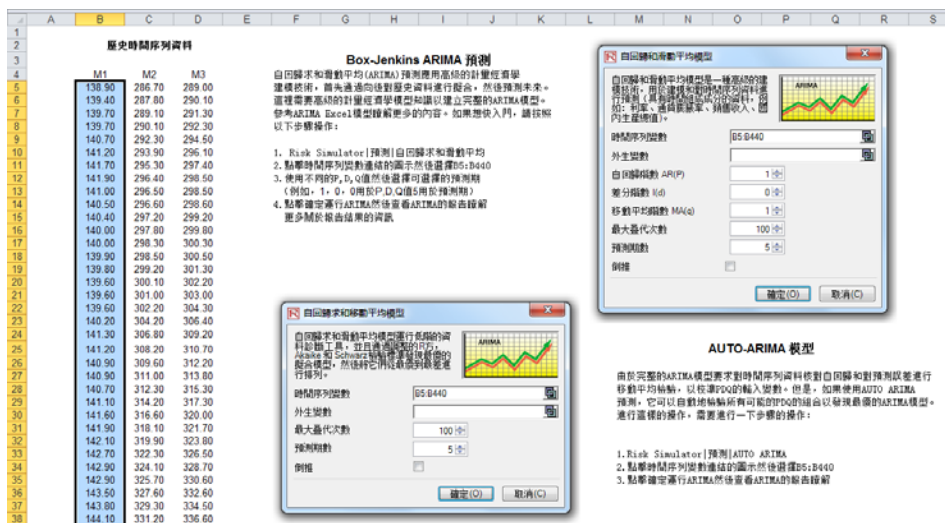


圖 3.14 自回歸求和移動平均模型

3.9 基本計量經濟學模型

理論：

計量經濟學涉及一系列的商業分析，建模和預測技術用於類比或者預測某個商業或者經濟變數。運行基本的計量經濟學模型類似於一般的回歸分析，而基本的計量經濟學模型中的變數和因變數在進行回歸前需要被修正。生成的報告類似於前面的多元回歸部分，對於報告的解釋也類似於多元回歸部分的解釋。

操作過程描述：

- 運行 Excel 輸入資料或者打開包含歷史資料需要預測的工作簿（如圖 3.15 顯示的示例文件）。[高級預測模型](#)在 Risk Simulator 的[示例文件](#)中
- 選擇基本計量經濟學資料 工作簿中的資料列，點擊 Risk Simulator | 預測 | 基本計量經濟學模型
- 輸入引數和因變數（如圖 3.15 所示）點擊**確定**運行模型和生成報告，或者點擊**顯示結果**，在生成報告之前來預覽結果，以免有對模型需要改進的地方。

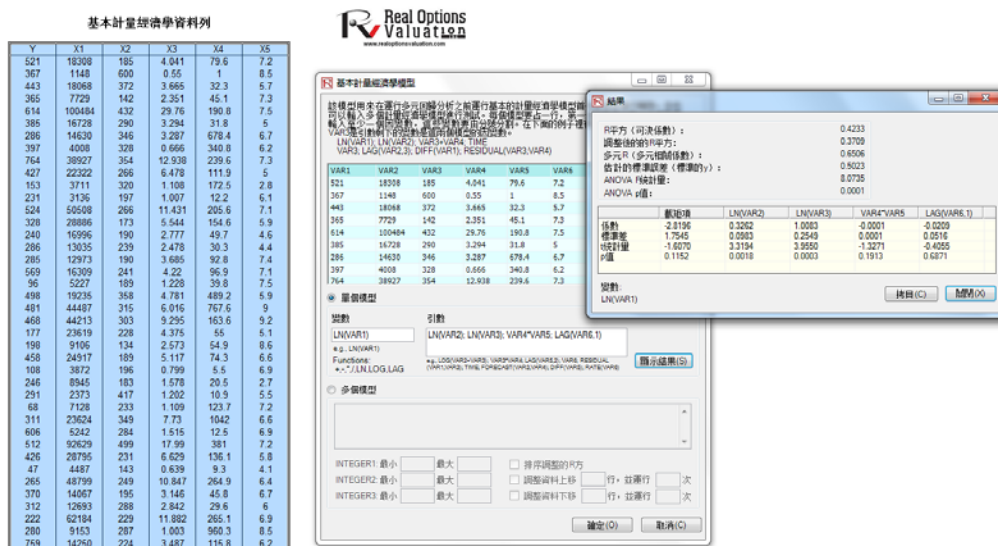


圖 3.15 基本計量經濟學模型

3.10 J-S 曲線預測

理論：

J-曲線或者指數增長曲線描述了這樣的增長現象，下一期的增長取決於本期增長的水平，而下一期將呈指數式地增長。這意味著隨著時間的增長，從一段時間到另一段時間，數值將很快地增加。該模型一般用於預測隨著時間變化生物的增長和化學的裂變。

操作過程描述：

- 運行 Excel 然後選擇 Risk Simulator | 預測 | JS 曲線
- 選擇 J 或者 S 曲線的類型輸入需要的輸入假設（如圖 3.16 和 3.17）點擊 **確定** 運行模型和生成報告

指數J增长曲线

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

生成J曲线预测，需要遵循下面的步骤：

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

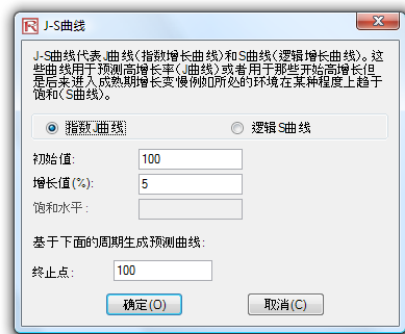


圖 3.16 J-曲線預測

S-曲線或者邏輯增長曲線開始啟動的時候類似於 J-曲線，具有指數式地增長。隨著時間地增長，環境不斷地飽和（例如，市場的飽和，競爭的加劇，堵塞），增長變得緩慢，最後預測值終止於飽和值或者最大水平。這個模型一般用於預測市場份額或者新產品從引入期到衰退期的銷售增長變化。圖 3.17 顯示了示例的 S-曲線。

邏輯函數或邏輯曲線可對某些變數的增長曲線建模。在初始極端的增長相當於指數增長，隨著競爭不斷升級，增長率不斷地降低，當處於飽和期時停止增長。這些函數可在生物學到經濟學的很多領域廣泛運用。舉例來說，在胚胎生長階段，受精卵分裂，細胞數快速增長：1，2，4，8，16，32，64等等。這是指數增長。但是胎兒只能長到子宮所能容納的大小：這樣別的因素就會減小細胞數的增長，增長率就下降了（當然，胎兒仍在成長）。經過一段時間後，胎兒被產出並且保持成長。最後，細胞數達到穩定；個體的高度變為常數，在成熟期，成長停止了。同樣的原理可被用到人類的人口增長和動物的種群增長理論中，也可用到市場滲透理論和產品的收益當中。在市場滲透中有初始的增長率，但隨著競爭的加強，市場的增長率下降，在成熟期市場達到飽和並保持它的規模。

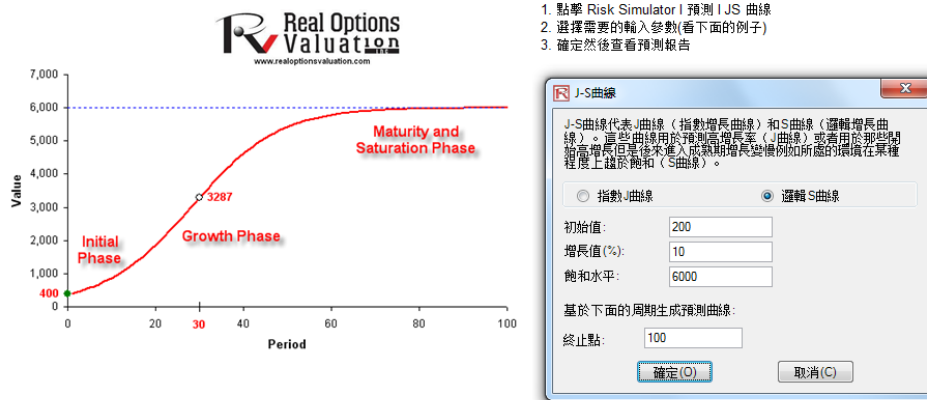


圖 3.17 S-曲線預測

3.11 GARCH 波動率預測

理論：

GARCH 模型可用在計算流動資產和可買賣資產的波動率上，比如金融期權裏的股票波動率；該模型有時也用在另外的可買賣資產上，比如石油和電力的價格。其缺點是需要大量資料，需要高級金融模型專家的意見並且易受人為操作的影響。該模型的優點是使用了嚴格的統計分析來得到最優擬合的波動率曲線，提供了隨時間變化的波動率估計。更多對於 GARCH 模型的討論超出了用戶手冊的範圍。關於 GARCH 模型的描述請參考 Johnathan Mun 博士所編著的“Advanced Analytical Models,” (Wiley 2008)一書。

操作過程描述：

- 運行 Excel 然後打開示例文件高級預測模型，轉到 廣義自回歸條件異方差模型 (GARCH) 工作簿選擇 Risk Simulator | 預測 | GARCH
- 點擊鏈結圖示選擇資料範圍區域，輸入需要的輸入假設（如圖 3.18）點擊 **確定** 運行模型和生成報告。

注意：一般的波動率預測要求 $P = 1, Q = 1$ ，周期數=每年的期數（12 用於月資料，52 用於周資料，252 或者 365 用於日資料），基數=1 到周期數之間的某個值，預測期=想要預測的波動率的期數。

(GARCH)廣義自回歸條件異方差模型(GARCH)

歷史資料

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

運行GARCH模型，輸入相應的時間序列資料，然後點擊Risk Simulator|預測|GARCH然後點擊資料鏈接圖示，選擇歷史資料範圍(例如，C8:C2428)。輸入要求的輸入變數(例如，P 1，Q 1，日交易週期為252，預測基數為 1， 預測期為10)然後點擊確定。然後查看生成的預測報告。



圖 3.18 GARCH 波動率預測

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

3.12 馬爾可夫鏈

理論：

當未來的狀態的概率水平取決於前期狀態的概率水平，從長遠的角度將它們彼此鏈結就形成了一條鏈，這條鏈就被稱為馬爾可夫鏈。該方法一般被用來預測具有市場份額的兩個競爭對手。需要輸入的輸入變數為顧客在第一家商店（第一種狀態）的概率值，顧客下次可能會去同一家商店，也可能轉向競爭對手的商店。

操作過程描述：

- 運行 Excel 然後選擇 Risk Simulator | 預測 | 馬爾可夫鏈
- 輸入需要輸入的假設（如圖 3.19）然後點擊 **確定** 運行模型和報告。

馬可夫過程在連續時間段的多次和重複試驗演進系統的研究中非常有用。該系統在特定時間的狀態是未知的，而我們感興趣的是出現這種特定狀態的概率。舉例來說，馬可夫鏈可用來計算特定的機器或設備將在下一個時期繼續工作的概率，或者計算客戶購買了A產品將在下一個時期繼續購買A產品或選中購買B品牌產品的競爭性產品的概率。這些品牌或產品記作處於某種狀態。通過輸入經過一段時間個體仍在狀態1的概率和現在在狀態2的個體在1期後仍在狀態2的概率，可以使用馬可夫鏈方法來確定未來的概率或所處的狀態和長期穩定的狀態（比如說最後的市場佔有率）。

生成一個馬可夫過程，需要進行以下的操作：

1. 點擊 Risk Simulator | 預測 | 馬爾可夫鏈
2. 輸入相應狀態的概率水準(例如，90%和80%) 然後點擊確定。
3. 查看生成的報告

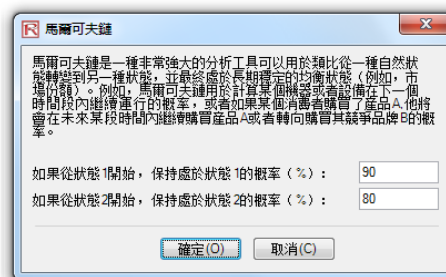


圖 3.19 馬爾可夫鏈（轉換體系）

3.13 最大似然估計模型（MLE）

理論：

有限因變數描述了這樣的情形因變數包含的資料在範圍上都是有限的，例如二元回應資料（0 或者 1），截取的資料，有序或者截斷資料。例如，假設有一系列獨立變數（例如，信用卡或者按揭貸款人的年齡，收入，教育水準），可以通過最大似然模型（MLE）類比違約發生率。響應或者因變數 Y 是二元的，即，只有兩種可能性發生，以 1 和 0 表示（例如，Y 表示有/沒有的條件，貸款違約/沒有違約，一些設備成功/失敗，問卷調查是/否，等等），這樣就有回歸獨立變數 X，假定會對 Y 產生影響。典型的普通最小二乘法對於此類問題變得失去效應，因為回歸誤差項具有非正態性和異方差性，估計的結果會失去意義，大於 1 或者 0。MLE 分析解決了這些問題，在因變數是有限的情況下，使用優化原則最大化對數可能性函數。

.Logit 或者邏輯回歸通過將資料擬合成邏輯曲線來預測事件的發生概率。它是一個廣義線性模型用於二元回歸，同大多數形式的回歸分析類似，借助於大量預測變數。MLE 應用二元多變數邏輯分析，該分析用於類比因變數來

決定某個可能組的期望成功概率。估計的係數都是對數比例的形式，不能直接解釋為概率。一個簡單的調整計算可以估計出最後的概率。

邏輯模型被定義為 $Y = \text{LN}[\text{Pi}/(1 - \text{Pi})]$ 或者相應的, $\text{Pi} = \text{EXP}(Y)/(1 + \text{EXP}(Y))$ 係數 β_i 是對數比例，通過反對數可以得 $\text{EXP}(\beta_i)$ 得到比率 $\text{Pi}/(1 - \text{Pi})$ 。可能性水準的變動率為 $dP/dX = \beta_i \text{Pi}(1 - \text{Pi})$ 。標準誤差估計了預測的精度係數，t 統計量是每個預測係數與標準誤差的商，是典型假設檢驗的參數用來估計每個估計量的顯著性水準。為了預測每個特定組的成功概率（例如，基於一個吸煙者每年吸煙的數量來預測一個吸煙者是否會有胸部的併發症），可以通過使用 MLE 係數來簡單計算估計的 Y 值。例如，假如模型是 $Y = 1.1 + 0.005 \times (\text{香煙})$ ，那麼某人每年吸煙 100 盒就有一個估計的 Y 等於 $1.1 + 0.005 \times (100) = 1.6$ 。因此，計算可能比率的反逆對數 $\text{EXP}(\text{估計 } Y)/(1 + \text{EXP}(\text{估計 } Y)) = \text{EXP}(1.6)/(1 + \text{EXP}(1.6)) = 0.8320$ ，可得，這個人在他的一生中有 83.20% 的可能會感染胸部併發症。

Probit 模型（也稱為 Normit 模型）作為對於二元回應模型的替代模型，應用 probit 函數估計使用最大似然估計和 probit 回歸。Probit 和邏輯回歸模型都是類似的預測模型，邏輯回歸的參數估計比 Probit 模型高出 1.6 到 1.8 倍。對於兩種模型的選擇都是十分方便的，主要的不同就是邏輯分佈都是具有高偏性的（肥尾效應）用來描述極值現象。例如，要類比房屋的所有權，回應變數是二元的（購房或者不購房），影響變數 X_i 為收入，年齡，等等，例如， $li = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$ ， li 的值越大，購房的可能性越高。對於每個家庭，關鍵的 I* 門檻值存在，如果超過，就購買住房，否則將不購買住房，影響概率（P）假設是正太分佈的，例如 $\text{Pi} = \text{CDF}(li)$ 使用標準正態累積概率密度分佈函數（CDF）。因此使用正態分佈估計係數（可以使用 Excel 中 NORMASDIST 函數或者 RiskSimulator 中分佈分析工具，選擇正態分佈設定均值為 0，標準差為 1）。最後，獲得 Probit 或者概率單位估計，設定 $li + 5$ （這是因為只要概率 $\text{Pi} < 0.5$ ，估計都是為負，這是因為正態分佈是在均值 0 左右對稱分佈的）。

Tobit 模型（截段模型）是計量經濟學或者生物統計模型用來描述非負因變數 Y_i 和一個或者多個引數 X_i 之間的關係。所謂截斷模型，是因為因變數由於小於 0 不被觀測到而被截斷。Tobit 模型假設存在潛在的未觀測到的因變數 Y^* 。這個變數與引數 X_i 線性相關，相關係數 β_i 的向量決定他們內部的關係。此外，存在正態分佈誤差項 U_i ，對他們的關係產生隨機影響。觀測變數 Y_i 等於潛在變數，當潛在變數大於零的時候或者 Y_i 假設等於零的時候。即， $Y_i = Y^*$, if: $Y^* > 0$ 以及 $Y_i = 0$, if: $Y^* = 0$ 。如果關係參數 β_i 通過最小二乘回歸得到，結果的回歸估計量就不一致，向下的偏移斜率和向上的偏移截距。只有 MLE 可以符合 Tobit 模型，存在輔助的統計量稱為 sigma，等同於標準最小二乘法的標準誤差的估計，估計的係數類似於回歸方法得到。

操作過程描述:

- 運行 Excel 然後打開示例文件高級預測模型, 轉到二元最大似然預測工作表, 選擇包含標題的資料列 Risk Simulator | 預測 | 最大似然估計
- 從下拉功能表學者因變數 (參看圖 3.20) 然後點擊**確定**運行模型和生成報告

二元最大似然邏輯預測

違約	年齡	教育水準	現任職位工作年數	現任地址居住年數	家庭收入(千元 \$)	負債收入比(%)	信用卡負債(千元 \$)	其他負債(千元 \$)
1	41	3	17	12	176	9.3	11.36	5.01
0	27	1	10	6	31	17.3	1.36	4
0	40	1	15	14	55	5.5	0.86	2.17
0	41	1	15	14	120	2.9	2.66	0.82
1	24	2	2	0	28	17.3	1.79	3.06
0	41	2	5	5	25	10.2	0.39	2.16
0	39	1	20	9	67	30.6	3.83	16.67
0	43	1	12	11	38	3.6	0.13	1.24
1	24	1	3	4	19	24.4	1.36	3.28
0	36							
0	27							
0	25							
0	52							
0	37							
0	48							
1	36							
0	43							
0	39							
0	41							
0	39							
0	47							
0	28							
0	29							
1	21							
0	25							
0	45							
0	43							
0	33							
0	26							
0	45							
0	30							
0	27							

Real Options Valuation
www.realoptionsvaluation.com

這裡的資料顯示了各種各樣的貸款, 債務, 或者負債問題。資料顯示了每筆貸款會發生違約現象, 包括貸款申請人的年齡, 教育水準(1-3說明高中, 大學, 或者研究生教育背景), 現任職位任職年數等等。現在對現有的歷史資料進行建模以發現哪些因素會對個體的違約行為發生影響。使用Risk Simulator的最大似然估計模型。結果模型會說明銀行或者信貸擔保者計算具有不同特點的信貸申請者的期望的違約率。進行該分析, 選擇左側的資料或者任何其他資料列(包括資料列的標題)然後確定資料列具有相同的長度, 不會有任何的缺失或者無效的資料。然後, 點擊Risk Simulator | 預測 | 最大似然模型。一系列的結果會顯示在RIS工作表中, 完成上述操作, 對個體的違約率進行計算。

圖 3.20 最大似然估計模型

3.14 樣條模型 (三次樣條內插和外插模型)

理論:

有時候在時間序列資料列中會有缺失值。例如, 1 到 3 年的利率存在, 下面就是 5 到 8 年的資料, 和第 10 年的資料。樣條曲線可以用來預測或者外推未來某個期間的資料。資料可以是線性也可以是非線性的。如圖 3.21 顯示了三次樣條是如何運行的, 圖 3.22 顯示了這個模組的預測報告。已知 X 值顯示在圖表的 x 軸 (在本例中, 是年份), 已知的 Y 顯示在 y 軸 (本例中, 是已知的利息率)。

三次樣條多項式內插和外推法

三次樣條多項式內插和外推法模型是用來“填滿”缺失值的缺口以及預測時間序列的資料，模型可以用來內插時間序列資料（例如，收益曲線，利率，宏觀經濟變數如通脹率，商品價格或者市場回報率）也可以在已知的或給定的範圍內外推資料，使之更好地用於預測。

運行三次樣條預測，需要點擊Risk Simulator | 預測 | 三次樣條插值然後點選連結圖示選擇C15:C25作為已知的X變數（時間序列圖的X軸）然後選擇D15:D25作為已知的Y變數（確定已知的變數X和變數Y的資料長度相同）。輸入想要預測的期數（例如，開始1，完成50，步長0.5）。點擊確定然後查看生成的預測結果和預測圖。

年份	收益點
0.0833	4.55%
0.2500	4.47%
0.5000	4.52%
1.0000	4.39%
2.0000	4.13%
3.0000	4.16%
5.0000	4.26%
7.0000	4.38%
10.0000	4.56%
20.0000	4.88%
30.0000	4.84%

這裡是已知的收益率作為三次樣條多項式內插和外推模型的已知輸入變數

三次樣條插值

三次樣條多項式內插和外推法模型是用來“填滿”缺失值的缺口以及預測時間序列的資料，模型可以用來內插時間序列資料（例如，收益曲線，利率，宏觀經濟變數如通脹率，商品價格或者市場回報率）也可以在已知的或給定的範圍內外推資料，使之更好地用於預測。

已知變數X:

已知變數Y:

基於下面的X值生成的樣條曲線

開始: 完成: 步長:

圖 3.21 三次樣條模組

操作過程描述:

- 運行 Excel 然後打開示例文件高級預測模型，轉到三次樣條多項式內插和外推法工作表，選擇包含標題的資料列 Risk Simulator | 預測 | 三次樣條插值
- 點擊鏈結圖示鏈結已知的 X 值和已知的 Y 值（參看圖 3.21），輸入開始值和完成值。點擊**確定**運行模型和報告（參看圖 3.22）。

三次樣條插值預測

三次樣條多項式內插和外推法模型是用來“填滿”缺失值的缺口以及預測時間序列的資料，模型可以用來內插時間序列資料（例如，收益曲線，利率，宏觀經濟變數如通脹率，商品價格或者市場回報率）也可以在已知的或給定的範圍內外推資料，使之更好地用於預測。

三次樣條內插和外推結果

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

下面是三次樣條內插和外推模型的已知變數：

觀測值	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	4.84%
11	30.0000	4.84%

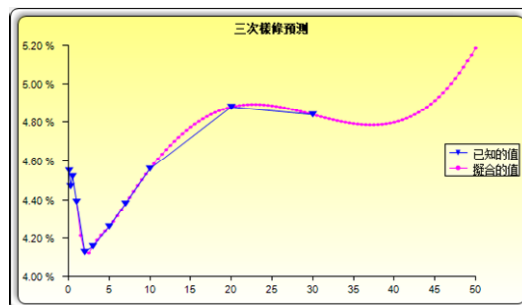


圖 3.22 樣條預測的結果

4

4 · 最佳化

本章節我們會更詳細地瞭解最佳化的過程和方法，它也是屬於 **Risk Simulator** 的功能之一。進行最佳化的方法包括連續型最佳化和離散型最佳化，以及靜態，動態和隨機最佳化。

4.1 最佳化方法

運行最佳化時存在很多運算法則，當最佳化與 **Monte Carlo** 模擬一起使用時存在很多不同的步驟。在 **Risk Simulator** 中，存在三種不同的最佳化過程、最佳化類型以及不同的決策變數類型。例如，**Risk Simulator** 可以處理**連續決策變數**（1.2535，0.2215 等等），**離散決策變數**（例如，1，2，3，4 或 1.5，2.5，3.5 等等），**二元決策變數**（用於停/走決策的 1 和 0 變數），**混合決策變數**（整數變數和連續變數）。**Risk Simulator** 還可以處理**線性最佳化**（例如，當目標和約束條件都是線性公式和函數時），以及**非線性最佳化**（例如，當目標和約束條件是線性和非線性函數和公式的時候）。

對於整個最佳化過程，**Risk Simulator** 首先可用於運行**離散最佳化**，也就是基於離散或靜態模型的最佳化，此時不運行模擬。換句話說，模型裏的所有輸入量都是靜態的和確定的。當模型包含的因素都是確定的時就可以使用這種最佳化類型。再則，在運行更高級的最佳化程式之前可以首先運行一個離散最佳化來確定可選方案及相關的決策變數的可選分佈。例如，在運行隨機最佳化問題之前，我們可以先運行一個離散最佳化以確定在進行更深入的分析之前是否存在其他解決最佳化問題的方案。

當 **Monte Carlo** 模擬和最佳化一起使用時，我們就要運用**動態最佳化**。此過程的另一個名稱叫做**模擬-最佳化**。也就是說，先運行一次模擬，接著在 **Excel** 模型中運用這些模擬結果，並對這些模擬結果進行最佳化。換句話說，在 **N** 次試驗中運行模擬，然後在 **M** 次疊代中運行最佳化程式直到得到一個最佳結果或發現一個不可行集。利用 **Risk Simulator** 的最佳化模型時您可以選擇要使用哪些預測和假設統計量，在運行模擬後的模型中替換。然後，這些預測統計量可以應用於最佳化程式中。在模型很大，裏面包含很多相互關聯的假定和預測，以及最佳化中要求某些預測統計量的情況下，這種

方法很有效。例如，如果在最佳化模型中需要假定或預測標準差（利用均值除以資產組合的標準差來計算資產分配和最佳化問題的夏普指數（Sharpe-Ratio）），此時就應該使用這個方法。

除了整個動態最佳化過程要重複 T 次以外，**隨機最佳化**過程與動態最佳化類似。也就是說，在 N 次試驗中運行模擬，然後在 M 次疊代中運行最佳化程式直到得到一個最佳結果。然後將上述過程重複 T 次。最後會生成一個每個決策變數有 T 個值的預測表。換句話說，運行模擬以及在最佳化模型中使用預測或假設統計量來尋找決策變數的最佳分配。然後，再運行一次模擬，生成不同的預測值，最佳化這些新更新值。因此最終每個決策變數都會有一個單獨的預測表，說明最佳決策變數的範圍。例如，您現在可以在動態最佳化程式中獲得決策變數的分佈，進而得到每個決策變數的最佳值範圍。該過程也被稱為隨機最佳化，它不僅僅是得到單點估計值。

最後，最佳化中還包括一種有效前沿最佳化過程，它在最佳化中應用了邊際增量和影子定價理論。意思就是，如果放寬其中一個約束條件，最佳化的結果會發生什麼變化？舉個例子來說，如果假定預算約束是 100 萬美元。那麼當它變為 150 萬，200 萬等時，資產結果和最佳決策會發生什麼變化？這就是金融投資領域的馬可維茨前沿概念，也就是當資產的標準差被允許有輕微的增大時，資產會產生哪些額外收益？除了只允許改變一個約束條件，並且在每次變化時，模擬和最佳化程式仍在運行以外，該過程與動態最佳化過程類似。運行這個過程最好利用 Risk Simulator 手動進行。也就是先運行一個動態或隨機最佳化，再在一個約束條件下重新運行另一次最佳化，最後重複幾次這個過程。這個手動過程是非常重要的，因為分析者可以在改變約束條件的情況下，發現結果是類似還是不同，以此來確定是否需要進其他額外的分析，或是確定為了得到目標和決策變數的一個顯著變化，約束所需的邊際增加程度。

有一點值得注意。有其他一些軟體在表面上可以進行隨機最佳化，但實際上它們是不可以的。例如，它們可能是在運行一次模擬後，生成了一個最佳化過程的疊代，再運行另一次模擬，生成第二個最佳化疊代，依此類推，這樣非常浪費時間和資源。在最佳化過程中，模型要經過一系列嚴格運算法則的檢驗，此時我們就需要運用多重疊代（範圍從幾次到上萬次疊代）來得到最佳的結果。因此，每一次隻生成一次疊代非常浪費時間和資源。與這種需要幾個小時的方法相比，運用 Risk Simulator 可以在一分鐘內得到同樣的資產組合結果。同時這種模擬-最佳化方法一般會得到不好的結果，並不是一種隨機的最佳化方法。當在我們的模型中應用最佳化方法時一定要非常注意這類方法。

接下來我們來看看兩個最佳化問題的例子。一個例子使用的是連續型決策變

數，另一個使用的是離散整數型決策變數。兩個模型均可應用離散最佳化，動態最佳化，隨機最佳化，甚至影子定價的最佳邊界等方法。因此為了簡便起見，我們僅僅介紹一下模型的設置，由使用者自己決定到底運行哪種最佳化程式。另外，連續型模型使用非線性最佳化方法（因為用於計算的資產組合風險是一個非線性函數，所以用資產回報除以資產風險表示的目標也是非線性函數），第二個例子中的整數型最佳化則使用線性最佳化模型（它的目標和所有約束條件都是線性的）。因此，這兩個例子囊括了上述所有的方法。

4.2 連續型決策變數的最佳化

圖 4.1 中是連續型最佳化模型的例子。可以依次點擊**開始程式**Real Option ValuationRisk Simulator**示例**找到連續型最佳化文件。在本例中一共存在 10 種典型的資產類型（例如，不同種類的共同基金，股票或資產），目標在於如何最有效的分配這些資產以取得最佳的效果。也就是說給定每種資產的固有風險，如何得到可能的最佳資產收益。為了更好的理解最佳化的概念，我們必須進一步研究這個示例模型來學習如何更好的利用最佳化程式。

模型中顯示了十種資產的分類及每種資產的年收益率和年波動率。這些收益和風險都是以年均值來表示的，以便不同的資產類型之間可以相互進行比較。收益率用相關收益的幾何均值來計算，而風險則使用相關股票收益對數法來計算。更多關於股票或其他資產年波動率和年收益率計算的詳情請參見本章的附錄。

A	B	C	D	E	F	G
2						
3						
4		資產配置優化模型				
5	資產描述	年收益率	風險波動率	分配權重	最小分配額	最大分配額
6	資產1	10.54%	12.36%	10.00%	5.00%	35.00%
7	資產2	11.25%	16.23%	10.00%	5.00%	35.00%
8	資產3	11.84%	15.64%	10.00%	5.00%	35.00%
9	資產4	10.64%	12.35%	10.00%	5.00%	35.00%
10	資產5	13.25%	13.28%	10.00%	5.00%	35.00%
11	資產6	14.21%	14.39%	10.00%	5.00%	35.00%
12	資產7	15.53%	14.25%	10.00%	5.00%	35.00%
13	資產8	14.95%	16.44%	10.00%	5.00%	35.00%
14	資產9	14.16%	16.50%	10.00%	5.00%	35.00%
15	資產10	10.06%	12.50%	10.00%	5.00%	35.00%
16						
17	組合總量	12.6419%	4.58%	100.00%		
18	風險收益率	2.7596				
19						

圖 4.1 連續型最佳化模型

E 列的權重對決策變數有要求，需要檢查變數的總份額為 100%（單格 E17）。一般情況下在開始最佳化時，我們會將這些單格設成統一的值，在本例中，從單格 E6 到 E15 我們均設為 10%。除此以外，每個決策變數的變化範圍還

有具體的限制。在本例中，我們從 F 和 G 列中可以看出最低權重和最高權重分別為 5% 和 35%。這意味著每類資產都有自己的權重邊界。H 列代表的是風險收益率，即收益率除以風險率，這個值越大，花同樣的錢取得的效果就越大。模型的餘下部分顯示了個別資產的收益率、風險、風險收益率和份額排名。換句話說，這些排名告訴我們哪類資產擁有低風險或高收益。

投資的資產總體收益率在單格 C17 中用命令 Sumproduct (C6:C15, E6:E15) 來計算，那也就是份額乘以每種資產的年收益率的和。我們可以用如下等式表示： $R_p = \omega_A R_A + \omega_B R_B + \omega_C R_C + \omega_D R_D$ ，其中 R_p 代表資產總的收益率， $R_{A,B,C,D}$ 是專案的個體收益率， $\omega_{A,B,C,D}$ 代表每個專案的權重或資金份額。

另外，單格 D17 中資產組合的風險通過等式：

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

來計算。這裏， ρ_{ij} 代表資產種類之間的相關係數——因此，如果相關係數為負，那麼它們就有分散風險的作用，資產的風險就降低了。但為了簡化計算，我們假設在計算資產風險時，資產之間的相關性為 0，只在稍後的收益模擬中假設存在相關性。這樣就不用在這不同的資產收益之間運用靜態相關性，而是只在模擬假設時應用相關性，在模擬的收益值之間創造出一種更為動態的關係。

最後還要計算資產的風險收益率或夏普指數 (Sharpe-Ratio)。這個值位於 C18 單格，代表本最佳化模型中的最大化目標。總結一下，本模型包含如下一些屬性：

目標：最大化風險收益率 (C18)

決策變數：權重 (E6:E15)

決策變數的限制：最小值和最大值 (F6:G15)

約束條件：資產總權重為 100% (E17)

步驟：

- 打開案例文件通過點擊模擬新建模擬文檔新建一個新文件，並命名。
- 最佳化的第一步就是設置決策變數。選定單格 E6 開始設置第一個決策變數 (**模擬最佳化設置決策變數**)，點擊連接圖表選擇命名單格 (B6)，同時在單格 F6 和 G6 中選擇上限和下限。然後，使用 Risk Simulator 的複製功能，複製單格 E6 的決策變數，並將其粘貼到剩下的單格 E7 至 E15 中。

- 最佳化的第二步是設置約束條件。這裏只有一個約束條件，就是資產的總份額之和為 100%。點擊**模擬|最佳化|約束條件**...選擇增加添加一個新約束。然後選中單格 E17，輸入=100%。完成後點擊**確定**。
- 最佳化的最後一步是設置目標函數，選擇目標單格 C18，依次點擊**模擬|最佳化|運行最佳化**，然後選擇最佳化類型（靜態最佳化，動態最佳化或者隨機最佳化），最後就可以開始模擬了。開始可以選用靜態最佳化。再檢查確保已經在單格 C18 中設置了目標，並選擇最大化。如果需要的話還可以檢查一下決策變數和約束條件，或是直接點擊**確定**運行靜態最佳化。
- 最佳化完成之後，您可以選擇**恢復**來回到決策變數和目標的原始值，或是選擇**替換**直接利用最佳化後的決策變數。一般情況下，我們會在最佳化完成後選擇**替換**功能。

圖 4.2 是以上這些步驟的圖示。可以在模型的收益和風險（C 列和 D 列）處增加模擬假定，然後使用動態最佳化和隨機最佳化。



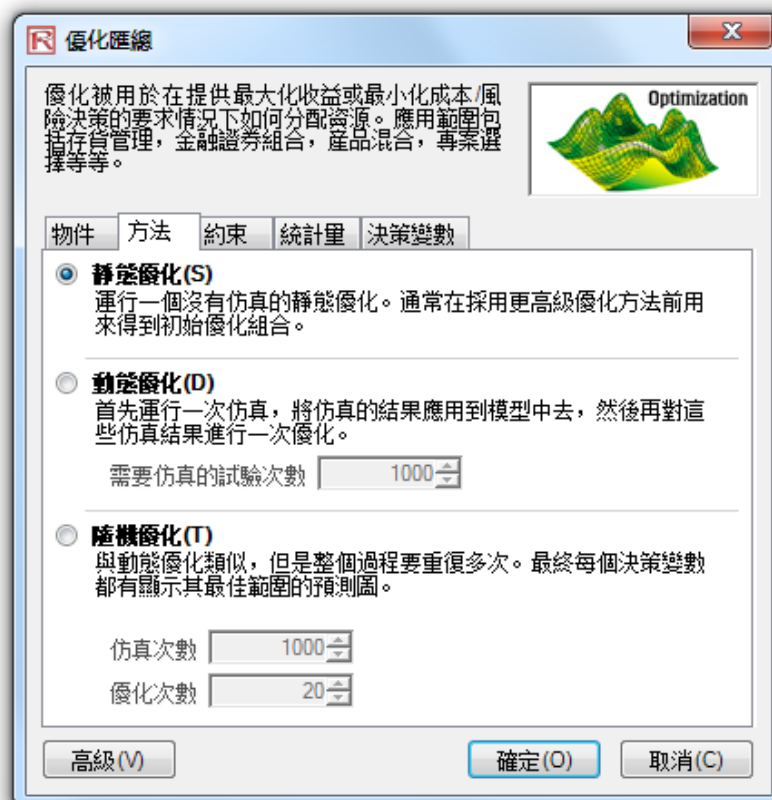
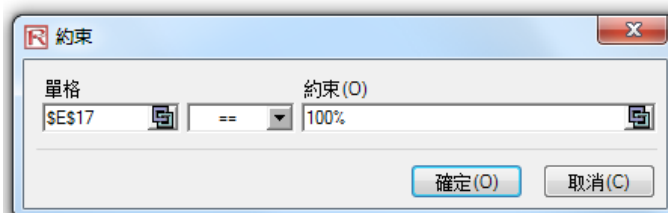
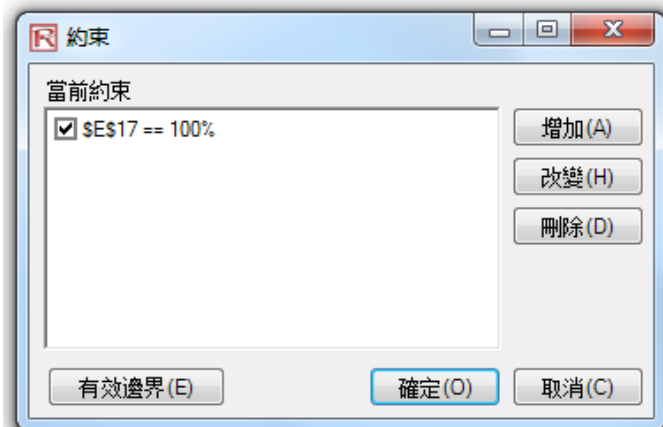


圖 4.2 在 Risk Simulator 中運行連續型最佳化

結果解析：

圖 4.3 中顯示了最佳化的最終結果，資產的最佳組合參見單格 E6 至 E15。也就是說給定每種資產份額在 5%到 35%之間波動以及份額總和為 100%這兩個限制條件，圖 3.4 中顯示了使得風險收益率最大的資產組合。

在回顧最佳化過程和結果之前一些值得注意的事項：

- 運行最佳化的正確方式是最大化風險收益率（夏普指數（Sharpe-Ratio））或最大化花同樣錢獲得的收益。
- 如果我們不要求最大化整體組合的收益，那麼最佳組合結果就沒有太大意義，我們也就不需要最佳化這一步驟了。也就是說，收益最低的 8 類資產的份額為 5%（允許的最小值），最高收益資產的份額設為 35%，剩下的 25%分配給收益次優的資產。不需要進行最佳化。但是，當這樣投資分配時，其風險比最大化風險收益率時的風險高很多，儘管此時資產組合的收益更高些。
- 相反的，您可以使得整體資產組合的風險最小，此時對應的收益也會小些。

表格 4.1 是三種不同目標情況下運行最佳化的結果

目標	資產組合 收益	資產組合風 險	資產組合的風險收益 率
最大化風險收益 率	12.69%	4.52%	2.8091
最大化收益	13.97%	6.77%	2.0636
最小化風險	12.38%	4.46%	2.7754

表 4.1 最佳化結果

從表中我們可以看出，最佳的方法是最大化風險收益率，也就是，在同等風險下，這種

組合的收益最大。相反的，在同等收益下，這種組合的可能風險是最小的。這種花同樣的錢獲得收益或風險收益率方法是現代投資組合理論中馬可維茨有效前沿理論的基石。意思就是如果我們限制投資組合整體的風險水平，並讓其隨時間逐漸增大，我們會得到幾個不同風險水平的最佳投資組合。因此，對風險的不同偏好會得到不同的最佳投資組合。

資產描述	年收益率	風險波動率	分配權重	最小分配額	最大分配額	風險收益率	收益評級 (高-低)	風險評級 (高-低)	風險收益評級 (高-低)	分配評級 (高-低)
資產1	10.54%	12.36%	11.09%	5.00%	35.00%	0.8524	9	2	7	4
資產2	11.25%	16.23%	6.86%	5.00%	35.00%	0.6929	7	8	10	10
資產3	11.84%	15.64%	7.78%	5.00%	35.00%	0.7570	6	7	9	9
資產4	10.64%	12.35%	11.23%	5.00%	35.00%	0.8615	8	1	5	3
資產5	13.25%	13.28%	12.09%	5.00%	35.00%	0.9977	5	4	2	2
資產6	14.21%	14.39%	11.04%	5.00%	35.00%	0.9875	3	6	3	5
資產7	15.53%	14.25%	12.30%	5.00%	35.00%	1.0898	1	5	1	1
資產8	14.95%	16.44%	8.90%	5.00%	35.00%	0.9094	2	9	4	7
資產9	14.16%	16.50%	8.37%	5.00%	35.00%	0.8584	4	10	6	8
資產10	10.06%	12.50%	10.35%	5.00%	35.00%	0.8045	10	3	8	6
組合總量	12.6919%	4.52%	100.00%							
風險收益率	2.8091									

圖 4.3 連續型最佳化結果

4.3 離散整數型最佳化

有時候，決策變數是離散型（如 0 和 1）而不是連續的。這意味著我們可以使用像轉換開關或執行/不執行來決策那樣的最佳化。圖 4.4 是一個包含 20 個專案的專案選擇模型。本例使用的是**開始|程式|Real Option Valuation|Risk Simulator**示例中的**離散最佳化**文件。跟以往一樣，每個專案都有自己的收益（ENPV 擴展淨現值和 NPV 淨現值——ENPV 就是 NPV 與任何戰略實物期權值之和），運作成本和風險等。如果需要的話，還可以對模型進行修改，加入所需的全工時評量法（FTE），其他各種函數資源，以及針對這些附加資源的附加約束。本模型中的輸入量基本都是從其他表中的模型鏈結而來的。例如，每個專案的投資模型都會有自己的折現現金流或收益大小。此處是應用在一定的預算下最大化投資組合的夏普指數(Sharpe-Ratio)，或是最小化風險，或是選擇的專案總數不超過 10 個的情況下增加額外的約束等等。所有這些情況都可以利用這個現有的模型。

步驟：

- 打開案例文件，通過點擊**模擬|新建模擬**生成一個新文檔，並命名。
- 最佳化的第一步就是設置決策變數。選定單格 J4 開始設置第一個決策變數（**模擬|最佳化|設置決策變數**），點擊連接圖示選擇命名單格（B4），選擇**二元變數**。然後，使用 Risk Simulator 的複製功能，複製單格 J4 的決策變數，並將其粘貼到剩下的單格 J5 至 J23 中。在只存在幾個決策變數的情況下，這是最佳的方法，可以為每個變數提供一個特殊的名字以方便以後的查找。
- 最佳化的第二步是設置約束條件。這裏有兩個約束條件：資產組合的總預算要小於\$5000，專案的總數不能超過 6 個。點擊**模擬|最佳化|約束條件**…選擇**增加**，添加一個新約束。然後選中單格 D17，輸入小於等於（<=）5000。通過 J17<=6 來重復設置。
- 最佳化的最後一步是設置目標函數，通過選擇目標單格 C19（或 C17），**模擬|最佳化|運行最佳化**，然後選擇最佳化類型（靜態最佳化，動態最佳化或者隨機最佳化），最後就可以開始模擬了。開始可以選

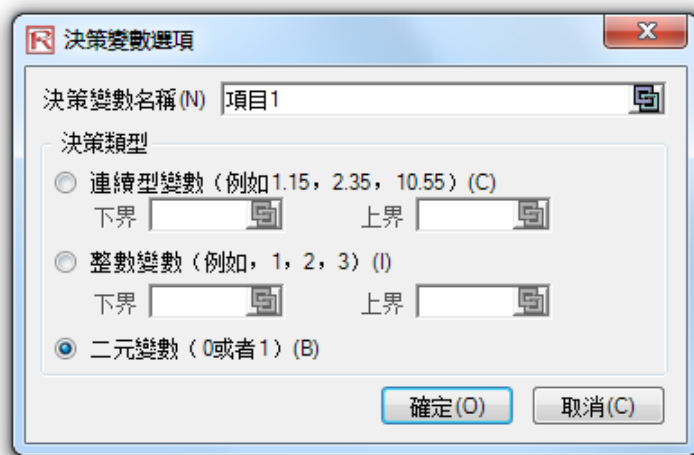
用靜態最佳化。再檢查確保目標單格中是夏普指數(Sharpe-Ratio)或風險收益率，然後選擇**最大化**。如果需要的話還可以回顧一下決策變數和約束條件，或是直接點擊**確定**運行靜態最佳化。

圖 4.5 是以上這些步驟的圖示。您可以在模型的 ENPV 和風險 (C 列和 F 列) 處增加模擬假設，然後練習使用動態最佳化和隨機最佳化。

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

擴展淨現值是備選項目的預期淨現值，然而可以使管理和需要持有來管理總備選項目的總成本，風險是備選專案擴展淨現值的變異係數。

圖 4.4 離散最佳化模型



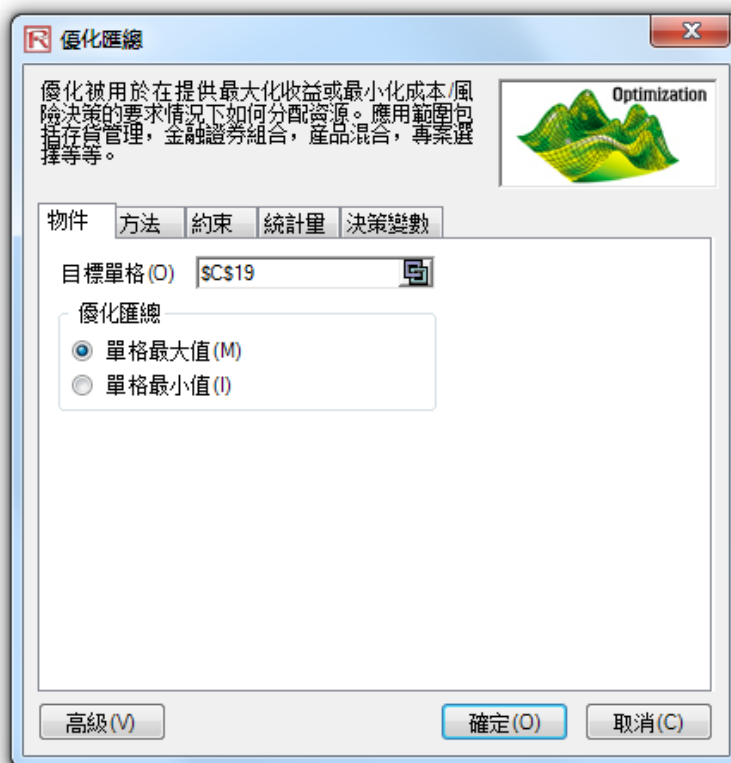
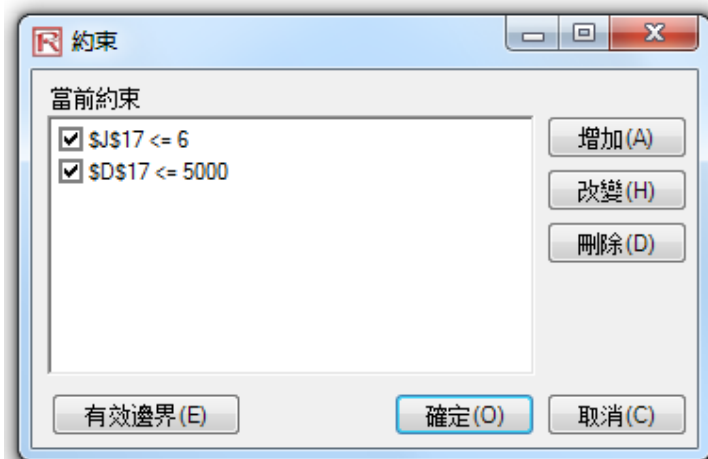


圖 4.5 在 Risk Simulator 中運行離散型最佳化

結果解析：

圖 4.6 中顯示了通過最大化夏普指數（Sharpe-Ratio）的方法來選擇最佳專案的一個示例。您也可以選擇最大化總體收益，並簡單地從清單中不斷選擇最高收益的專案直到資金用完或是超出預算約束，這將是一個非常繁瑣的過程。但是這樣可能會選到一些理論上不歡迎的專案，因為收益越高，風險越大。使用 Risk Simulator 將會使這一切變得非常方便。現在如果需要的話，您可以通過在輸入假設中加入 ENPV 和風險值，然後利用隨機或動態最佳化來

重復最佳化過程。

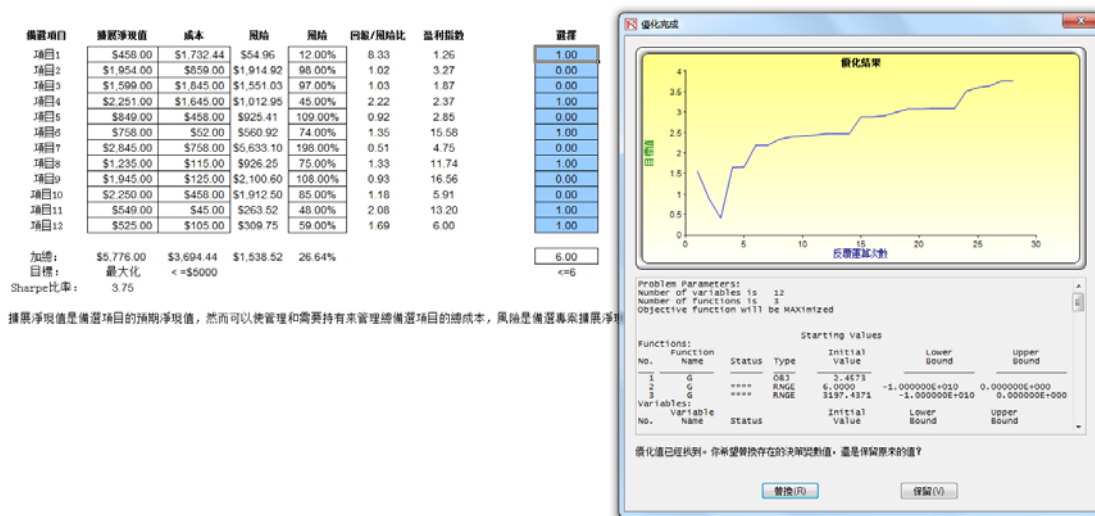


圖 4.6 最大化夏普指數（Sharpe-Ratio）條件下的最佳專案選擇

如果想要瞭解多關於最佳化的例子，可以參考《實物期權分析：工具和方法》，第二版一書的第十一章綜合風險分析，（Wiley Finance, 2005）。這個案例學習解釋了如何得到一個有效前沿，以及如何將預測，模擬，最佳化和實物期權融入到分析過程之中。

4.4 有效邊際和高級最佳化選項

圖 4.7 的第二個圖表顯示了最佳化的限制條件。這裏，如果在設定某些限制條件之後，點擊了**有效邊際**按鈕，現在可以改變這些限制因素。即，每個限制因素可以在最大值和最小值之間。例如，限制因素單格 J17 <= 6 可以設定在 4 和 8 之間運行最佳化（圖 4.7）。即將運行五組最佳化，每組都有以下的限制條件：J17 <= 4, J17 <= 5, J17 <= 6, J17 <= 7 和 J17 <= 8。最優的結果將繪製成有效邊際然後創建報告（圖 4.8）。需要指出的是，下面顯示了創建變動限制的步驟：

- 在一個最佳化模型中（例如，包含目標值，決策變數，和設定好的限制條件）點擊 *Risk Simulator* | **最佳化** | **限制** 然後點擊**有效邊際**。
- 選擇想要變動的**限制條件**（例如，J17）然後輸入參數作為**最大值**，**最小值**和**最佳化步數**（圖 4.7）然後點擊**添加**然後**確定**。
- 運行最佳化（*Risk Simulator* | **最佳化** | 運行最佳化）。可以選擇**靜態**，**動態**或者**隨機**。
- 結果將在用戶手冊中顯示（圖 4.8）。點擊**創建報告**生成報告工作簿包含所有的詳細資訊和最佳化結果。

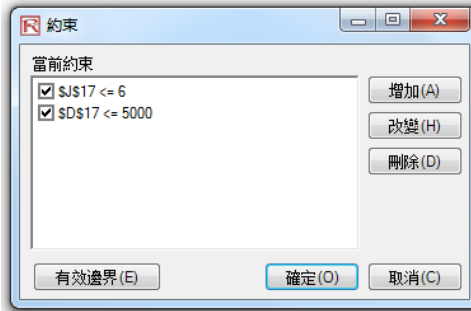


圖 4.7

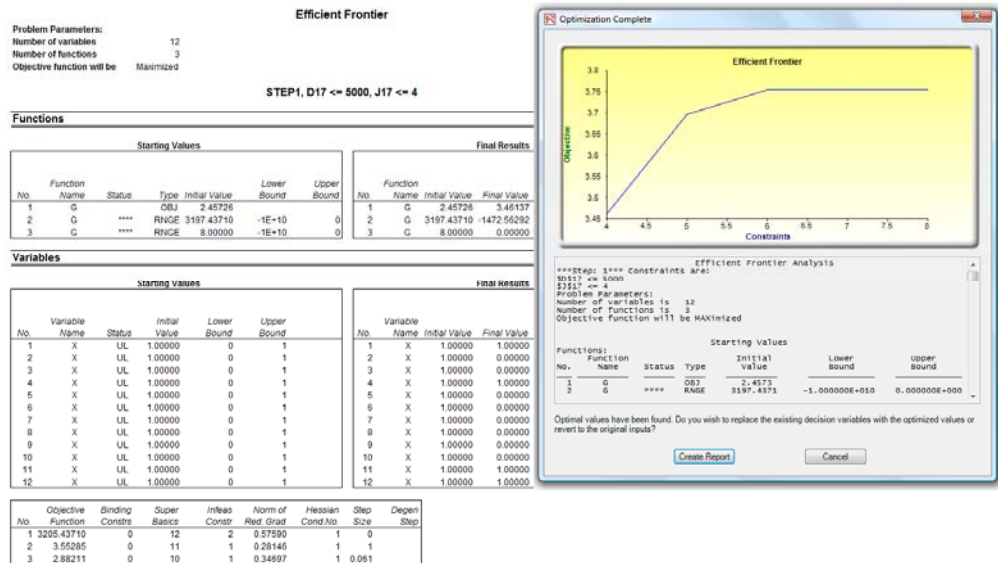


圖 4.8



5 · 風險模擬分析工具

本章涉及的是 Risk Simulator 的分析工具。我們會通過逐步分析的方法來討論 Risk Simulator 軟體中的一些案例應用，以確定分析工具的作用。這些工具對於風險分析領域的分析者來說是非常有價值的。本章將會詳細介紹每種工具的應用。

5.1 模擬中的颶風和敏感性分析工具

理論：

模擬中的重要工具之一就是颶風圖分析——它描述的是每個變數對模型結果的靜態作用。這個工具會自動擾動模型中每個變數的預設值，記錄模型預測或最終結果的波動，並按照擾動重要性的高低對結果進行排序。圖 5.1 至圖 5.6 是颶風分析應用的圖示。例如，圖 5.1 是一個折現現金流模型，它的輸入假設如圖所示。問題是哪一種關鍵因素對模型輸出的影響最大？也就是，什麼控制著\$93.63 的淨現值或哪種輸入變數對這個值的影響最大？

通過依次點擊**模擬工具**→**颶風分析**就可以使用颶風圖工具了。繼續第一個例子，打開案例文件夾中的颶風和敏感性圖（線性）文件。如圖 5.2 中所示的模型，單格 G6 中的淨現值是需要被分析的目標結果。模型中目標單格的引用部分被用於生成颶風圖。引用變數是所有影響模型結果的輸入量和中間變數。例如，如果模型中包含 $A=B+C$ ，並且 $C=D+E$ ，那麼 B，D，E 就是 A 的引用變數（C 是中間計算值，不是引用變數）。圖 5.2 中還顯示了用於估計目標結果的每個引用變數的檢驗範圍。如果引用變數僅僅是輸入量，那麼檢驗範圍將會是基於所選擇範圍（例如，誤差範圍在±10%之間）的一個簡單擾動。每個引用變數都可以根據需要在不同的百分比範圍內擾動。大範圍比較重要，因為相對於圍繞期望值的小幅度擾動來說，它可以涉及極值的檢驗。在某些情況下，極值可能會造成比較大，比較小，或是不穩定的影響（例如，對於一個較大或較小的變數值，當它們的規模經濟效應和範圍經濟效應增加或減少時就容易發生非線性事件），只有在大範圍情況下才能反映出這種非線性影響。

Discounted Cash Flow 模型

基年	2011	總現值收益	\$1,896.63
貼現率	15.00%	總現值投入	\$1,800.00
風險中性概率	5.00%	淨現值	\$96.63
銷售增長額	2.00%	內部收益率	18.80%
價格侵蝕	5.00%	投資回報	5.37%
稅率	40.00%		

	2011	2012	2013	2014	2015
產品A的價格	\$10.00	\$9.50	\$9.03	\$8.57	\$8.15
產品B的價格	\$12.25	\$11.64	\$11.06	\$10.50	\$9.98
產品C的價格	\$15.15	\$14.39	\$13.67	\$12.99	\$12.34
產品D的價格	50.00	51.00	52.02	53.06	54.12
產品E的價格	35.00	35.70	36.41	37.14	37.89
產品F的價格	20.00	20.40	20.81	21.22	21.65
總利潤	\$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
總利潤	\$1,046.99	\$1,014.53	\$983.08	\$952.60	\$923.07
運營成本	\$157.50	\$160.65	\$163.86	\$167.14	\$170.48
管理及辦公室費用	\$15.75	\$16.07	\$16.39	\$16.71	\$17.05
營運收入 (EBITDA)	\$873.74	\$837.82	\$802.83	\$768.75	\$735.54
折舊	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00
攤銷	\$3.00	\$3.00	\$3.00	\$3.00	\$3.00
EBIT	\$860.74	\$824.82	\$789.83	\$755.75	\$722.54
利息費用	\$2.00	\$2.00	\$2.00	\$2.00	\$2.00
EBT	\$858.74	\$822.82	\$787.83	\$753.75	\$720.54
稅金	\$343.50	\$329.13	\$315.13	\$301.50	\$288.22
淨收入	\$515.24	\$493.69	\$472.70	\$452.25	\$432.33
成本減值	\$13.00	\$13.00	\$13.00	\$13.00	\$13.00
淨營運資本的改變量	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00
資本支出	\$0.00	\$0.00	\$0.00	\$0.00	\$0.00
自由流動金	\$528.24	\$506.69	\$485.70	\$465.25	\$445.33
投資	\$1,800.00				

財務分析					
自由現金流的現值	\$528.24	\$440.60	\$367.26	\$305.91	\$254.62
投資應付的現值	\$1,800.00	\$0.00	\$0.00	\$0.00	\$0.00
淨現金流	(\$1,271.76)	\$506.69	\$485.70	\$465.25	\$445.33

表 5.1：樣本模型

步驟：

- 在 Excel 模型中選擇一個單一的輸出單元（例如，一個有函數或方程的單格）（本例中的單格 G6）
- 選擇模擬工具|颶風圖分析
- 檢查一下引用變數然後重新加以適當命名（使用更加簡短的名稱以便得到颶風圖和蛛網圖好的視覺效果），點擊**確定**



圖 5.2——運行颶風分析

結果解析:

圖 5.3 是颶風分析的結果報告，圖中顯示資本投資對淨現值的影響最大，其次是稅率，生產線所需的平均銷售價格和數量等等。整個報告包含四個基本因素：

- 統計概述列出了運行的步驟。
- 敏感性表（圖 5.4）中顯示開始的 NPV 基準值為 96.63，以及每個輸入量的變化（例如，投資以 10% 的升幅從 1800 美元增加到 1980，以 10% 的降幅從 1800 降到 1620）。結果最低值和最高值分別為 -83.37 和 276.63，變化總量為 360，是對 NPV 影響最大的變數。引用變數是按照影響力大小來排列的。
- 蛛網圖（圖 5.5）用圖來解釋了這些影響。Y 軸代表淨現值的目標值，X 軸代表每個引用變數變動的百分比（中心點是位於 96.63 的基準值，對於每個引用變數的偏離為 0 個百分點）。一條正斜率的直線意味著正相關的關係，負斜率的直線意味著負相關的關係（例如，投資圖是負斜率的意味著投資水平越高，淨現值越低）。斜率的絕對值代表影響的力度（一條比較陡的直線意味著對給定 X 軸上引用變數一定百分比的變化，Y 軸上淨現值的變化較大）。

- 颶風圖從另外一種圖形角度進行了解釋，影響最大的引用變數被置於頂部。X 軸代表淨現值，中心值為圖形的基準情況。圖中的綠色條代表正的影響，紅色條代表負的影響。因此對於投資來說右邊的紅色條意味著投資對高淨現值的負作用——換句話說，資本投資和淨現值是負相關的。反之產品 A 到 C 的價格和數量的作用亦然（圖右邊的綠色條）。

颶風圖和蛛網圖

統計匯總

仿真的有效工具之一是颶風圖——它分析了每個變數對模型結果的靜態影響。也就是說，這個工具會讓模型中的每個引用變數按照用戶自定義的值得進行擾動，進而記錄模型預測結果或最終結果的波動。最後按照重要性次序對擾動結果進行排序。引用變數是有可能對模型結果造成影響的輸入變數和中間變數。例如，如果模型中包含 $A=B+C$ ，並且 $C=D+E$ ，那麼 B、D、E 就是 A 的引用變數（C 是中間計算值，不是引用變數）。變數的範圍和數量是用戶自定義的，可以用於極值的檢驗，而不僅僅是圍繞期望值的小範圍的擾動。在某些情況下，極值可能會造成比較大，比較小，或是不穩定的影響（例如，對於一個較大或較小的變數值，當它們的規模經濟效應和範圍經濟效應增加或減少時就容易發生非線性事件），只有在大範圍情況下才能反映出這種非線性影響。

颶風圖中列出了所有影響模型的輸入因素，從對結果影響最大的輸入量開始排列。通過將每個引用變數每次在一定範圍內（例如，基準的 $\pm 10\%$ 之間）擾動，再將得到的結果與基準進行比較就可以得到颶風圖了。蛛網圖的形狀看起來像一隻蜘蛛的中間身體部分和它產生的線。正斜率意味著正相關關係，負斜率意味著負相關關係。同時，蛛網圖還可以詳細為我們顯示線性和非線性關係。颶風圖和蛛網圖有助於找出輸出單格的關鍵影響因素以便發現用於仿真的輸入量，識別出的那些關鍵的不確定的變數就是需要被仿真的變數，不要浪費時間去仿真那些確定的或是對結果影響很小的變數。

結果

引用變數名稱	基準值: 96.8261638553219			輸入量變化		
	輸出下限	輸出上限	有效範圍	輸入下限	輸入上限	基準值
C36: 投資	276.62616	-83.373836	360.00	\$1,620.00	\$1,980.00	\$1,800.00
C9: 稅率	219.72693	-26.474599	246.20	36.00%	44.00%	40.00%
C12: 產品A的價格	3.4255424	189.82679	186.40	\$9.00	\$11.00	\$10.00
C13: 產品B的價格	16.706631	176.5457	159.84	\$11.03	\$13.48	\$12.25
C15: 產品D的價格	23.177498	170.07483	146.90	45.00	55.00	50.00
C16: 產品E的價格	30.533	162.71933	132.19	31.50	38.50	35.00
C14: 產品C的價格	40.146587	153.10574	112.96	\$13.64	\$16.67	\$15.15
C17: 產品F的價格	48.047369	145.20496	97.16	18.00	22.00	20.00
C5: 折現率	138.23913	57.029841	81.21	13.50%	16.50%	15.00%
C7: 價格侵蝕	116.80381	76.640952	40.16	4.50%	5.50%	5.00%
C8: 銷售增長額	90.588354	102.68541	12.10	1.80%	2.20%	2.00%
C24: 折舊	95.084173	98.168155	3.08	\$9.00	\$11.00	\$10.00
C25: 攤銷	96.163566	97.088761	0.93	\$2.70	\$3.30	\$3.00
C27: 利息費用	97.088761	96.163566	0.93	\$1.80	\$2.20	\$2.00

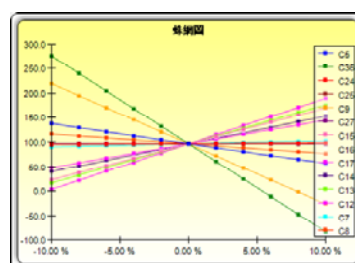


圖 5.3——颶風圖分析報告

注意：

記住颶風分析是一種靜態的敏感性分析，應用於模型中的每一個輸入變數——也就是說，每個變數都被單個擾動，其影響結果被製成表格。在運行模擬之前運行颶風分析是一個關鍵的步驟。風險分析中非常重要的第一步就是確定哪些是模型中對結果最重要的影響因素。下一步就是找出這些關鍵影響因素中哪些是不確定的。這些不確定的影響因素是專案成功的關鍵因素，模型的結果就取決於這些因素。這些變數是需要被模擬的。不要浪費時間去模擬那些確定的或是對結果的影響很小的變數。颶風分析可以很快很容易的

幫助我們找出那些關鍵影響因素。在假設投資和有效稅率已知並保持不變的前提下，本例中需要模擬的變數應該是價格和數量。

引用變數單格	基準值: 96.6261638553219			輸入量變化		
	輸出下限	輸出上限	有效範圍	輸入下限	輸入上限	基準值
C36: 投資	276.62616	-83.373836	360.00	\$1,620.00	\$1,980.00	\$1,800.00
C9: 稅率	219.72693	-26.474599	246.20	36.00%	44.00%	40.00%
C12: 產品A的價格	3.4255424	189.82679	186.40	\$9.00	\$11.00	\$10.00
C13: 產品B的價格	16.706631	176.5457	159.84	\$11.03	\$13.48	\$12.25
C15: 產品D的價格	23.177498	170.07483	146.90	45.00	55.00	50.00
C16: 產品E的價格	30.533	162.71933	132.19	31.50	38.50	35.00
C14: 產品C的價格	40.146587	153.10574	112.96	\$13.64	\$16.67	\$15.15
C17: 產品F的價格	48.047369	145.20496	97.16	18.00	22.00	20.00
C5: 貼現率	138.23913	57.029841	81.21	13.50%	16.50%	15.00%
C8: 價格侵蝕	116.80381	76.640952	40.16	4.50%	5.50%	5.00%
C7: 銷售增長額	90.588354	102.68541	12.10	1.80%	2.20%	2.00%
C24: 折舊	95.084173	98.168155	3.08	\$9.00	\$11.00	\$10.00
C25: 攤銷	96.163566	97.088761	0.93	\$2.70	\$3.30	\$3.00
C27: 利息費用	97.088761	96.163566	0.93	\$1.80	\$2.20	\$2.00

圖 5.4——敏感性表格

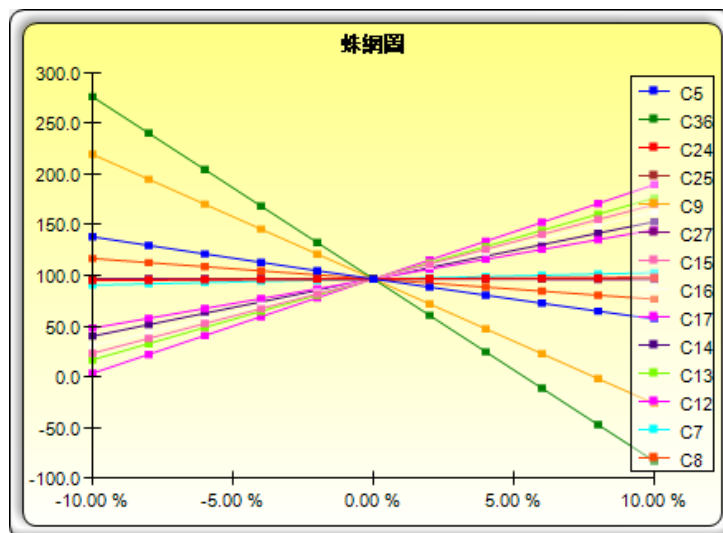


圖 5.5——蛛網圖

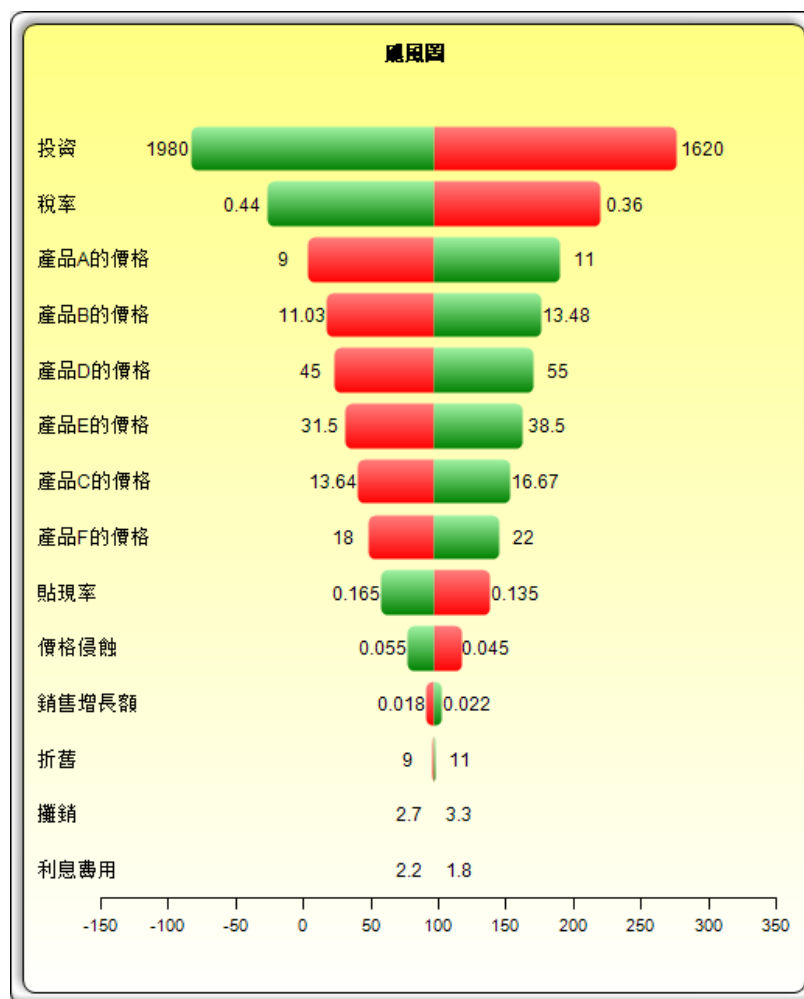


圖 5.6——颶風圖

儘管颶風圖很容易理解，但是蛛網圖有一個重要的優勢就是確定模型是否存在非線性關係。例如，圖 5.7 是蛛網圖的另一個例子，其中有很明顯的非線性關係（表中的線條不是直線而是曲線）。使用的示例文件是*颶風和敏感性表（非線性）*，它使用了 Black-Scholes 期權定價模型作為示範。颶風圖不能確定這些非線性關係，雖然它們可能是模型中的重要資訊或者可以讓決策者更深入地觀察模型的動態趨勢的因素。

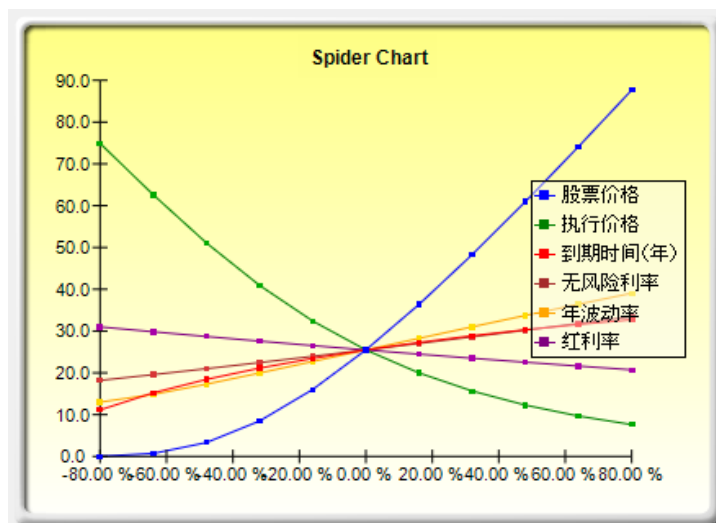


圖 5.7——非線性蛛網圖

更多關於颶風圖的注釋：

圖 5.2 顯示了颶風圖分析工具的用戶介面。注意在 Risk Simulator v4 或者更高的版本開始會有很多的增強功能。下面是關於運行颶風圖分析和有關增強功能的貼士。

- 颶風圖分析不應該僅僅運行一次。這就意味著，作為模型分析工具理想狀態下應該運行多次。例如，在一個很大的模型中，颶風圖分析可以使用默認選項運行第一次，然後顯示所有的引用單元（選擇**顯示所有的變數**）。這樣就會出現一個很大的報告和很長的颶風圖（不是很利於分析和顯示）。儘管如此，它卻為選擇多少個引用單元提作為關鍵因素供了一個參考（例如，颶風圖可以顯示 5 個具有重大影響的變數，而剩下的 200 個變數的影響則可能很小），這樣颶風圖運行第二次顯示更少的變數（例如，選擇**顯示前 10 個變數**如果前 5 個是關鍵的，這樣就創建了一個更為美觀的報告和颶風圖，顯示了關鍵因素和次關鍵因素之間的比較，即顯示不包含次關鍵因素的颶風圖）。最後，默認設置點可以增加±10%到某個大一些的值測試用於檢驗非線性（蜘蛛圖可以顯示非線性的線，而如果引用變數的影響是非線性的，颶風圖將偏向一邊）。
- **使用單格位址**對於一個較大的模型是一個很好的辦法，可以讓你確定引用變數的位置（工作簿名稱和單格地址）。如果未選該選項，軟體應用自身的模糊邏輯找到每個引用變數的名稱（有時候名稱可能會發生衝突或者過長，影響颶風圖的美觀）。
- **分析當前工作簿和分析所有的工作簿**選項允許客戶控制引用變數是否是現有工作簿的一部分或者所有的工作表包含在同一個工作簿中。這個選項可以用在只分析當前表格中的輸入輸出變數還是進行全員搜索所有的引用變數。

- **使用全員設置**當分析一個較大的模型的時候十分有用，可以檢驗所有的引用變數，比如 $\pm 50\%$ 而非默認的 10% 。不改變每個引用變數的大小，用戶可以選擇這個選項，改變一個設置然後點擊別的位置，然後整個引用變數的列表都會變化。未選中這個選項允許用戶每次更改一個引用變數的測試點。
- **忽略零或者空值**是一個默認選項如果引用變數包含零值或者空值則不會在颶風圖分析中考慮。這是一個典型設置。
- **著重指出可能的整數值**選項可以快速確定所有包含整數的可能引用變數單格。這個選項有時候十分重要（例如，函數 IF 在單格顯示中可能會出現 1，或者類似的整數 1, 2, 3，這些值可能不需要進行檢驗）。例如， $\pm 10\%$ 對於標記 1 的結果可能會是 0.9 或者 1.1，實際上這對於模型來說是不正確的，Excel 會顯示公式是錯誤的。這個選項一旦選擇，將著重指出颶風圖分析可能包含潛在問題的區域，可以手動選擇或者不選這個引用變數，或者可以使用**忽略可能的整數值**同時關閉他們。

5.2 敏感性分析

理論：

另一個相關的特徵就是敏感性分析。颶風分析（颶風圖和蛛網圖）是用在運行模擬之前的靜態擾動，而敏感性分析則是在運行模擬之後運行的動態擾動。颶風圖和蛛網圖是靜態擾動後的結果，每個引用變數或假設變數每次按照一個事先設置的值進行擾動，生成的擾動結果被製成表格。相反的，敏感性分析是動態擾動的結果，多個變數假設同時被擾動，它們在模型中的相互作用和變數之間的相關性都在結果的波動中反映出來。颶風圖用於識別對結果影響最大，最適合模擬的因素；敏感性分析則是確定多個變數在模型中同時被模擬時對結果的影響。圖 5.8 詳細的解釋了這一效果。注意到關鍵影響因素的排名和上例中颶風圖裏的很類似。但是，如果再加上變數之間的相關性，圖 5.9 中出現完全不同的情況。比如我們注意到價格侵蝕對淨現值的影響很小，但是如果某些輸入量之間相關的話，它們之間的相互作用可能會增大價格侵蝕的影響力。

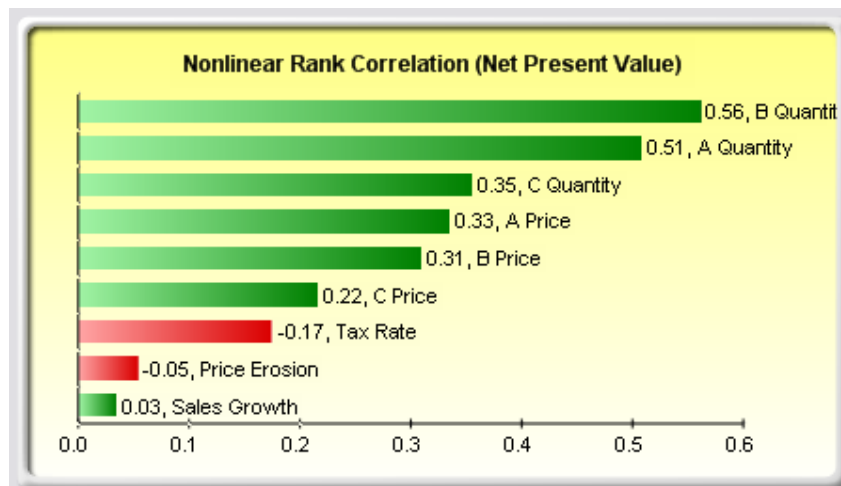


圖 5.8——不存在相關性的敏感性圖

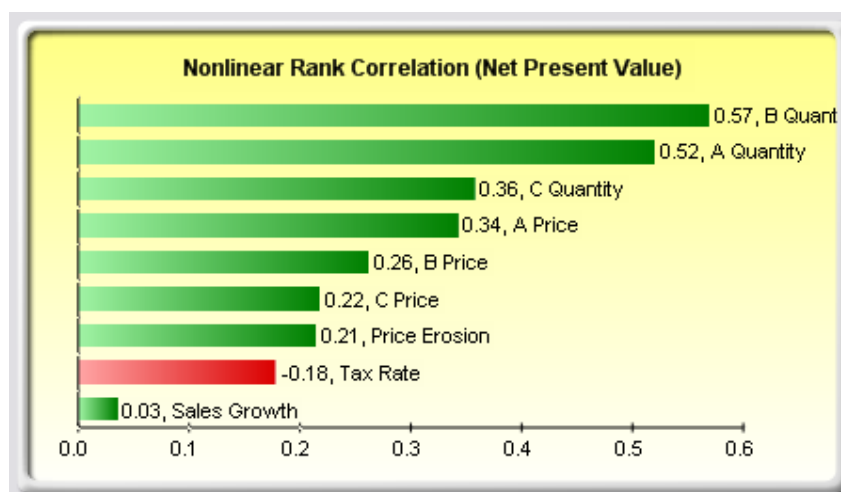


圖 5.9——存在相關性的敏感性圖

步驟：

- 打開或新建一個模型，定義輸入和預測，運行模擬（本例使用的是**颶風和敏感性圖（線性）**示例文件）
- 選擇模擬工具**敏感性分析**
- 選擇用於分析的預測點擊**確定**（圖 5.10）

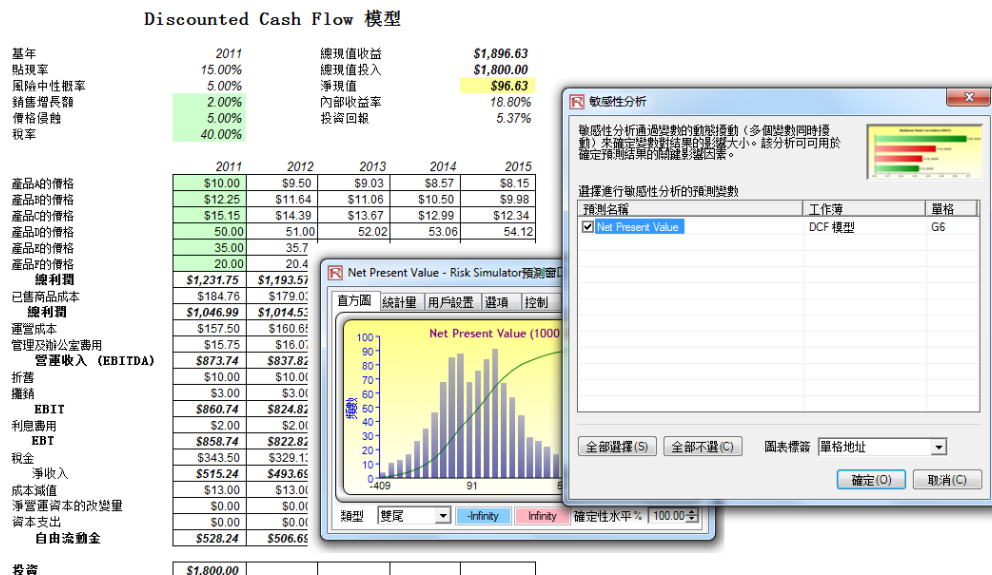


圖 5.10——運行敏感性分析

結果解析：

敏感性分析的結果包括一份報告和兩個關鍵圖。首先是非線性秩相關（圖 5.11），按照假設量-預測相關性由高到低來排列。這些相關關係是非線性和非參數的，這樣它們就不用遵守任何分佈要求（例如，一個服從韋伯分佈的假設量可以和另一個服從 β 分佈的變數相比較）。除了一點以外，從圖中得出的結果與之前颶風分析的結果很類似（當然不包括資本投資這個變數，因為我們假定它是已知的，所以不需要模擬）。另外，與颶風圖（圖 5.6）相比，稅率在敏感性分析圖（圖 5.11）中的位置相對低一些。這是因為如果僅是看稅率，它對結果影響很大，但是一旦模型中的其他變數有相互作用，稅率的影響就明顯降低（這是因為由於歷史稅率資料的波動很小，導致其分佈也較小，同時稅率是稅前收入的一定百分比值，其他引用變數對稅前收入有很大影響）。這個例子證明敏感性分析在運行模擬之後來確定模型中是否存在相互作用以及這些作用是否會持續具有重要作用。第二個圖（圖 5.12）說明了變異的百分比解釋。也就是說，給出預測的波動，並考慮到變數之間的相互作用之後，每個假設變數對變異的解釋程度如何？注意通常所有變異的解釋總和都會接近 100%（有時會有其他因素對模型產生影響，但是不能被直接觀察出），如果存在相關性，那麼總和有時可能會超過 100%（這取決於累積的相互作用）。

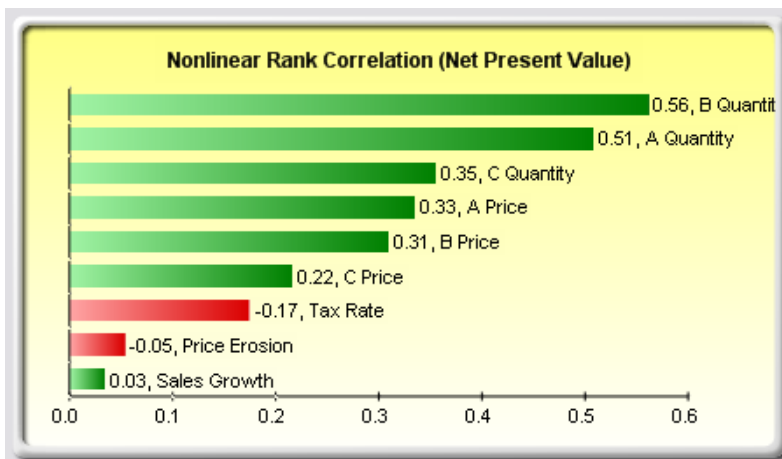


圖 5.11——相關性排列圖

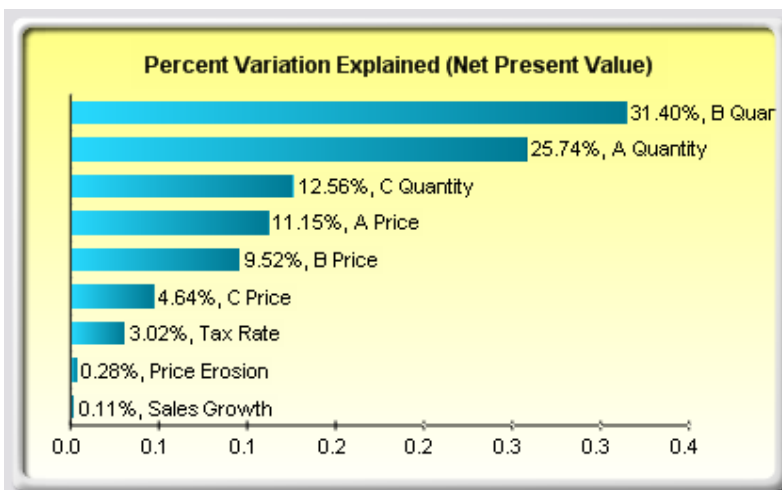


圖 5.12——方差貢獻圖

注意：

颶風分析是在運行模擬之前，而敏感性分析是在運行模擬之後。颶風分析中的蛛網圖可以處理非線性情況，敏感性分析中的相關性排列圖可以處理非線性和自由分佈情況。

5.3 分佈擬合：單變數和多元變數

理論：

另一個有效的模擬工具是分佈擬合。對於模型中的一個具體輸入變數分析者該使用哪種分佈呢？分佈的相關參數有哪些？如果變數不存在歷史資料，那麼分析者必須對未知變數作一些相關假定。其中一種方法就是利用 Delphi 法，也就是一組專家來估計每個變數的變化。例如，一組機械工程師要通過嚴格的試驗或推測來估計螺旋彈簧在極端情況下的直徑。這些值可被看作變數的輸入參數（例如，極值為 0.5 和 1.2 的均勻分佈）。當試驗不能進行時（如市場份額和收益增長率），管理層依然可以對可能的結果作一些估計以提供

最佳案例情景，最可能案例情景和最差案例情景。

然而，如果可以得到歷史資料，那我們就可以進行分佈擬合。假設這種歷史趨勢自身是不斷重復的，那麼就可以利用歷史資料來找到最佳擬合分佈及其相關參數，以便更好的定義用於模擬的變數。圖 5.13 到 5.15 是一個分佈擬合的例子。使用的是示例文件夾中的資料擬合模型。

步驟：

- 打開一個包含資料的工作簿
- 選擇想要擬合的資料（資料必須在同一列）
- 選擇模擬工具|分佈擬合（單變數）
- 選擇希望擬合的分佈類型或是接受預設值，選擇所有的分佈，點擊**確定**（圖 5.13）
- 查看擬合的結果，選擇想要的相關分佈點擊**確定**（圖 5.14）

正態	正態（多元）	均勻	二項
93.75	87.53	45.29	6.00
109.52	99.66	46.94	6.00
101.17	108.75	45.96	6.00
102.29	87.41	52.09	8.00
105.58			
99.55			
86.79			
105.20			
113.63			
105.90			
90.68			
96.20			
79.74			
91.49			
98.28			
97.70			
97.85			
93.73			
92.06			
85.51			
103.21			
87.45			
96.40			
92.41			
82.75			
103.65			
90.19			
112.42			
103.22			
91.56			
86.04			
115.40			
	95.27	53.87	6.00
	95.33	46.24	7.00
	102.26	54.80	6.00

圖 5.13——單變數分佈擬合

結果解析：

用於檢驗的零假設是：樣本的擬合分佈與總體的分佈是一致的。因此，如果計算出的 p 值小於臨界 α 水平（一般為 0.10 或 0.05），那麼這個擬合的分佈是不可信的。反之，p 值越大，分佈擬合的越合理。概略的，您可以將 p 值看作是解釋百分比，也就是說，如果 p 值為 0.9727（圖 5.14），那麼設置一個均值為 99.28，標準差為 10.17 的常態分佈可以解釋 97.27% 的資料變異，那麼說明這是一個很好的擬合。結果（圖 5.14）和報告（圖 5.15）中都有檢驗量，p 值，理論統計量（基於所選擇的分佈），經驗統計量（基於原始資料），原始資料（用於記錄使用的資料），以及假設變數的相關分佈參數（例如，如果選擇了自動生成假定選項並且已經存在模擬文件）。結果中還對所有選擇的分佈按照它們的擬合情況進行了排名。

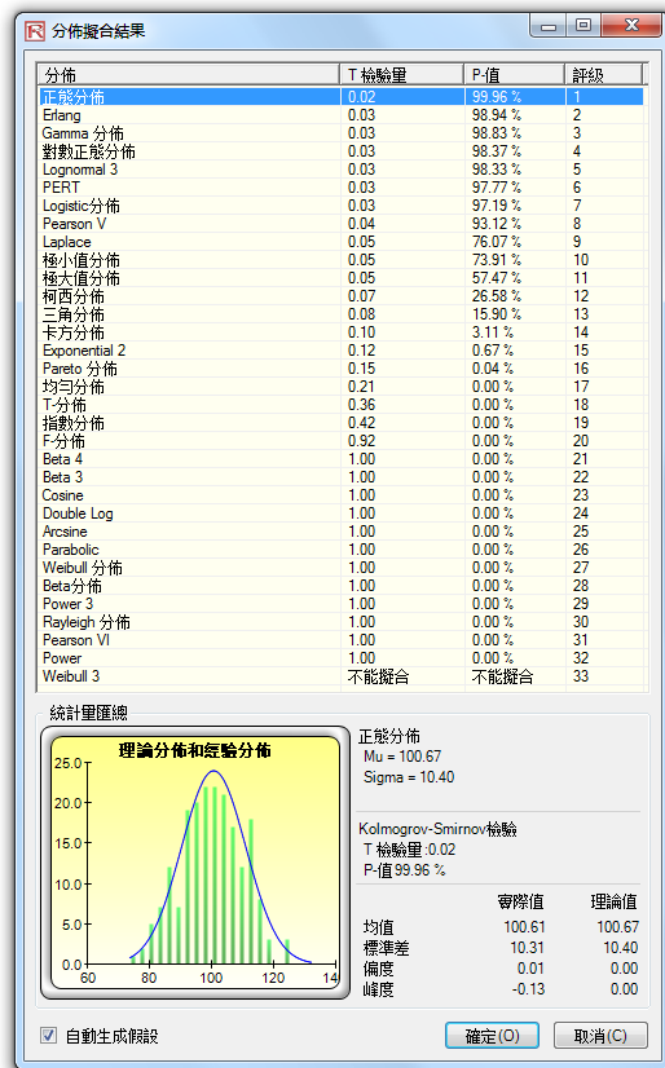


圖 5.14：分佈擬合結果

单变量分布拟合

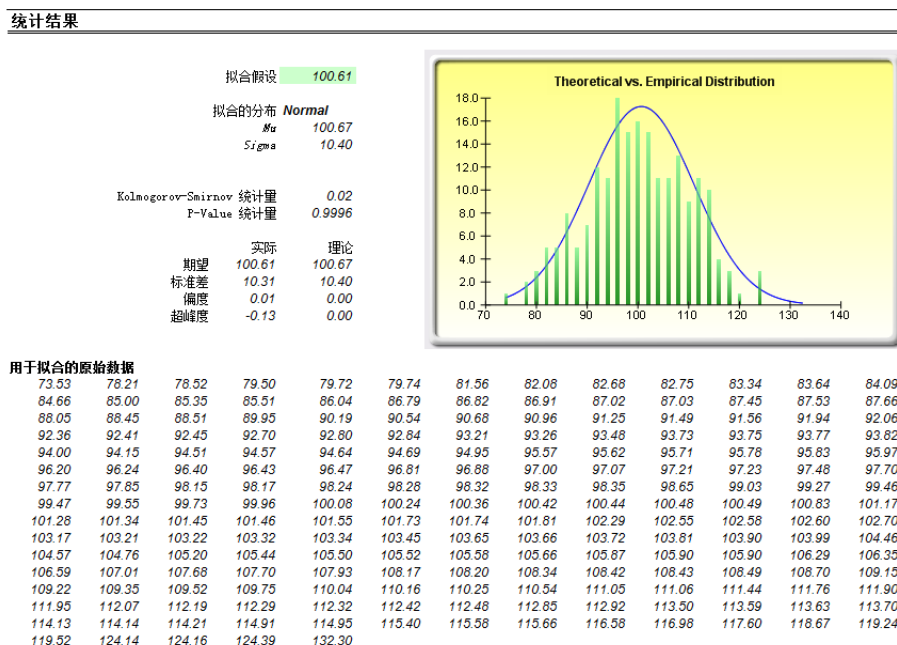


圖 5.15：分佈拟合報告

多元變數的拟合過程與單個變數的拟合非常類似。但是，資料必須按列排列（每個變數被分配在一列）並且所有變數一次只能拟合一個概率分佈。

步驟：

- 打開一個包含資料的工作簿
- 選擇您想要拟合的資料（資料必須是同一列的）
- 選擇模擬工具|分佈拟合（多元變數）
- 查看拟合的結果，選擇您想要的相關分佈點擊**確定**

注意：

注意到分佈拟合程式中所使用的統計排名方法是卡方核對總和柯爾莫諾夫斯米爾諾夫檢驗。前者用於檢驗離散分佈，後者用於檢驗連續分佈。簡單來說，具有內部最佳化規則的假設檢驗被用於尋找每種被檢驗分佈的最優拟合參數，並將結果由好到壞排列。

5.4 Bootstrap 模擬（拔靴法）

理論：

Bootstrap 模擬是用來估計統計預測量或其他樣本原始資料可靠性或精確度的一種簡單方法，一般來說 Bootstrap 模擬被用於假設檢驗中。過去傳統的方法都是依靠數學公式來描述樣本統計量的精確度。這些方法假定樣本統計量的分佈接近常態分佈，這樣統計量的標準誤差或置信區間的計算就相對容易

一些。但是，當統計樣本的分佈不屬於常態分佈或不容易發現時，這些傳統的方法就無法使用了。相反，Bootstrap 通過反復取樣並從每次取樣的不同樣本中創造分佈來對樣本統計進行經驗分析。

步驟：

- 運行模擬
- 選擇模擬工具|非參數 Bootstrap
- 選擇一個預測來進行 Bootstrap，選擇 Bootstrap 的統計量，輸入需要進行 Bootstrap 的次數點擊**確定**（下圖 5.16）

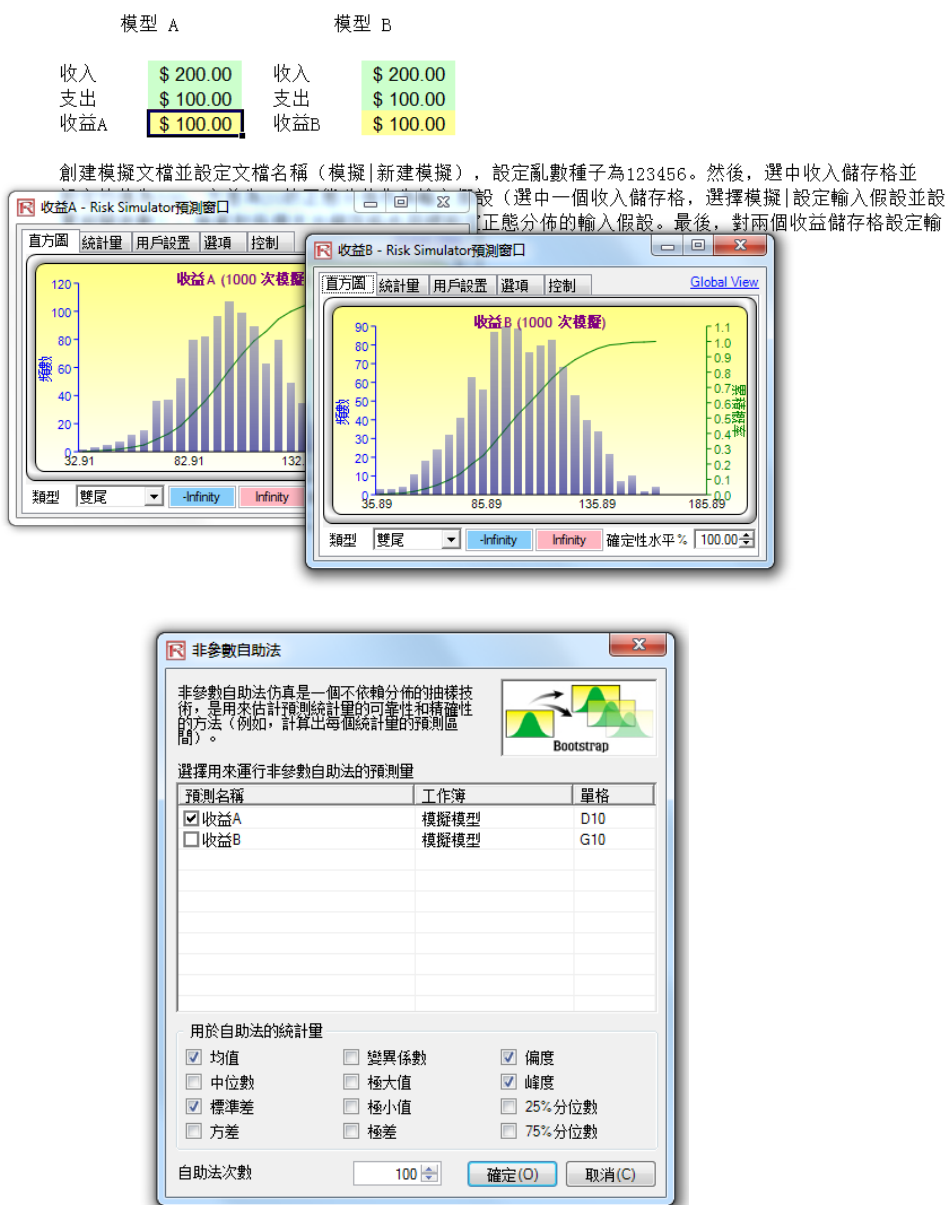


圖 5.16——非參數 Bootstrap

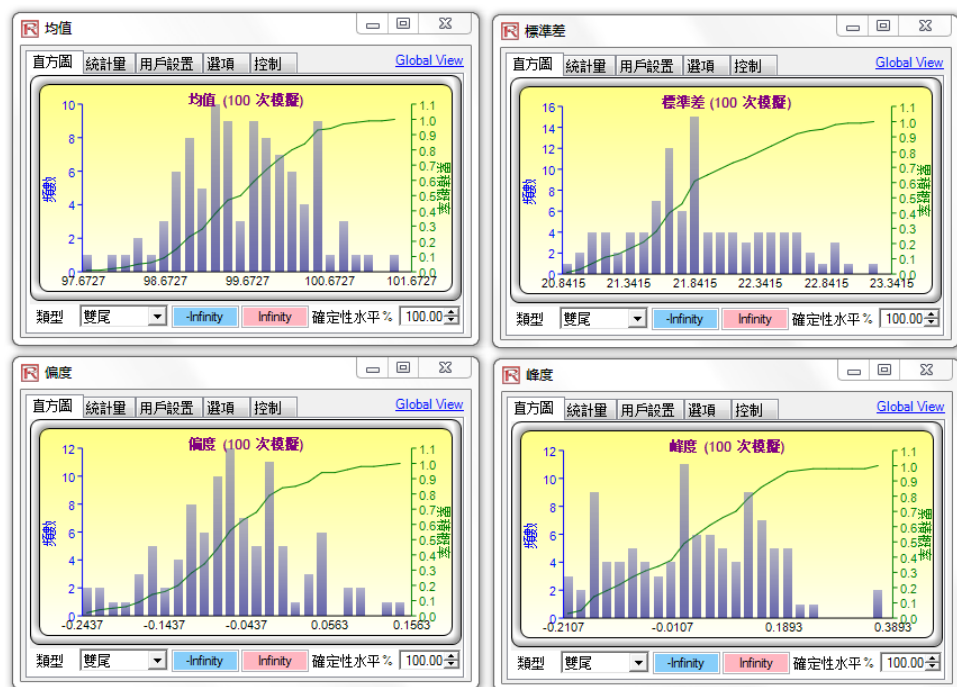


圖 5.17——Bootstrap 類比結果

結果解析:

實際上，非參數 Bootstrap 模擬可以被看作是基於模擬的模擬。因此，在運行一次模擬之後，結果的統計量會顯示出來，但是有時這些統計量的精確度和它們的統計顯著性卻讓人懷疑。例如，如果一次類比後得到的偏度值為 -0.10，那麼這個分佈真的是負偏的，還是因為隨機的影嚮造成的輕微負偏呢？-0.15，-0.20 等等呢？也就是說這個分佈被認為是負偏的程度是多少呢？其他的統計量也存在同樣的問題。如果計算出的統計量相同，那麼可以說一個分佈在統計上與另一個分佈是相同的，或是它們是顯著不同的？圖 5.17 是部分 Bootstrap 的結果。例如，置信度為 90% 的偏度統計值位於 -0.2233 和 0.1543 之間，0 值落在這個區間，意味著在 90% 的置信水平上，預測的偏度統計上不顯著區別於 0，或是這個分佈可以被認為是對稱的和非偏的。相反的，如果 0 值落在這個區間之外，那麼反面就成立，分佈是有偏的（如果預測值為正，就是正偏，如果預測值為負，就是負偏）。

注意:

Bootstrap 這個詞來自於一句俗語，“拎著鞋帶把自己提起來”，這種方法利用自身統計量的分佈來分析統計量的精確程度。非參數模擬就是簡單的從一個大籃子裏隨機取出高爾球然後放回，每個高爾夫球都基於一個歷史資料點。假設籃子裏一共有 365 個高爾夫球（代表 365 個歷史資料點）。想象一下將您每次隨機取出的球都記錄在一塊大黑板上。有放回取出的 365 個球的結果被記錄在黑板上的第一列，一共有 365 行。計算出這 365 行的相關統計量（例如均值、中值、標準差等），然後將這個過程重復 5000 次。現在黑板上有 365 行和 5000 列資料。所以我們會得到 5000 個被製成表格的統計量（有

5000 個均值、5000 個中值、5000 個標準差等等) 及它們的分佈，也計算出統計量的其他相關統計量，從這些結果中我們可以看出這些模擬統計量的置信度。換句話說，在一個 10000 次試驗的模擬中，得出預測結果的均值為 5.00 美元。那麼這個結果的可信度是多少呢？Bootstrap 法允許使用者計算均值的置信區間，統計量的分佈等等。由於根據統計學中的大數定理和中心極限定理，樣本均值的均值是無偏估計量，當樣本空間增大的時候，它趨近真實的總體均值，所以 Bootstrap 的結果是重要的

5.5 假設檢驗

理論：

假設檢驗就是通過檢驗兩個分佈的均值和方差來判斷這兩個分佈在統計上是否是一致的。也就是找出對均值和方差的不同預測是由於隨機的原因還是由於它們之間的統計顯著性差別的原因造成的。

步驟：

- 運行模擬
- 選擇模擬工具|假設檢驗
- 每次只選擇兩個預測量來進行檢驗，選擇您想要運行的假設檢驗類型，點擊**確定**（圖 5.18）

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

創建模擬文檔並設定文檔名稱（模擬|新建模擬），設定均值為200，方差為20的正態分佈作為輸入假設置相關參數）。接著對每個支出儲存格也同樣設定出預測並進行模擬。

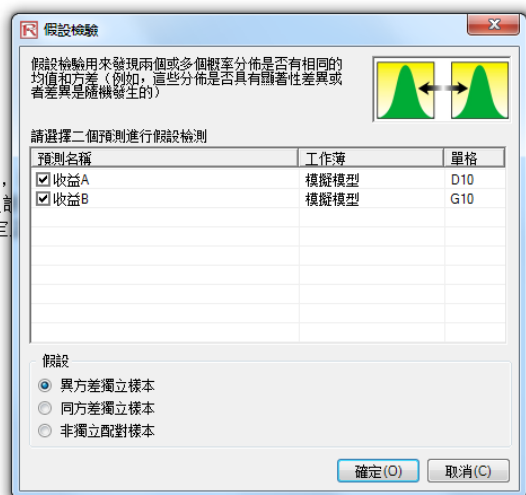


圖 5.18——假設檢驗

結果解析：

雙尾假設檢驗使用零假設（ H_0 ）：兩變數的總體均值是統計一致的。它的備擇假設是兩者的均值是不一致的。如果計算的 p 值小於或等於 0.01、0.05 或 0.10，這意味著我們要拋棄零假設，也就是說預測平均值在 1%、5% 和 10% 的顯著性水平是統計顯著不一致的。當 p 值較高時，我們就不能拋棄零假設，此時兩個預測的分佈就是統計一致的。再對兩個預測的方差使用 F 檢驗進行相同的分析。如果得到的 p 值很小，說明方差（標準差）統計不一致的，相反的，對於較大的 p 值，兩方差是統計一致的。

兩預測變數均值和方差的假設檢驗

統計匯總

假設檢驗就是通過檢驗兩個分佈的均值和方差來判斷這兩個分佈是否在統計上是一致的還是不同。也就是說，對均值和方差的不同預測是由於隨機的原因還是由於它們之間的統計顯著性差別的原因造成的。當預測的分佈來自於不同的樣本空間時（例如，在兩個不同的地點，在兩個不同的商業運作單元搜集的資料等等），使用不等方差的雙變數 t 檢驗（預測 1 的樣本空間方差與預測 2 的樣本空間方差是不一致的）。當預測的分佈來自兩個類似樣本空間時（從類似規格的兩個不同機械部件上搜集的資料等等），使用等方差的檢驗（預測 1 的樣本空間方差與預測 2 的樣本空間方差是一致的）。當預測的分佈來自同一個樣本空間時（在不同情況下從同一組客戶那裏取得的資料等）可以使用配對雙變數 t 檢驗。

雙尾假設檢驗使用的零假設（ H_0 ）是：兩變數的總體均值是統計一致的。備擇假設是兩者的均值是不一致的。如果計算的 p 值小於或等於 0.01、0.05 或 0.10，這意味著我們要拋棄零假設，也就是說預測平均值在 1%、5% 和 10% 的顯著水平是統計顯著不一致的。當 p 值較高時，我們就不能拋棄零假設，此時兩個預測的分佈就是統計一致的。再對兩個預測的方差使用 F 檢驗進行相同的分析。如果得到的 p 值很小，說明方差（標準差）統計不一致的，相反的，對於較大的 p 值，兩方差是統計一致的。

結果

假設檢驗假定	異方差
t 統計量	0.484334
t 統計量的 P 值	0.628202
F 統計量	1.067792
F 統計量的 P 值	0.300091

圖 5.19——假設檢驗結果

注意：

當預測的分佈是來自於不同的樣本空間時（例如，從兩個不同的地點，兩個不同的商業運作單元搜集的資料等等），使用不等方差的雙變數 t 檢驗（預測 1 的樣本空間方差與預測 2 的樣本空間方差是不一致的）。當預測的分佈來自兩個類似的樣本空間時（從類似規格的兩個不同機械部件上搜集的資料等等），使用等方差的 t 檢驗（預測 1 的樣本空間方差與預測 2 的樣本空間方差是一致的）。當預測的分佈來自同一個樣本空間時（在不同情況下從同一組客戶那裏取得的資料等）可以使用配對雙變數 t 檢驗。

5.6 資料輸出和保存模擬結果

使用 Risk Simulator 的資料提取功能可以很容易地提取出模擬的原始資料。假設和預測都可以被提取，但是首先要運行一次模擬。提取的資料可以用於其他的一些分析。

步驟：

- 打開或新建一個模型，定義輸入和預測，運行模擬
- 選擇模擬工具|資料提取
- 選擇想要的假設和預測提取資料，點擊**確定**

可以按照不同的形式來提取資料：

- 新工作簿中的原始資料，其中的類比值（假設和預測）可以被保存並用於進一步分析
- 保存為純文字檔案，這樣可以直接導入其他的分析軟體
- Risk Simulator 文件，可以在稍後通過選擇**模擬工具|打開資料|導入**來恢復結果資料（假設和預測）

第三個選項是最常用的選項，也就是說，將類比的結果保存為*.risksim 文件，這樣稍後

可以恢復資料，不必重新運行一次類比。圖 5.21 是提取、導出和保存模擬結果的對話方塊。

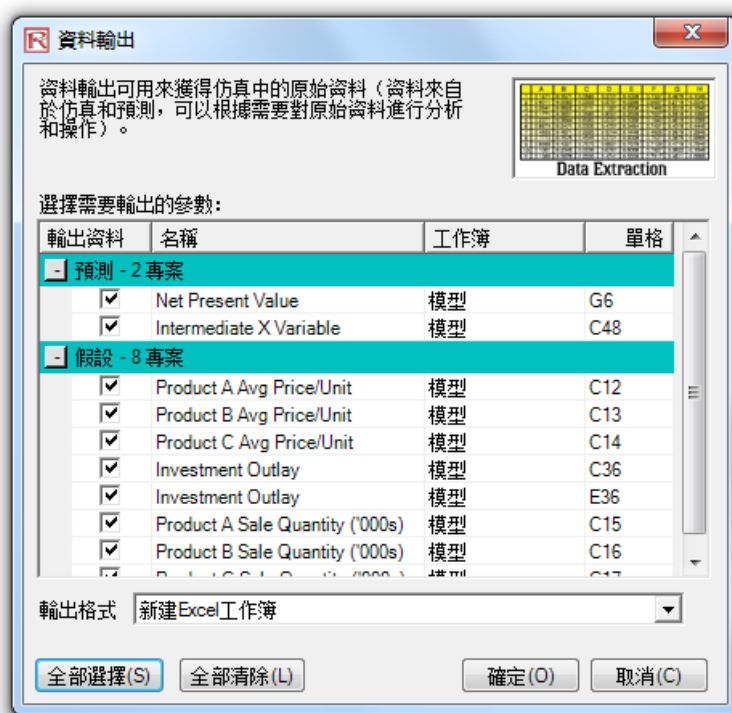


圖 5.21——樣本類比報告

5.7 創建報告

在運行模擬之後，您可以生成一份關於假設，預測和模擬結果的報告。

步驟：

- 打開或新建一個模型，定義假設和預測，然後運行模擬
 - 選擇模擬|生成報告

仿真 - DCF Model

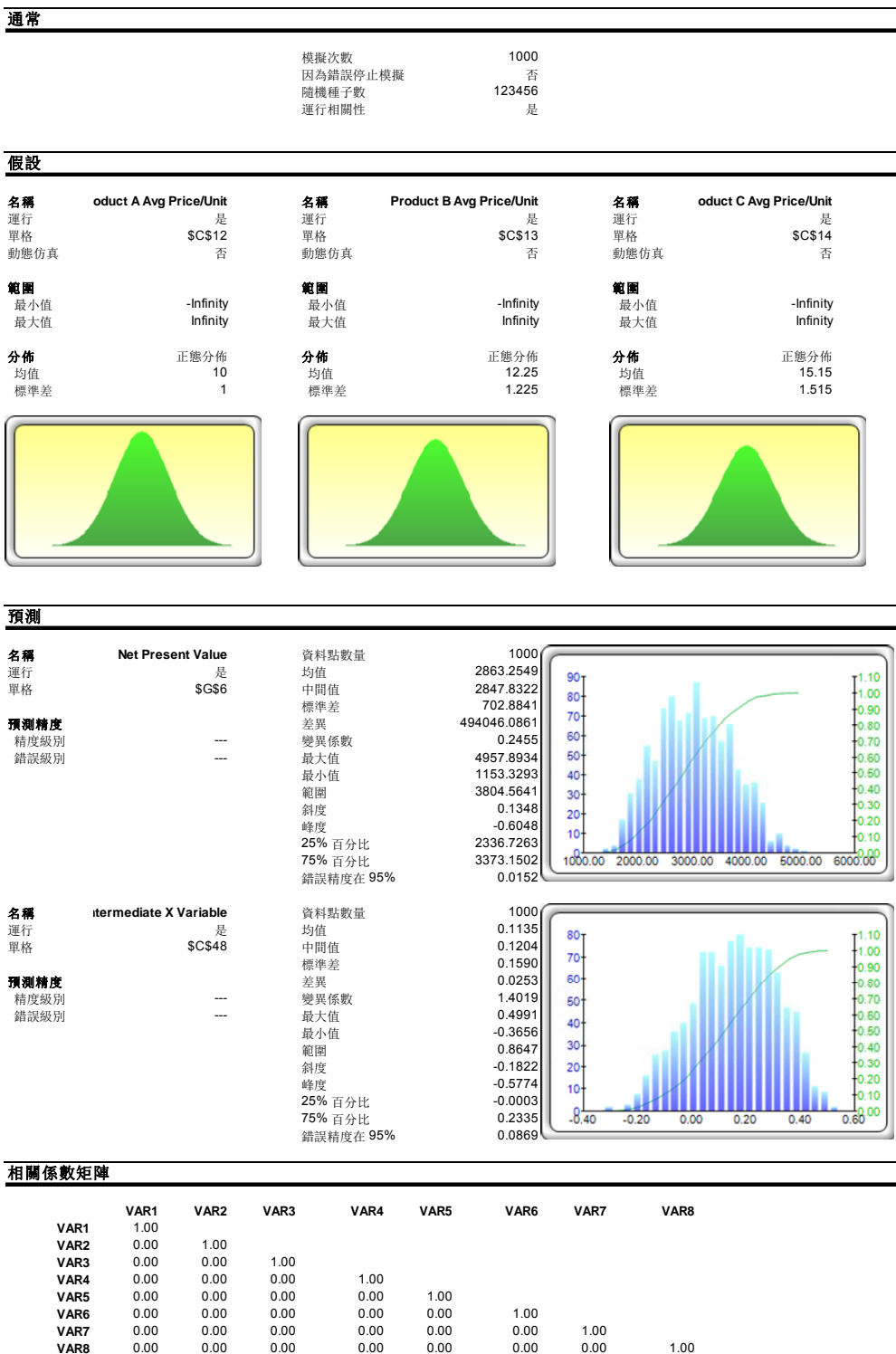


圖 5.21——樣本模擬報告

5.8 回歸和預測診斷工具

Risk Simulator 中的高級分析工具可以用來決定資料的計量經濟學特性。診斷工具包括測定資料的異方差性，非線性，異常性，規格誤差，微數缺測性，平穩和隨機性，誤差的正態性和球性，以及多重共線性。每個檢驗在各自模型的報告中都有詳細地描述。

操作過程描述：

- 打開示例模型 (Risk Simulator | 示例模型 | 回歸診斷) 點擊時間序列資料工作表然後選擇資料包括資料變數的名稱(單格 C5:H55)。
- 點擊 Risk Simulator | 工具 | 診斷工具。
- 點擊資料在下拉功能表中選擇因變數 Y。點擊**確定**完成選擇 (圖 5.22)。

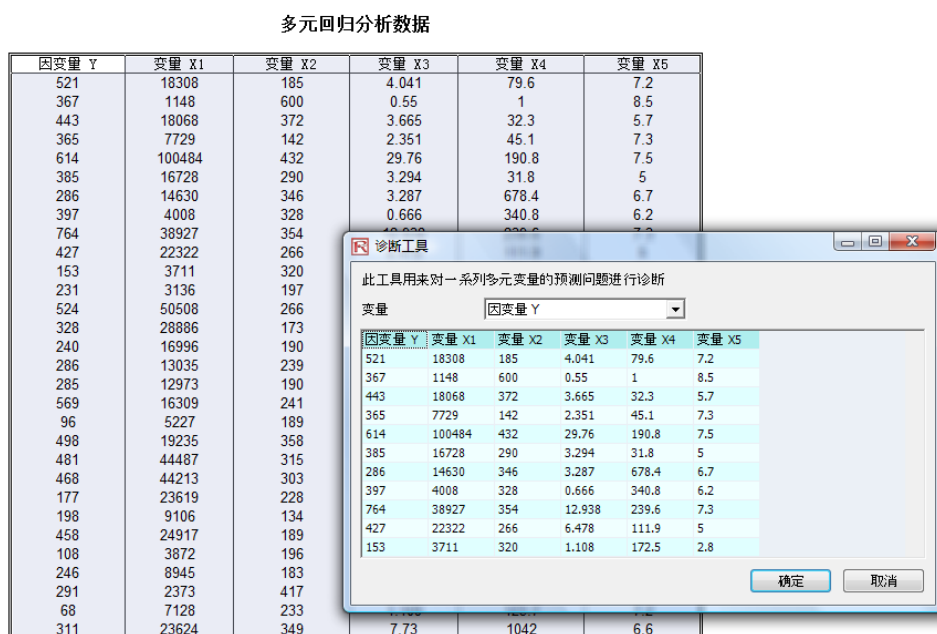


Figure 5.22 - 運行資料診斷工具

在預測和回歸分析中最常見的錯誤是異方差，也就是說，誤差的標準差隨著時間的增加不斷變大。(參看圖 5.23 使用診斷工具作為測試的結果)。視覺上來說，資料在豎直方向上的波動寬度隨著時間不斷增大或成扇形散開，並且明顯地，可決係數 (R 方) 當異方差存在時顯著下降。如果因變數的標準差不是一個常數，誤差的標準差也將不是一個常數。除非因變數的異方差性是顯著的，否則它的效果不會非常劇烈：最小二乘法估計仍然是無偏的，當誤差是常態分佈時，斜率和截距的估計將是常態分佈的。當誤差不是常態分佈時，斜率和截距的估計也將是漸進常態分佈的 (當資料點的個數很大時)。斜率方差和整體方差的估計量將是不精確的，但是如果引數的值是關於它們

的均值對稱，這種不精確性就可能不那麼重要了。

如果資料量很少（微數缺測性），那麼就很難判斷是否違背使用這些模型的假設。而且在存在違背模型假設的情況下，非正態性或方差的異方差性也是很難察覺的。尤其對線性回歸模型而言，資料點比較少時，很難保證不違背模型假設。同時，這時候往往很難決定是用直線擬合數據點效果好，還是使用非線性函數（曲線）效果好。即使所有假設檢驗都是符合的，小樣本的線性回歸也可能沒有足夠有效性來判斷斜率是否為 0。這種有效性與殘差項、引數的方差，假設檢驗的置信水平和數據點的個數有關。當殘差增大時，或置信水平減小時（比如，假設檢驗更嚴格），這種有效性減小。當引數的方差增加或資料點個數增加時，這種有效性增加。

由於異常的存在，資料可能分佈上並不一致。異常是資料中那些不正常的值。異常可能會對擬合的斜率和截距產生很強的影響，並且使得對大塊資料的擬合較差。異常的存在傾向於增大預測的殘差，減小拒絕零假設的概率。比如，產生更高的預測誤差。異常的產生可能由於記錄誤差造成的（可糾正的），也可能由於因變數的值並不是全都從同一個分佈中取樣的。顯然地，異常也可能是由於因變數的值是從一個非正態的總體樣本中取得而造成的。但是，引數和因變數的散列點裏面的不尋常的值也可能並不是異常。在回歸分析中，擬合的直線對異常是非常敏感的。換句話說就是，最小二乘回歸和擬合斜率的估計抗異常的能力較差。一個數據點從另外的一些點中豎直地移動下來會造成擬合的直線更貼近這個點，而不是跟隨剩下這些點的線性趨勢，尤其當這個點是水平地遠離另外一些點的中心時。

但是，當去除這些異常時，必須非常小心。儘管在大多數情況下，當異常被去除後，回歸的結果往往看起來更好，但是在此之前必須進行先驗論證。舉例來說，在對某一特別公司的股票收益表現進行回歸時，由於股票市場低迷造成的異常應該被保留下來。這些並不是真正的異常，只是體現了商業周期的必然性。在回歸時去掉這些異常來預測如基於公司股票的退休金將有可能會產生不正確的結果。但是，如果這些異常是由於不可重現的商業條件（比如，企業的並購）及不可重現的商業結構的變化造成時，往往需要在使用回歸分析前去除這些異常點。這裏所講述的內容只涉及如何辨認異常資料，在實際應用中仍需要由使用者根據實際情況來決定保留還是去除這些異常點。

有時候，因變數和引數之間的關係用非線性關係來描述比線性關係更合適。那麼，對於這些情況，使用線性回歸將不是最優的選擇。如果線性模型不是正確的選擇，那麼斜率和截距的估計和線性回歸的擬合值就是有偏的，同時擬合的斜率和截距估計量將是沒有意義的。當規定引數和變數的範圍時，非線性模型可以近似地看成線性模型（事實上，這是線性截距的主要部分），但是要精確預測就要選擇對資料解釋合理的模型。在使用回歸錢，必須先對資料進行非線性變換。這方面，一個簡單的方法是對引數進行對數運算（另外的方法包括對因變數進行開根或平方、三次方運算），並且對預測量使用

這些非線性變化後的資料進行回歸。

诊断结果

变量	异方差		微数缺测性		异常		潜在的异常数	非线性	
	W检验 p值	假设检验 结果	近似 结果	本性 下屈	本性 上界	非线性检验 p值		假设检验 结果	
变量 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	

圖 5.23 - 異常性，異方差性，微數缺測性，和非線性的檢驗結果

在預測時間序列資料時一個典型的問題是這些引數是相互獨立的還是存在某種相互關係。因變數的時間序列資料可能是自相關的。對那存在序列相關的因變數資料，斜率和截距的估計是無偏的，但是得到的預測值和方差是不可靠的，因此擬合的統計檢驗有效性是有缺陷的。譬如，利率、通貨膨脹、銷售量、收入、和其他諸如此類的時間序列資料明顯是自相關的，現階段的值和前一階段的值有關（顯然，三月的通貨膨脹資料和二月份的通貨膨脹資料有關，二月份的和一月份的有關，這種相互關係可一直找尋下去）。

如果忽略這種關係，那得到的預測是有偏的，也是不精確的。對這樣的情況，使用自回歸模型或自回歸求和滑動平均效果將會更好（Risk Simulator 預測 IARIMA）。最後，需要指出，對那些非穩態資料的自相關函數趨向於較慢的衰減（參見非穩態報告）。

例如，如果自相關 AC (1) 為非零值，意味著序列是一階序列相關的。如果 AC 隨著滯後的增加呈幾何下降趨勢，這意味著序列遵循一個低階自回歸過程。如果經過幾次滯後之後 AC 值趨於 0，這意味著序列遵循一個低階移動平均過程。相反的，PAC 衡量了在移除了滯後干擾後的 k 階相關值。如果自相關模式可以通過小於 k 階的自回歸解決，那麼 k 階滯後的部分自相關值趨近於 0。報告裏同時還提供了 Ljung-Box 的 k 階滯後 Q 統計值和 p 值，此時被檢驗的原假設是 k 階時不存在自相關。自相關的虛線圖近似在兩個標準差的範圍。如果自相關值在此範圍之內，那麼在 5% 的顯著性水平內它不顯著區別於 0。尋找到合適的 ARIMA 模型需要嘗試和經驗。AC, PAC, SC 和 AIC 都是識別正確模型的有效診斷工具。

自相關用來測量引數 Y 的現在的數值和過去的數值的相互關係。與這種分佈上滯後相對應的是引數和不同因變數 X 之間的時間滯後關係。舉例來說，抵押率的走勢往往會跟從聯邦準備金率的走勢，但是會有時間上的滯後性（典型的是 1 到 3 個月）。有時，時間滯後帶有周期性和季節性（例如，冰激凌的銷售量會在夏季的月份裏達到最大，因此會和 12 月前夏季的銷售量有關）。下面的分佈滯後分析顯示了在各種時間滯後下（這裏的滯後是同時發

生的)，因變數和每個引數之間的關係，並判斷那些時間滯後是統計上顯著的，應該被考慮。

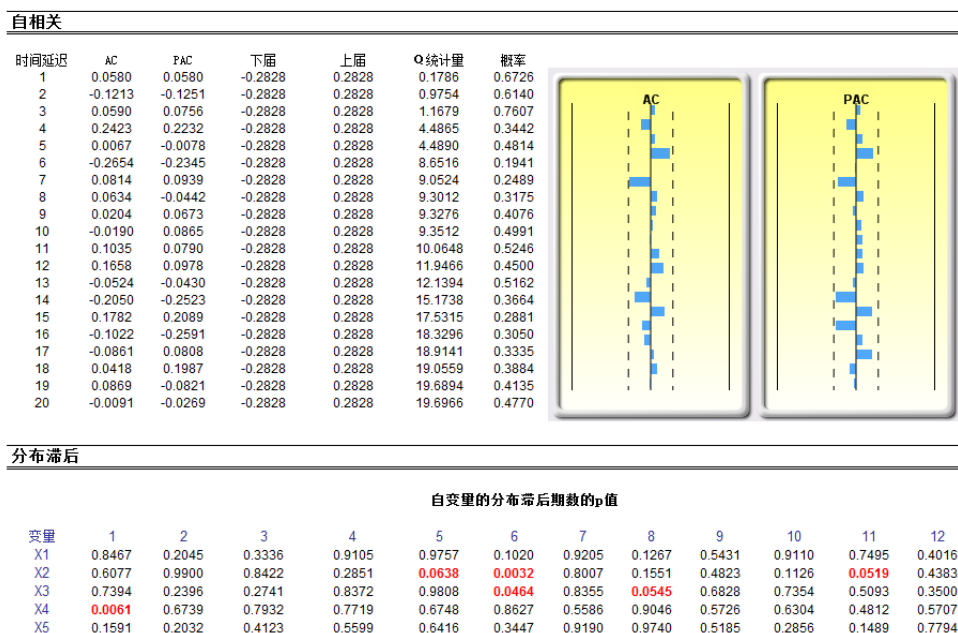


圖 5.24 - 自回歸分佈滯後結果

在運行回歸模型時的另外一個假設是誤差項的正態性和球形假設。如果正態性假設違背或出現異常，那麼線性回歸模型的擬合度檢驗（用來表明是否採取線性擬合）就可能不是判別模型好壞的最有效的檢驗方法。如果誤差不是獨立的和常態分佈的，這就可能表明資料可能是自相關的，或含有非線性或者另外的更具破壞性的誤差。誤差的獨立性也可通過異方差性檢驗（圖 5.25）來探測。

誤差的正態性檢驗是一個非參數檢驗，該方法並不需要樣本總體形狀的假設，對小樣本資料的分析也可使用。該檢驗使用的零假設是樣本誤差服從常態分佈的，備擇假設是樣本誤差不是常態分佈的。如果在各種有效性水平下，計算得到的 D 統計量都大於或等於 D 臨界值，那就拒絕零假設並接受備擇假設（誤差不是常態分佈的）。相反地，如果 D 統計量小於 D 臨界值，那麼就不拒絕零假設（誤差是常態分佈的）。該檢驗依靠兩類累積頻率：一種來自於樣本資料集，另一種來自於基於樣本均值和方差的理論分佈。

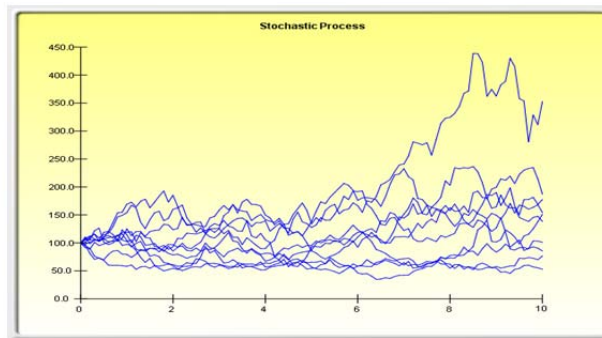
检验结果

		误差	相关频率	观察值	预期值	O-E
回归平均误差	0.00					
误差的标准差	141.83	-219.04	0.02	0.02	0.0612	-0.0412
D 统计量	0.1036	-202.53	0.02	0.04	0.0766	-0.0366
1%下的D临界值	0.1138	-186.04	0.02	0.06	0.0948	-0.0348
5%下的D临界值	0.1225	-174.17	0.02	0.08	0.1097	-0.0297
10%下的D临界值	0.1458	-162.13	0.02	0.10	0.1265	-0.0265
零假设：误差是正态分布的。		-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
结论：误差是正态分布的在 1% 显著性水平下。		-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

圖 5.25 - 誤差的正態性檢驗

有時，某種類型的時間序列資料由於其代表的事件在本質上是隨機的，從而使得除了隨機過程外不能使用其他的方法來建模。舉例來說，由於股票價格，利率，石油價格和其他商品的價格是高度不確定和波動的，你並不能通過使用單變數回歸模型來充分地類比和預測這些值。換句話說，這些過程不是平穩的。平穩性這裏通過使用 Runs 檢驗來測試，並且在自相關報告中也可得到其他的視覺化效果（自相關係數趨向於緩慢地減少）。隨機過程是指服從某一概率分佈的一系列事件和軌迹的集合。就是說，隨機事件雖然隨著時間發生但這些事件都符合某一特殊的統計和概率法則。我們主要碰到的隨機過程包括隨機遊走（布朗運動），均值回復，跳躍-擴散這些。這些過程能用來預測大多數體現隨機傾向但又服從概率分佈的變數的變化。產生該過程的方程是實現已知的，但是產生的結果確實未知的。（圖 5.26）

隨機遊走（布朗運動）能用來預測股票的價格，商品的價格和任何沿著漂移路徑有著漂移或增長率和波動率的隨機時間序列資料。均值回復可以通過遠期目標水平來減小隨機遊走的波動。該過程可以用來預測如利率，通脹這些有長期目標水平（這些長期目標水平由權威機構和市場提供）的時間序列變數。跳躍-擴散過程可以用來預測如石油價格，電力價格（個別外部事件的發生能使價格往上跳躍或下降）這些偶爾伴有隨機跳躍的時間序列資料。最後，這三類隨機過程可根據需要相互配合使用。



统计汇总

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

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

随机模型拟合的概率： 46.48%
高的拟合值意味着随机模型比传统模型要好

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

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

圖 5.26 - 隨機過程參數檢驗

當引數之間存在線性關係時，就稱之存在多重共線性。當處於這樣的情況時，就不能使用回歸模型來估計了。在共線性情況下，回歸估計是有偏的，結果是不精確的。當使用逐步回歸時，得到的結果就會像上述描述的情況，統計上有效的引數會很早地從回歸模型中剔除，使得得到的結果既无效，也不精確。一種在多元回歸模型中判斷多重共線性的快速有效的方法是查看當 t 統計量比較小時， R^2 的值是否比較大。

另一種快速的檢驗方式是創建引數的相關性矩陣。相互的相關性比較高意味著潛在的自相關性。當相關係數的絕對值大於 0.75 時，就認為引數間存在劇烈的多重共線性。多重共線性的另一種檢驗方法是通過計算方差膨脹因數（VIF）。可通過對每個引數用其他引數來回歸後，得到 R^2 來計算 VIF。如果 VIF 大於 2.0，表明存在劇烈的多重共線性。如果 VIF 大於 10.0，表明存在破壞性的多重共線性。（圖 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

圖 5.27 - 多重共線性誤差

相關性矩陣通過變數之間的 Pearson 乘積項（一般被稱為 Pearson 係數）來表示變數之間的相互關係。這些相關係數取值在-1 到 1 之間，包括兩個端點。它們的符號反映了變數之間聯繫的方向性，大小則反映了這種聯繫的強弱。Pearson 的相關係數僅僅測量了線性相關性，對非線性相關性不是很有效。

雙尾假設檢驗用來判斷變數之間的相互關係是否是顯著的，檢驗結果的 P 值被顯示出來。P 值小於 0.1, 0.05 和 0.01 的用藍色顯示，表明是統計上有效的。換句話說，某個相關變數的 p 值小於給定的有效性水平時，表明變數之間的相互關係是統計上顯著不同於 0 的，表明兩個變數之間存在顯著的線性關係。

兩個變數（x 和 y）的 Pearson 乘積項係數（R）是和協方差（cov）有關，運

算式為：
$$R_{x,y} = \frac{COV_{x,y}}{s_x s_y}$$
。這裏協方差除以兩個變數的標準差（s）的優勢是

可以把相關係數控制在-1 到 1 的範圍內。這使得這種測量方法能較好地反映不同變數之間的相互關係（尤其對那些不同單位和量級的變數）。Spearman 基於排列的非參數相關性也在下面給出。Spearman 的相關係數是通過將資料先排列，然後在求排列的這種相關性得到的。當變數之間存在非線性關係時，排列的相關性能提供一種更好的估計。

需要指出的是，存在顯著的相關性並不意味著變數之間含有因果關係。變數之間的這種聯繫並不表示改變一個變數的值，另一個變數也會改變。當兩個變數以相關的路徑各自獨立地變動時，它們可能是相關的，但這種相關關係可能是沒有道理的（例如，太陽黑子的數目和股票的市場價格存在較強的相關性，但是這裏面並不存在因果關係，這種相關性完全是偽造的）。

5.9 統計分析工具

Risk Simulator 軟體中另外一個非常有用的工具就是統計分析工具，它可以發現資料的統計特徵。診斷工具運行包括對資料的統計特性進行描述，對亂數據進行基本的描述性統計核對總和校正。

操作過程描述：

- 打開示例模型 (Risk Simulator | 示例模型 | 統計分析) 回到資料工作簿，選擇資料包括變數名稱(單格 C5:E55)。
- 點擊 Risk Simulator | 工具 | 統計分析 (圖 5.28)。
- 點擊 **資料類型**，選擇的資料來自於一列或者多列。本例中，資料來自於多列。點擊**確定**完成選擇。
- 選擇想要進行的統計檢驗。建議 (軟體默認) 是選擇所有的檢驗。點擊**確定**完成選擇 (圖 5.29)。

請參閱生成的報告更好的理解統計檢驗結果的意義。(示例報告顯示如下圖 5.30-5.33)

變數 X1	變數 X2	變數 X3
521	18308	185
367	1148	600
443	18068	372
365	7729	142
614	100484	432
385	16728	290
286	14630	346
397	4008	328
764	38927	354
427	22322	266
153	3711	320
231	3136	197
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
606	5242	284
512	92629	499
426	28795	231
47	4487	143
265	48799	249
370	14067	195
312	12693	288
222	62184	229
280	9153	287
759	14250	224
114	3680	161

統計分析

此工具用來預測和發現一系列原始資料之間的統計關係。

選取資料

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

資料來自單一變數
 資料包含多種中的多個變數

確定(O) 取消(C)

統計分析

選擇需要運行的分析

運行: 所有分析

描述性統計

分佈擬合

直方圖和其他圖表

假設檢驗

非線性外推

正態性

隨機過程參數估計

周期性: 年

時間序列自回歸

時間序列預測

季節性 (周期/圖圈): 4

預測 (周期): 4

趨勢線預測

TrendLines: 4

確定 取消

圖 5.28 - 運行統計分析工具

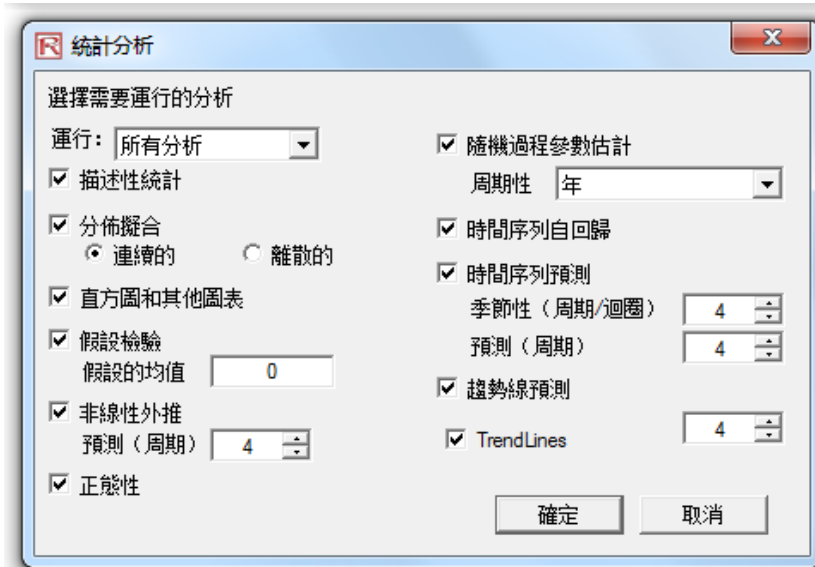


圖 5.29 - 統計檢驗

統計分析		描述性統計																																																					
<p>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。</p> <p>基本平均数通过加总所有的数据值再除以数据个数得到发生的平均值。几何平均数就是数据乘积的基本根, 并且数据都必须为正数。在计算百分比或者比例的时候, 使用几何平均数则更为精确。例如, 使用几何平均数根据变化率计算平均增长率。截尾平均数是截取两边极大值之后的代数平均数。当两边的取值存在的时候, 均值会向重要差异的方向倾斜, 而截尾平均数在有偏分布中避免了这个问题。</p> <p>均值的标准误差计算了样本均值的误差大小。样本越大, 误差越小, 对于无限大的样本总体, 误差就越接近于零, 说明已经估计到了样本参数。由于抽样的误差, 需要提供95%的置信区间用于均值的估计。基于样本数据的分析, 实际的样本均值落在上下区间之内。</p> <p>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, 中位数受分布两边影响最小。对称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。</p> <p>第二矩用来测量分布的幅度或离散度, 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散程度。它通常作为与风险相关的一个测定值 (越高的标准差意味着越广的分布, 越高的风险, 或者数据在均值附近分散的更广) 它的单位等于原始数据的单位。样本标准差不等于总体标准差, 前者使用了对于小样本的自由度。总体样本的标准差会落入这个区间。如果数据包含整个样本总体, 使用总体标准差。因此, 两个方差也就是各自标准差的平方。</p> <p>变异系数是样本的标准差除以样本的均值, 提供了一个无量纲的计算结果。可以在多个分布之间比较 (可以将以百万作为计数单位的分布和以十亿作为计数单位的分布做比较, 或者以米和千米做单位的分布进行比较, 等等)。第一分位数用来度量按照数据从小到大排列的第25百分数。第50分位数是七十五百分数。有时候, 分位数可以作为分布的上下限, 因为它截取了数据不去考虑分布的两端。四分位数间距不同于第一和第三分位数, 常用来度量分布中心范围的宽度。</p> <p>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。</p> <p>峰度体现了分布与正态分布相比较平坦或者尖峭的特性。是分布的第四矩。正的峰度说明分布相当的尖峭。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。较适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。</p>																																																							
<p>统计汇总</p> <table border="1"> <thead> <tr> <th>统计量</th> <th>变量 X1</th> <th></th> <th></th> </tr> </thead> <tbody> <tr> <td>观测值</td> <td>50.0000</td> <td>标准差 (样本)</td> <td>172.9140</td> </tr> <tr> <td>算术平均数</td> <td>331.9200</td> <td>标准差 (总体)</td> <td>171.1761</td> </tr> <tr> <td>几何平均数</td> <td>281.3247</td> <td>标准差的下置信区间</td> <td>148.6090</td> </tr> <tr> <td>截尾平均</td> <td>323.1739</td> <td>标准差的上置信区间</td> <td>207.7947</td> </tr> <tr> <td>基本平均数的标准差</td> <td>24.4537</td> <td>方差 (样本)</td> <td>29599.2588</td> </tr> <tr> <td>均值的下置信区间</td> <td>283.0125</td> <td>方差 (总体)</td> <td>29301.2726</td> </tr> <tr> <td>均值的上置信区间</td> <td>380.8275</td> <td>变异系数</td> <td>0.5210</td> </tr> <tr> <td>中位数</td> <td>307.0000</td> <td>第一四分位数 (Q1)</td> <td>204.0000</td> </tr> <tr> <td>众数</td> <td>47.0000</td> <td>第三四分位数 (Q3)</td> <td>441.0000</td> </tr> <tr> <td>极大值</td> <td>764.0000</td> <td>四分位数间距</td> <td>237.0000</td> </tr> <tr> <td>极小值</td> <td>717.0000</td> <td>偏度</td> <td>0.4838</td> </tr> <tr> <td>极差</td> <td></td> <td>峰度</td> <td>-0.0952</td> </tr> </tbody> </table>				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	323.1739	标准差的上置信区间	207.7947	基本平均数的标准差	24.4537	方差 (样本)	29599.2588	均值的下置信区间	283.0125	方差 (总体)	29301.2726	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
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	323.1739	标准差的上置信区间	207.7947																																																				
基本平均数的标准差	24.4537	方差 (样本)	29599.2588																																																				
均值的下置信区间	283.0125	方差 (总体)	29301.2726																																																				
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																																																				

圖 5.30 - 示例統計分析報告

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): $\mu =$ 假设的均值 备择假设 (Ha): $\mu <>$ 假设的均值 注意: "<>" 代表 右尾检验中 "大于的意思", 代表作为检验 "小于的意思", 或者 "不等于" 用于双尾检验。	
假设均值	0.00		
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。			

圖 5.31 - 示例統計分析報告 (單變數假設檢驗)

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	数据	相关频率	观察值	估计值	O-E
标准差	172.91	47.00	0.02	0.0497	-0.0297
D 统计量	0.0859	68.00	0.02	0.04	-0.0235
1%显著性水平的D值	0.1150	87.00	0.02	0.06	-0.0183
5%显著性水平的D值	0.1237	96.00	0.02	0.08	-0.0062
10%显著性水平的D值	0.1473	102.00	0.02	0.10	0.0918
零假设: 数据的分布服从正态分布		108.00	0.02	0.12	0.0977
		114.00	0.02	0.14	0.1038
		127.00	0.02	0.16	0.1180
		153.00	0.02	0.18	0.1504
		177.00	0.02	0.20	0.1851
		186.00	0.02	0.22	0.1994
		188.00	0.02	0.24	0.2026
		198.00	0.02	0.26	0.2193
		222.00	0.02	0.28	0.2625
		231.00	0.02	0.30	0.2797
		240.00	0.02	0.32	0.2975
		246.00	0.02	0.34	0.3096
		251.00	0.02	0.36	0.3199
		265.00	0.02	0.38	0.3494
		280.00	0.02	0.40	0.3820

结论: 样本数据是正态分布的 1% 显著性水平.

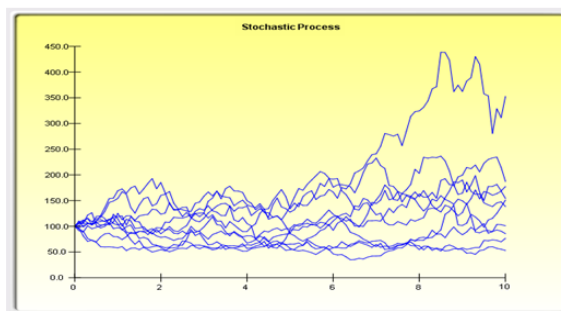
圖 5.32 - 示例統計分析報告(正態檢驗)

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飙升或者下降）。最终，这三种随机过程可以混合出现。



统计汇总

下面是在给定数据的条件下估计随机过程的参数。这可由个人决定该是否应该使用随机过程来拟合（类似于拟合优度的计算），如果是的，这个随机过程是一个随机游走，均值回复，还是跳跃扩散的模型，或者是一个三者混合在一起的模型。需要根据以往的经验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%
波动性	88.84%	长期值	327.72
		跳跃率	20.41%
		跳跃大小	237.89
随机模型的概率拟合			46.48%

圖 5.33 - 示例統計分析報告(隨機參數估計)

5.10 分佈分析工具

這是 Risk Simulator 軟體中的統計概率密度工具，在進行一系列的設定之後會十分的有用，可以用來計算概率密度函數（PDF），和離散資料的概率密度函數（PMF），兩者可以交替使用。再給出了某個分佈的參數，我們就可以決定某些事件 x 發生的概率水平。此外，累積的概率密度函數（CDF）也可以計算出來，它就是對發生事件 x 的 PDF 的加總。最後，逆累積概率密度函數（ICDF）被用來計算在給定概率發生水平時的 x 值。

可以通過點擊 **Risk Simulator | 工具 | 分佈分析**。作為示例，圖 5.34 顯示了二項分佈（例如，一個事件只有兩種可能性水平，例如擲錢幣，只有頭像和背面兩種可能性，這裏的事件可以是頭像也可以是背面，事先定義好頭像出現的可能性水平）。假設投擲錢幣兩次，設定頭像出現為成功，使用二項分佈，試驗次數為 2（投擲兩次），概率水平=0.50（成功的概率或者頭像出現的概率）。選擇 PDF，設定 x 的範圍，從 0 到 2 步長為 1，（這意味著將 0, 1，

2 作為 x 值)，結果以圖像和表格的形式輸出，也包括理論的分佈四矩。出現的結果可能是頭像-頭像，背面-背面，頭像-背面，背面-頭像，因此頭像不出現的概率為 25%，一個頭像出現的概率為 50%，兩個頭像出現的概率為 25%。

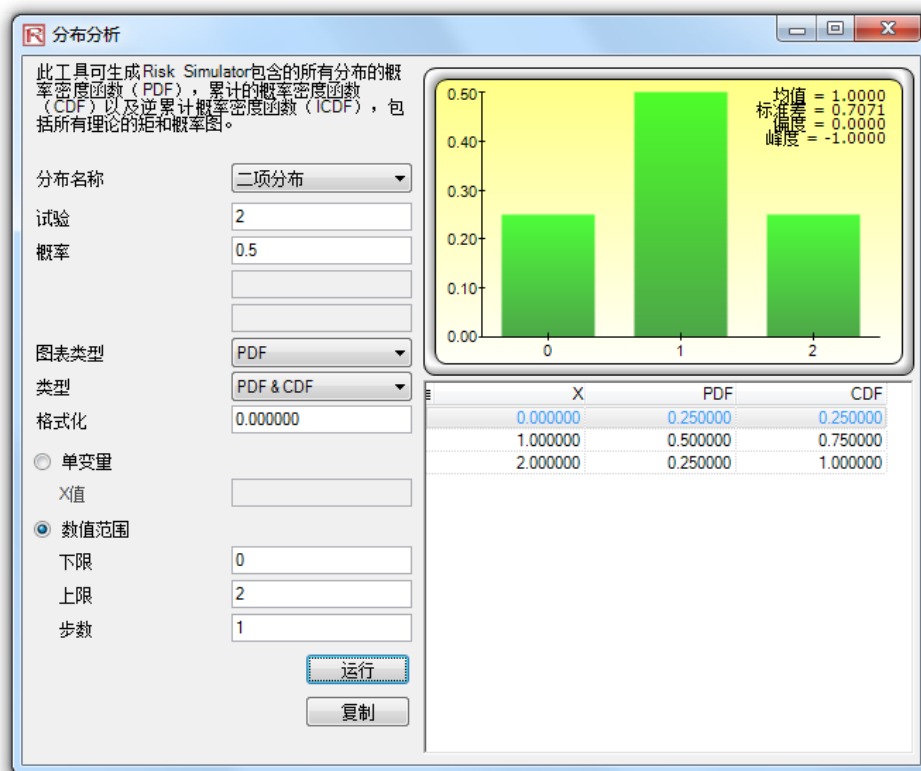


圖 5.34 - 示例統計分析報告(2 次試驗的二項分佈)

同樣地，我們可以得到投擲錢幣的概率，以 20 次為例，如圖 5.35。結果同樣以表格和圖形的形式輸出。

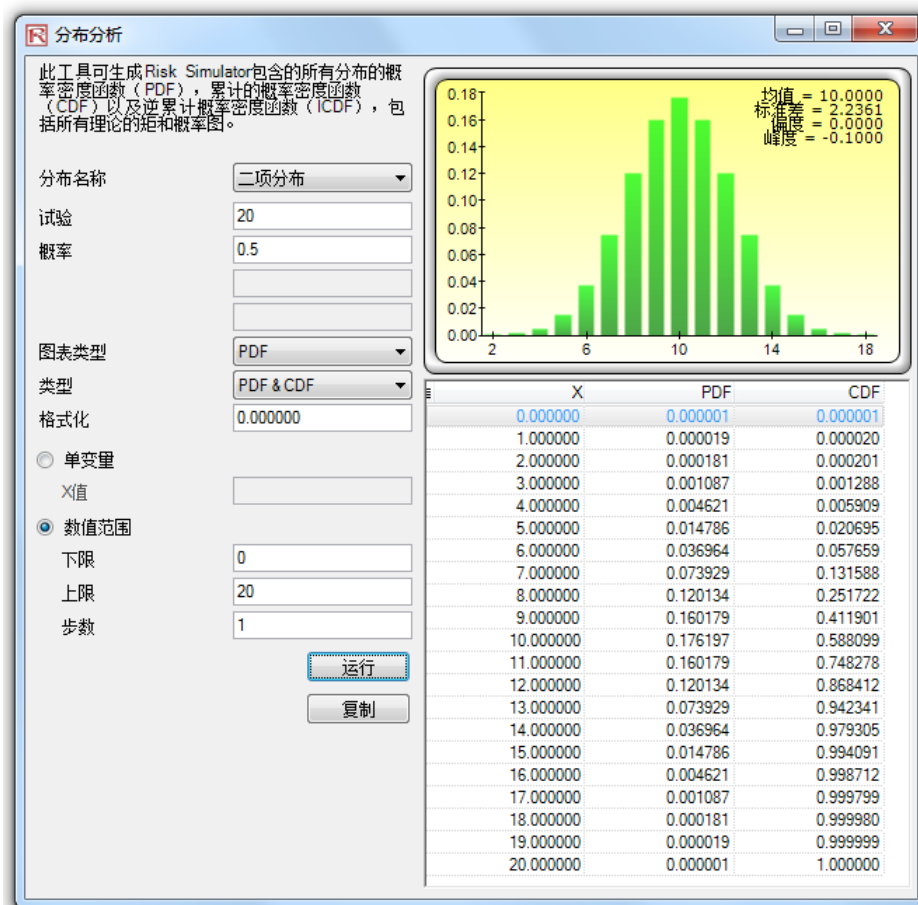


圖 5.35 - 示例統計分析報告(20 次試驗的二項分佈)

圖 5.36 顯示了二項分佈以及如何計算 CDF 的。CDF 就是對每個 x 點對應的 PDF 進行加總。例如，在圖 5.35，我們看見對於試驗 0, 1, 2 對應的概率為 0.000001, 0.000019, 和 0.000181，總和為 0.000201，也就是 $x = 2$ 時的 CDF 值，如圖 5.36。PDF 計算了出現兩次頭像的概率，CDF 計算出不超過兩次頭像的概率（或者出現 0, 1, 2 次頭像的概率）。相減之後（例如， $1 - 0.00021$ 得到 0.999799 或者 99.9799%）提供了出現 3 次或者超過 3 次頭像的概率。

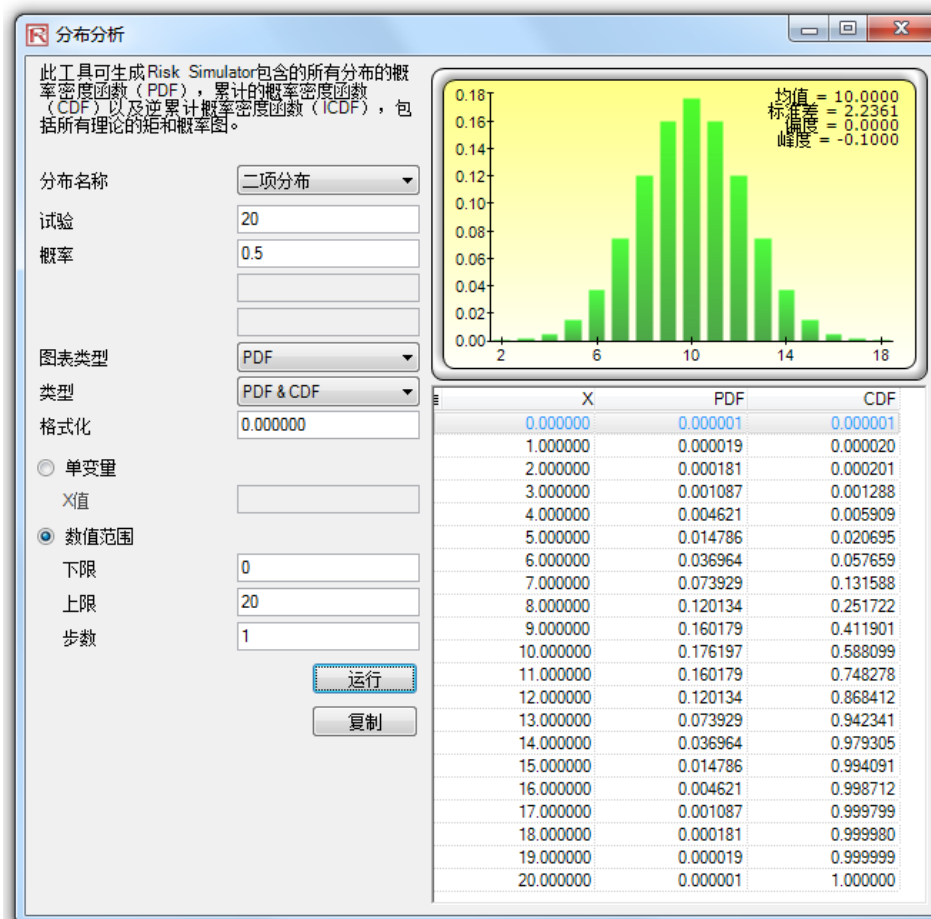


圖 5.36 - 示例統計分析報告(20 次試驗二項分佈的 CDF)

使用分佈分析工具，可以對一些高級的分佈進行分析，例如 gamma 分佈，beta 分佈，逆二項分佈，等其他一些包含在 Risk Simulator 中的分佈。下面是對於連續型概率分佈在該工具應用的示例，圖 5.3.7 顯示了標準常態分佈（均值為 0，標準差為 1 的常態分佈），這裏使用 ICDF 發現對於累積的概率 97.5%（CDF）對應的 x 值。即，單尾 97.5% 的 CDF 等於雙尾的 95% 的置信區間（即 2.5% 概率水平在左尾，2.5% 的概率水平在右尾，中間或者置信域的水平為 95%，就等於單尾 97.5% 的區域）。結果類似於 Z 值的 1.96。因此，使用分佈分析工具，可以得到標準的分佈值，具體的和累積的概率值，十分地方便和簡單。

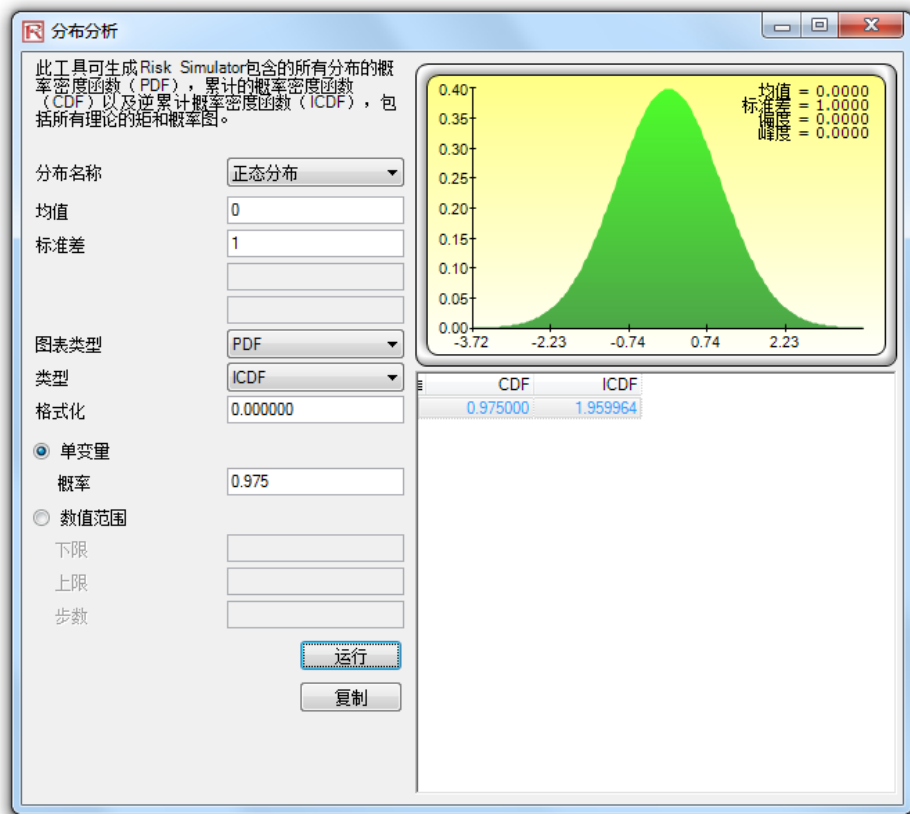


圖 5.37 - 示例統計分析報告(常態分佈的 ICDF 和 Z 值)

5.13 Risk Simulator 2011/2012 新工具

5.14 亂數產生，蒙特卡羅與拉丁超立方，以及相關性 Copula 方法

啟動 Risk Simulator 2011/2012 版,有 6 種亂數產生器,3 種相關性 Copula 方法,及 2 種模擬抽樣方法供以選擇(圖 5.41)。這些選項可以通過打開位置 *Risk Simulator*選擇進行設置。

亂數產生器 (RNG) 是任何模擬軟體的核心。基於產生的亂數，可以構造不同的數學分佈。本軟體預設使用方法是專有的 ROV 風險模擬方法，它提供了最好的和最可靠的亂數。共有 6 種亂數發生器提供支援，在一般情況下，ROV Risk Simulator 的預設方法和高級減法隨機打亂法是推薦使用的兩種方法。除非你的模型或分析方法有特定的要求，不需要應用其它的方法，即使如此，我們仍然建議針對這兩種方法對結果進行測試。越向 RNG 方法清單的下面，演算法越簡單，運行的速度越快，相反，越向 RNG 方法清單的上面，結果越可靠。

在相關性部分，支援三種方法：正態 Copula，T-Copula，和准偽正態 Copula。這些方法依賴數學積分技術，當你有疑問的時候，正態 copula 方法提供最安全和最保守的結果，T - Copula 方法提供模擬分佈尾部的極值，而偽正態 copula 方法返回這些極值之間的結果。

在模擬方法部分，支援蒙特卡羅模擬 (MCS) 和拉丁超立方抽樣 (LHS) 方法。請注意，Copula 方法和其它多元函數與 LHS 是不相容的。這是因為 LHS 可以應用到一個單變數而不是一個聯合分佈。實際上，一個模型的分佈越多，LHS 對模型輸出結果的精度只有非常有限的影響，因為只是分別應用到每個分佈。如果使用者在開始的時候沒有確定樣本的數量，LHS 的優勢會受到削弱，例如，當用戶在模擬的過程中停止運行模擬的時候。LHS 也會對一個有著大量輸入的模型有沉重的負擔，因為在一個分佈中運行第一個樣本之前，它需要在每個分佈生成和組織樣本。這在運行一個較大的模型的時候產生一個較長時間的滯後，但只提供非常有限的額外準確性。最後，LHS 最好應用在當分佈表現良好，對稱並且沒有任何相關性的時候。然而，LHS 是一個功能強大的方法，可以產生一個抽樣均勻的分佈，當應用 LHS 的時候，與一個抽樣更加均勻的分佈相比（分佈的每一個部分都進行抽樣），MCS 有時候可以產生塊狀分佈（抽樣資料有時可能產生在一個集中的區域）。

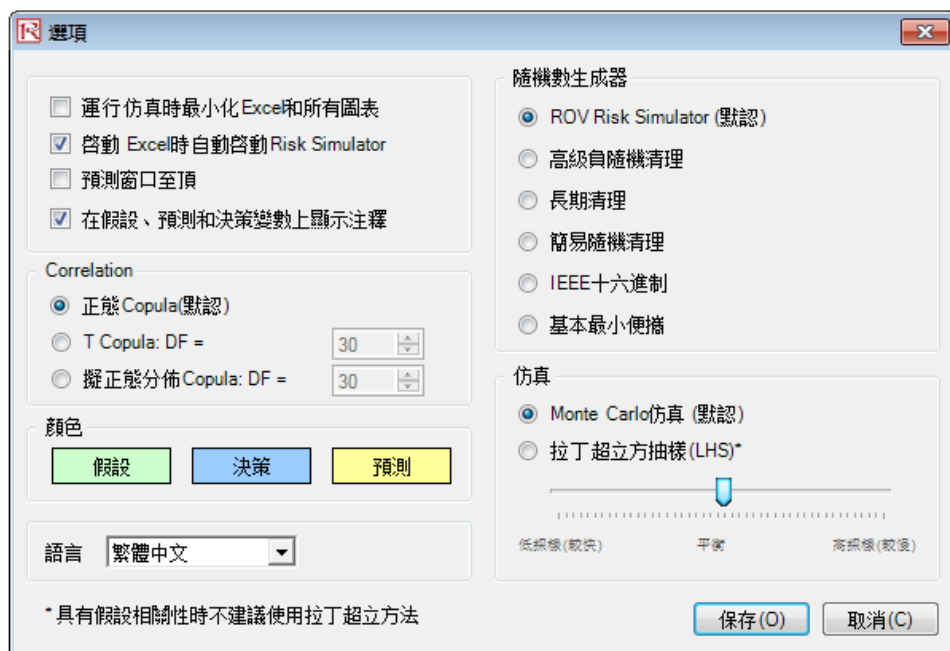


圖 5.41 - Risk Simulator 選項

5.15 資料季節性因素和趨勢性因素消除

此工具能夠對你的原始資料進行季節性因素和趨勢性因素消除處理，意消除季節性和趨勢性的部分（圖 5.42）。在預測模型中，該過程通常用於消除包括來自於季節性和趨勢性的累積影響，僅顯示數值的絕對變化，在消除一系列時間序列資料一般的漂移，趨勢，扭曲，彎曲和季節性週期的影響之後確定潛在的迴圈模式。例如，從一個消除趨勢性之後的資料集可以更加準確地瞭解到一個公司的銷售情況，通過將整個資料集從一個斜坡轉移到一個平面上，可以更好地理解潛在的迴圈和波動性。

很多時間序列資料呈現出季節性特點，其中某些事件在一段時間或者季節性週期之後會重現（滑雪勝地在冬天的收入比夏天的收入高，並且這個可預見的事件在每年冬季會重現）。季節性週期代表在迴圈自身重複之前需要經過多少個期間（例如，一天 24 小時，1 年 12 個月，一年 4 個季度，1 小時 60 分鐘，等等）。此工具可以將你原來的資料消除週期性影響。一個季節性指數大於 1 表明是季節迴圈的一個高峰值，同樣一個季節性指數小於 1 表明是季節迴圈的一個低峰值。

操作流程（節性因素和趨勢性因素消除）：

- 選擇您要分析的資料（例如，B9：B28）然後點擊 Risk Simulator 工具消除資料季節因素和消除趨勢因素

- 選擇**消除資料季節因素**和/或**消除資料趨勢因素**，選擇你希望運行的任何消除趨勢模型，並輸入相關的處理選項（例如，多項式，階移動平均線，差分，比率），然後點擊確定
- 檢查這兩份詳細報告，瞭解提供的方法，應用程式，以及由此產生圖表和消除季節性因素/消除趨勢性因素的資料

操作流程（節性因素和趨勢性因素消除）：

- 選擇您要分析的資料（例如，B9：B28）然後點擊 Risk Simulator|工具|資料季節性因素檢驗
- 輸入進行測試的最大季節性期間。也就是說，如果你輸入 6，該工具將測試以下的季節性時期：1，2，3，4，5，6。時期 1 當然意味著資料沒有季節性
- 檢查這兩份詳細報告，瞭解提供的方法，應用程式，以及由此產生圖表和季節性測試結果。最好的季節性週期排列在第 1 位（通過最低的 RMSE 錯誤測量結果來進行排序）和包含所有的相關錯誤測量以進行比較：均方根誤差（RMSE），均方誤差（MSE），平均絕對差（MAD）的，及平均絕對百分比誤差（MAPE）

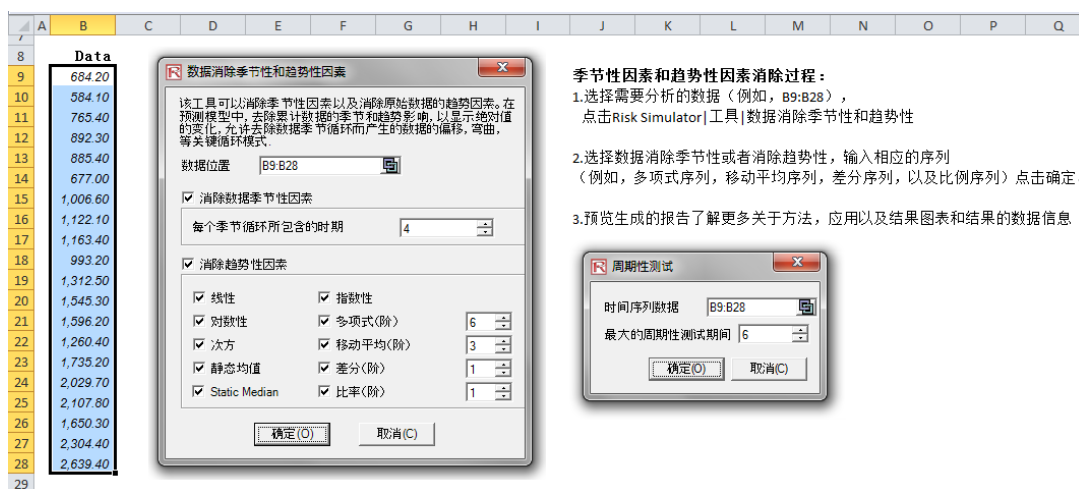


圖 5.42 - 消除資料季節因素和消除趨勢因素

5.16 主成分分析

主成分分析是一種識別資料模式的方式，並且重新塑造資料以突出其共性和差異（圖 5.43）。當多個變數存在的時候，資料模式不容易被辨認。一旦發現了資料的模式以後，可以壓縮資料使其維度減少。資料維度的減少並不意味著信息量的減少。相反，通過較少的變數，也可以獲取相同水準的信息量。

操作程式:

- 選擇資料進行分析（例如，B11：K30），點擊 *Risk Simulator|工具|主成分分析* 和按一下確定
- 檢查生成的報告獲取計算結果

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
10	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	过程:				
11	96.998	87.223	102.443	112.765	111.984	117.331	78.164	97.658	110.950	89.133	1.选择需要分析的数据（例如，B11：K30），				
12	93.098	83.096	81.531	90.224	92.265	78.821	94.321	95.960	101.349	96.345	点击Risk Simulator 工具 关键成分分析，点击确定				
13	96.730	96.298	113.426	99.147	98.138	94.868	119.722	108.657	123.757	93.451	2.预览生成的报告和计算的结果				
14	116.615	83.876	105.389	109.022	119.189	99.155	94.762	106.751	96.187	107.576					
15	85.558	91.528	84.784	96.371	99.675	100.281	96.773	121.945	82.575	92.635					
16	74.224	114.477	87.202	93.464	107.577	104.667	108.746	105.957	86.282	88.843					
17	106.940	103.226	90.602	97.591	101.315	105.578	101.387	90.890	118.848	104.872					
18	100.722	108.298	108.620	93.635	90.768	111.112	87.988	84.411	107.113	106.384					
19	122.057	114.438	113.039	101.130	100.020	104.537	99.745	89.453	82.252	108.283					
20	104.442	106.179	102.135	89.731	112.382	96.888	91.601	91.789	95.710	95.466					
21	94.762	108.494	105.132	93.917	113.050	82.391	105.506	98.837	100.417	93.459					
22	94.504	108.493	108.030	104.564	106.914	116.306	103.039	105.890	118.528	96.644					
23	110.383	101.435	111.410	98.517	92.202	110.760	94.182	105.339	105.458	96.836					
24	95.592	86.340	119.930	94.335	100.861	97.657	128.354	112.520	108.809	113.322					
25	101.879	105.420	97.504	87.789	112.667	97.111	86.941	107.643	107.843	104.282					
26	104.039	93.519	107.231	105.253	110.750	72.306	104.638	114.671	82.774	100.455					
27	113.540	116.882	102.387	101.451	118.545	99.574	93.431	109.074	99.901	110.392					
28	104.347	114.534	98.788	90.383	84.614	74.349	101.032	102.992	99.822	102.005					
29	102.582	114.762	100.853	88.833	86.101	101.915	109.511	84.912	93.900	105.235					
30	97.832	96.564	98.365	95.603	91.974	106.448	100.588	112.635	102.622	100.571					

圖 5.43 - 主成分分析

5.17 結構性中斷分析

結構性中斷分析可以檢驗不同資料組的係數是否相同，在時間序列分析中常用於檢驗資料是否存在結構性突變（圖 5.44）。時間序列資料集可以分為兩個子集，然後可以對每個子集分別進行檢驗，統計學上確定資料在某個特定的時期是否開始突變。結構性突變檢驗經常被用來確定獨立變數是否對資料總體子集有不同的影響，進行的測試包括例如一個新的行銷計畫，活動，重大事件，收購，剝離等等對時間序列資料是否有影響。假設資料集有 100 個時間序列資料點，您可以設置各種中斷點進行檢驗，例如，資料點 10，30 和 51（這意味著有三個結構性中斷檢驗將在以下資料集進行：資料點 1-9 和 10-100；資料點與 1-29 和 30-100；以及 1-50 和 51-100，確實在第 10 個點，第 30 個點和第 51 個點開始的資料是否有結構性中斷）。一個單尾假設檢驗對零假設（ H_0 ）進行檢驗，檢驗兩個資料集之間在統計上是否顯著相同。可替換假設（ H_a ）就是兩個資料集之間在統計上顯著不同，這表明一個可能的結構性中斷。如果計算的 P 值小於或者等於 0.01，0.05 或者 0.1，這意味著假設被拒絕，這意味著在 1%，5% 和 10% 的顯著性水準，這兩個資料子集統計上顯著差異，較高 P 值表示沒有顯著的結構性突變。

操作流程:

- 選擇您要分析的資料（例如，B15:D34）和點擊 *Risk Simulator|工具|結構性中斷檢驗*，然後輸入你想要進行檢驗的點（例如，6，10，12），然後點擊 **確定**。
- 檢查報告以確定你的資料上哪些測試點在統計上存在顯著突變和哪些點不存在顯著突變

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

过程：

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

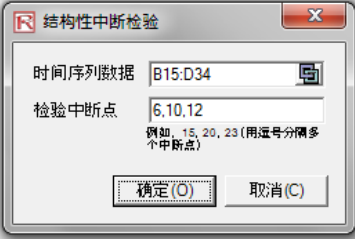


圖 5.44 - 結構性中斷分析

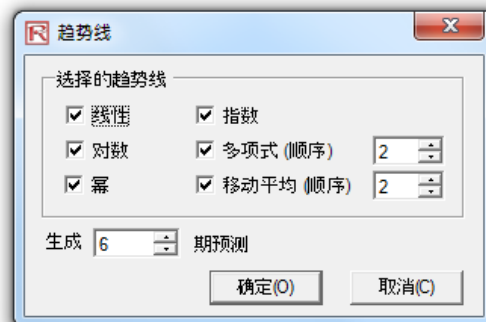
5.18 趨勢線預測

趨勢線可以用來確定一個時間序列資料集是否遵循任何明顯的趨勢（圖 5.45）。趨勢可以是線性或非線性（例如指數，對數，移動平均，電力，多項式，或幕）。

操作程式：

- 選擇您要分析的資料然後點擊 [Risk Simulator|預測趨勢線](#)，選擇你想應用於資料的相關趨勢線，輸入預測的期數（例如，6 期）然後點擊確定。
- 檢查報告確定哪些檢驗趨勢線針對你的資料提供了最優的擬合和最優的預測

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



圖示 5.45 - 趨勢線預測

5.19 模型檢驗工具

在模型創建好並且假設和預測設置好以後，你可以運行模擬或者運行模型檢驗工具（圖 5.46）來檢驗該模型是否被正確設置。另外，如果模型不能運行或者你懷疑模型的一些設置可能不正確，可以從 [Risk Simulator 工具/模型檢驗](#) 來確定你的模型可能存在的問題。請注意，這個工具檢驗的是最常見的模型問題以及在 Risk Simulator 假設和預測中的問題，當然不能全面測試所有類型的問題，這仍然取決於模型的開發者來確保模型是否能夠正常運行。

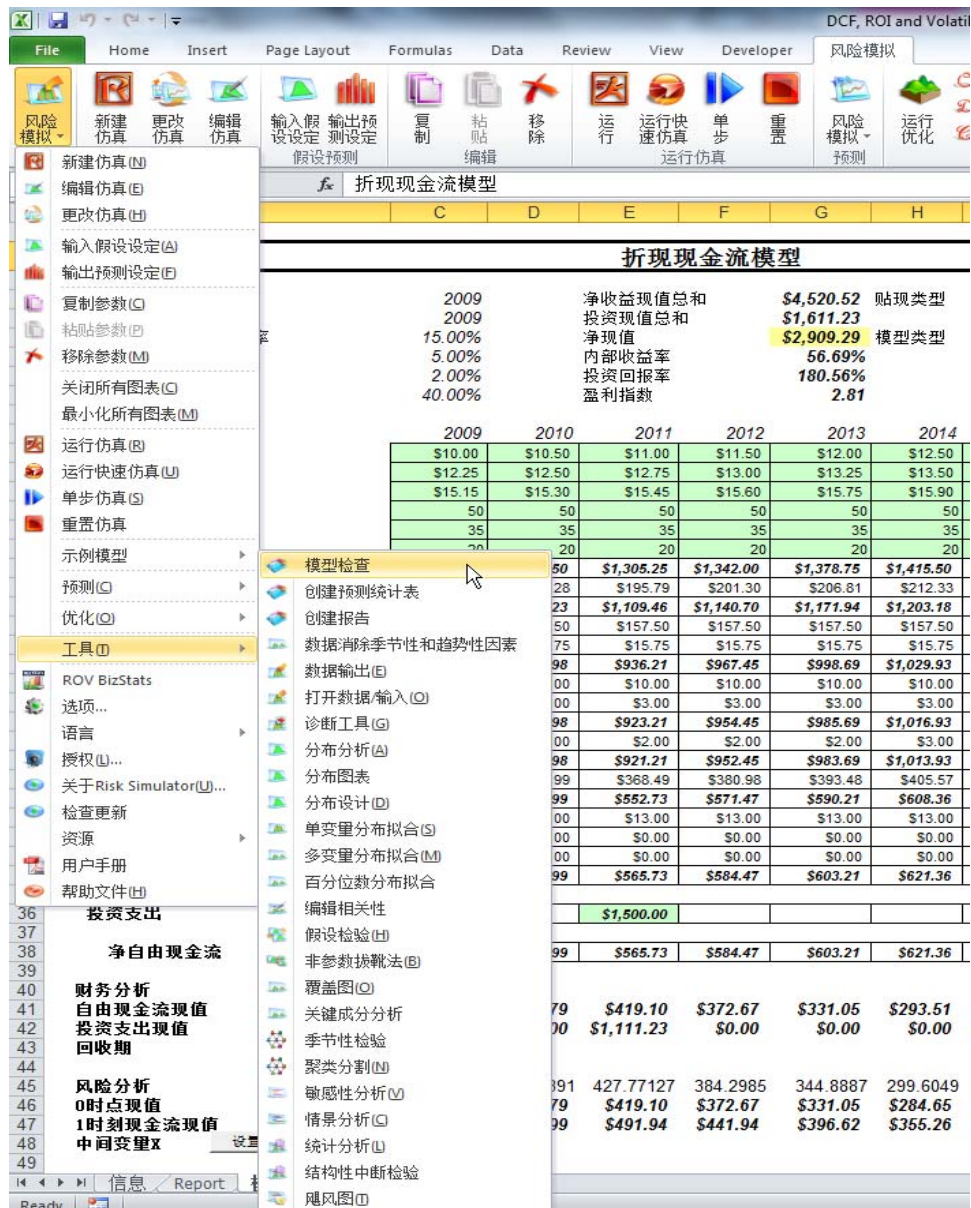


圖 5.46 - 模型檢驗工具

5.20 百分位分佈擬合工具

百分位分佈擬合工具（圖 5.47）是另外一種可替代的擬合概率分佈的方法。有幾個與此相關的分析工具，每一種工具都有自己的用途和優勢：

- 分佈擬合（百分位） - 用另一種輸入的方法（百分位和第一/第二矩組合），你可以找到一個特定分佈的最優擬合參數而不需要原始資料。當沒有足夠資料的時候，這種方法也適合使用，只有當百分位和矩可以獲得的時候，或者當只有兩個或者三個資料點但是分佈類型是

已知或者被假定的時候，這種方法可以作為一種方式來恢復整個分佈

- 分佈擬合（單變數） - 利用統計方法來擬合你的原始資料到所有 42 種分佈，然後進行比較以找出最佳擬合及其輸入參數。好的擬合需要多個數據點，分佈的類型可以不需要提前知道
- 分佈擬合（多變數） - 使用與單變數擬合相同的演算法，使用統計方法對多個變數同時擬合你的原始資料，但是可以在變數之間整合一個相關性矩陣。好的擬合需要多個數據點，分佈的類型可以不需要提前知道
- 定制分佈（設置假設） - 使用現存的原始資料，利用非參數重採樣技術，以生成一個定制分佈，及基於這個經驗分佈來模擬分佈。這種方法需要較少的資料，分佈類型不需要提前知道

操作程式：

- 點擊 *Risk Simulator* 工具 **分佈擬合（百分位）**，選擇概率分佈和你希望使用的輸入類型，輸入參數和點擊 **運行** 來獲取結果。檢查擬合的 R 方結果及比較實證的和理論的擬合結果來確定你的擬合是否擬合得較好。

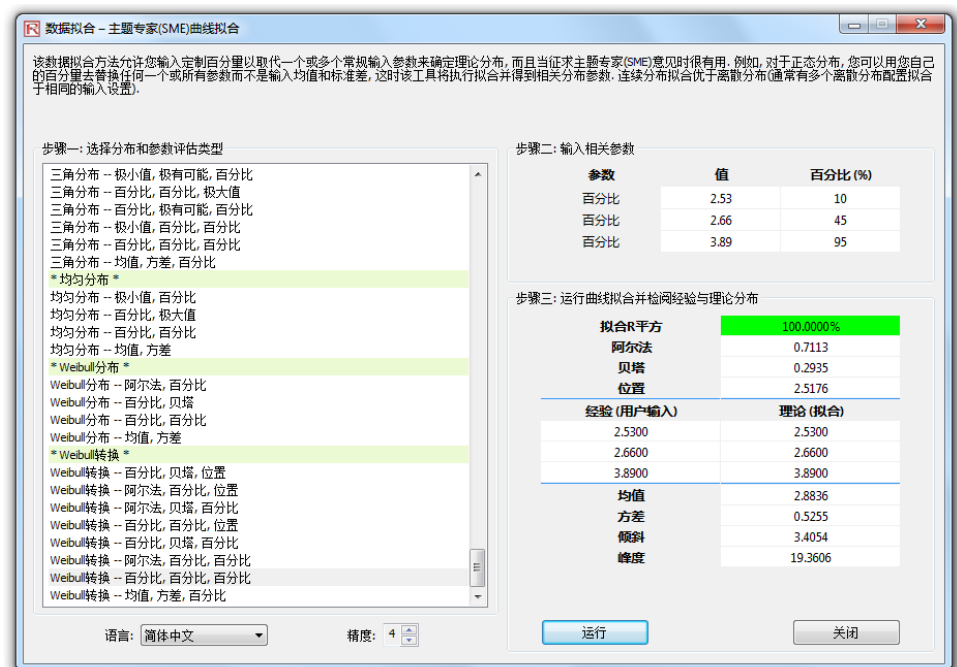


圖 5.47 - 百分位分佈擬合工具

5.21 分佈圖表：概率分佈工具

新的概率分佈工具非常地強大和快捷，可用于生成分佈圖表（圖 5.48-5.51）。注意 Risk Simulator 軟體中包含三種類似的工具，但是彼此的功能各不相同：

- 分佈分析—用於快速的計算 Risk Simulator 中 42 種分佈的 PDF，CDF，和 ICDF，並且返回這些值的概率表。
- 分佈圖表—概率分佈工具在這裡用來比較相同分佈的不同參數（例如，具有不同 Alpha，Beta 參數[2,2]，[3,5]，[3.5,8]的 Weibull 分佈的 PDF，CDF，ICDF 的形狀，並且將彼此重疊到一起）。
- 重疊圖—用於比較不同的分佈（理論輸入假設和試驗模擬輸出預測）並且將彼此重疊比較。

過程：

- 在 *Risk Simulator* 分佈圖表中運行 ROV BizStats，點擊 **應用全域輸入** 按鈕載入一系列輸入參數或者輸入相應的輸入，點擊 **運行** 計算結果。相應的 45 個分佈的四個矩和 CDF，ICDF，PDF 結果都進行了計算（圖 5.48）。



圖 5.48—概率分佈工具（45 種概率分佈）

- 點擊 **圖表欄**（圖 5.49），選擇分佈[A]（例如，正弦函數），選擇想要返回的 CDF，ICDF，或者 PDF[B]，輸入相應的參數，**點擊運行圖或者表**[C]。就可以在圖和表欄之間切換，查看運行的結果，或者點擊圖的圖示[E]

查看圖的效果。

- 可以更改兩個參數[H]輸入到**從到步**生成多個圖表或者使用**自訂**輸入點擊**運行**。例如，圖 5.50 展示的，運行 Beta 分佈，選擇 PDF[G]，選擇 Alpha 和 Beta 改變[H]使用自訂輸入[I]，輸入相應的輸入參數：2;5;5 用於 Alpha;5;3;5 用於 Beta 分佈[J]點擊**運行圖表**。就可以生成三個 Beta 分佈[K]：Beta (2,5)，Beta (5,3)，Beta (5,5) [L]。查看不同的圖形類型，網格，語言和小數設置[M]，以及再次運行分佈使用理論 vs 試驗模擬的值[N]。
- 圖 5.51 顯示了伯努利分佈生成的概率表，使用**從到步**選項[O]，選擇不同的試驗成功概率和試驗成功次數（隨機變數 X）。重複計算，點擊表欄[P]查看創建的概率密度函數結果。在案例中，伯努利分佈的輸入起始點，試驗=20，成功概率=0.5，成功的試驗次數 X=10,顯示的成功概率允許從 0,0.25,...,0.50 變動，如列變數所示，成功試驗次數允許從 0,1,2,...,8，如行變數所示。選擇 PDF，因此結果在表中展示。例如，進行 20 次試驗，每次試驗都 25%機會成功，因此 2 次成功的概率為 0.0669 或者 6.69%的概率值。

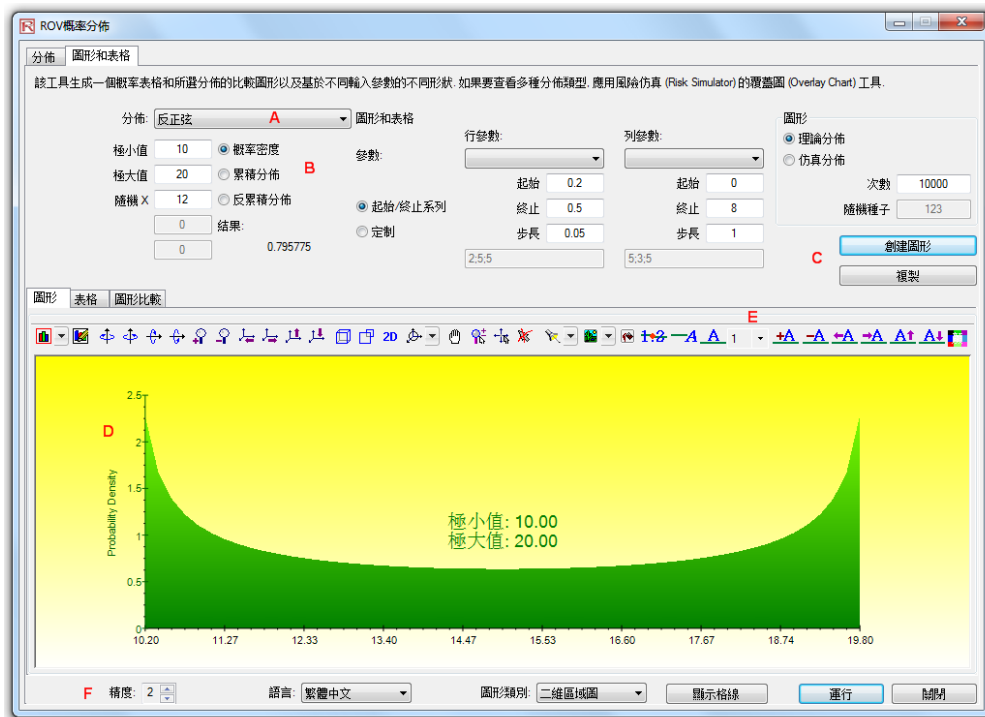


圖 5.49—ROV 概率分佈（PDF 和 CDF 圖）

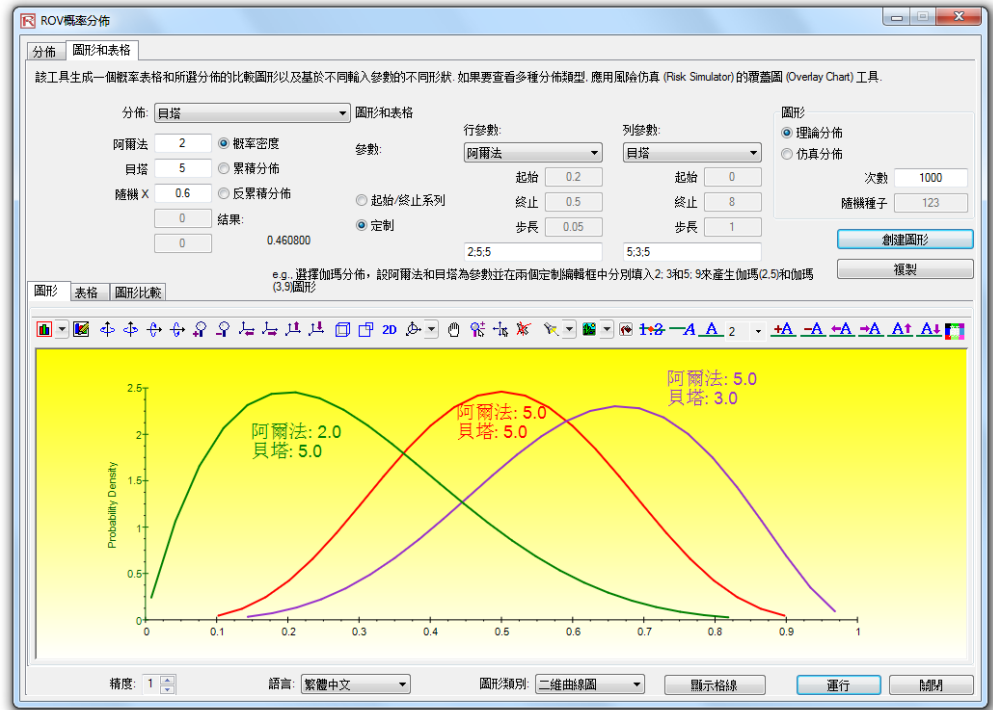


圖 5.50—ROV 概率分佈（多重重疊圖）

改變第一變量: 概率	改變第二變量: 隨機 X								類別	概率密度	
N	0.0000	1.0000	2.0000	3.0000	4.0000	5.0000	6.0000	7.0000	8.0000		
1	0.2000	0.0115	0.0576	0.1369	0.2054	0.2182	0.1746	0.1091	0.0545	0.0222	
2	0.2500	0.0032	0.0211	0.0669	0.1339	0.1897	0.2023	0.1686	0.1124	0.0609	
3	0.3000	0.0008	0.0068	0.0278	0.0716	0.1304	0.1789	0.1916	0.1643	0.1144	
4	0.3500	0.0002	0.0020	0.0100	0.0323	0.0738	0.1272	0.1712	0.1844	0.1614	
5	0.4000	0.0000	0.0005	0.0031	0.0123	0.0350	0.0746	0.1244	0.1659	0.1797	
6	0.4500	0.0000	0.0001	0.0008	0.0040	0.0139	0.0365	0.0746	0.1221	0.1623	
7	0.5000	0.0000	0.0000	0.0002	0.0011	0.0046	0.0148	0.0370	0.0739	0.1201	

圖 5.51—ROV 概率分佈（分佈表）

5.22 ROV BizStats

最新的 ROV BizStats 工具是 Risk Simulator 中非常強大和便捷的模組，用於進行商業統計和資料分析模型，它包含 130 多種商業統計和分析模型（圖 5.52-5.55）。下面提供了快速應用步驟和詳細的軟體資訊。

過程：

- 在運行 *Risk Simulator / ROV BizStats* ROV BizStats，點擊案例載入資料和模型文檔[A]或者手動輸入或者複製/粘貼資料到表格中 [D]（圖 5.52）。也可以在第一個注釋行中添加注釋或者變數名稱[C]。
- 選擇相應的模型[F]在第 2 步運行，使用示例資料登錄設置[G]，輸入相應的變數[H]。相同參數的不同變數用分號隔開，不同參數的另起一行。
- 點擊運行[I]計算結果[J]。可以在不同的欄目中查看任何相關的分析結果，圖表或者統計結果。
- 如果需要，可以提供模型的名稱在步驟 4[L]中保存為文檔。多個模型可以使用同一個文檔。現有的模型可以通過編輯或者刪除[M]，重新排列配置[N]，所有的更改都保存在 [O]一個單一文檔中以尾碼名.bizstats 結尾。

注意：

- 可以在功能表列中設置資料表的大小，資料表支援高達 1000 個變數行，1 百萬列的資料格。功能表也允許更改語言設置和資料的小數位。
- 開始，可以載入案例檔[A]包含完整的資料和預先創建好的模型[S]。可以按兩下任何一個模型在報告區得到結果[J]，有的時候會是圖表或者統計結果[T/U]。使用案例檔，可以瞭解輸入參數[H]是根據模型的描述[G]，可以創建任何自訂的模型。
- 點擊變數的標頭[D]一次選擇單個或者多個變數，按右鍵添加，刪除，複製，粘貼或者視覺化[P]選擇的變數。
- 模型也是可以通過使用命令列控制台程式輸入[V/W/X]。查看運行過程，按兩下運行模型[S]進入命令列控制台[V]。也可以重建模型或者新建模型，點擊運行命令[X]完成模型。控制台下的每行反應了模型和相關的參數。
- 整個.bizstats 文檔（當資料和多個模型已經創建和保存）可以直接通過功能表中的 XML 編輯器對 XML[Z]檔編輯。文檔中的更改可以程式化地應用，當檔被保存的時候發生作用。

提示：

- 點擊表格列的頭選擇整個表格或者變量，一旦選擇，可以鼠標右鍵自動調整適合列的大小，剪切，複製，刪除，或者粘貼數據。也可以點擊然後選擇列的頭部選擇多個變量，右鍵選擇可視化數據。

- 如果單元格包含很大的數據而無法完全顯示，點擊選擇，在單元格附近挪動鼠標，可以看到跳出的提示顯示整個值，或者簡單的調整變量列的大小（拖動數據列使之顯示更寬，雙擊數據列的邊緣自動調整大小，或者右鍵點擊數據頭部然後選擇自動擬合）。
- 使用上，下，左，右鍵在表格內移動，或者使用 Home，或者 End 鍵跳到行的最左端和最右端。也可以使用組合鍵例如：Ctrl+Home 跳到左側頂端，Ctrl+End 右側底端，Shift+Up/Down 選擇整個區域，等等。
- 可以在提示區域輸入簡短的提示。盡量使提示簡短。
- 在可視化欄嘗試點擊多個按鈕，改變試圖的外觀感（例如，翻轉，移動，放大，等等）。
- 在第 3 步模型運行完之後，複製按鈕可以用來複製結果，圖表，以及統計欄，否則的話複製的內容都會是空。
- 報告按鈕只會在第 4 步存在保存的模型或者表格中包含數據的時候生成報告，否則的話生成的報告也會是空的。運行數據輸出和生成報告的時候需要預先安裝 Microsoft Excel 和 PowerPoint。
- 在如何運行特定統計模型時可以查看示例文檔查看如何在第 1 步設置數據或者如何在第 2 步輸入參數。可以使用開始指南和模板存放數據和模型。
- 可以在語言菜單更改語言。目前軟件包含 10 種語言，更多地都會在未來加入。可是，有時候某些結果還是以英語顯示。
- 在第 2 步可以通過顯示的下拉列表，更改模型列表的顯示。可以按照首字母，內容，或者輸入的要求—注意在某些字符語言（例如，中文，日文，韓語），沒有首字母排列，所以這一項不可用。
- 軟件可以處理小數點和數字的設定（例如，一千零五十美分，可以寫作 1,000.50 1.000,50 或者 1'000,50）。可以在 ROV BizStats 菜單數據|小數設置選擇小數設置。可是，請更改計算機的區域設置到 English USA，然後在 ROV BizStats 保留默認的北美 1,000.50（這個設置用來保證能夠與 ROV BizStats 和默認示例兼容）。

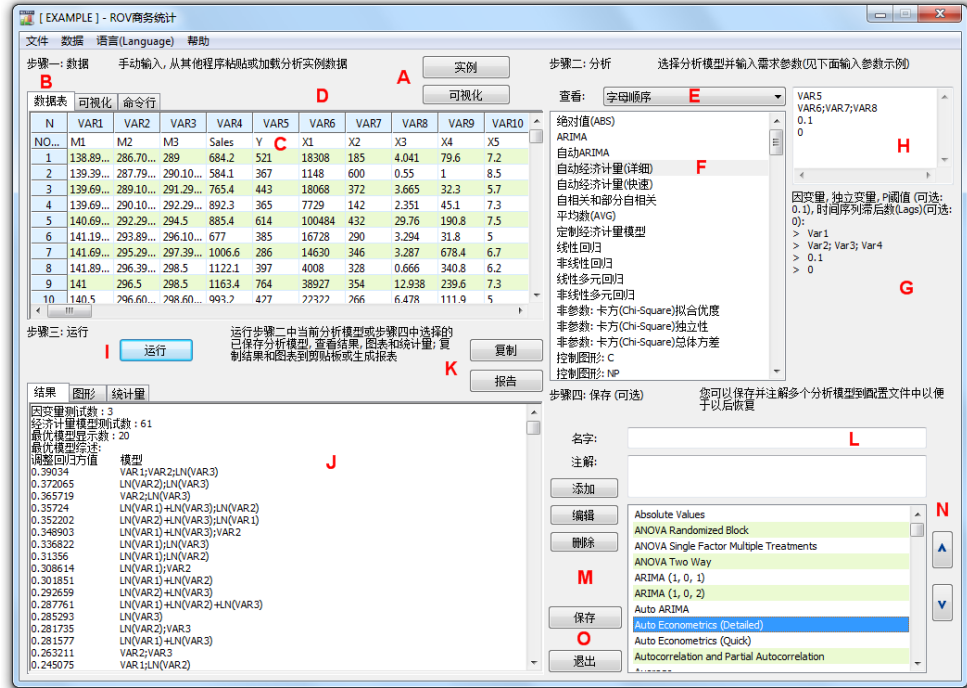


圖 5.52—ROV BizStats (統計分析)

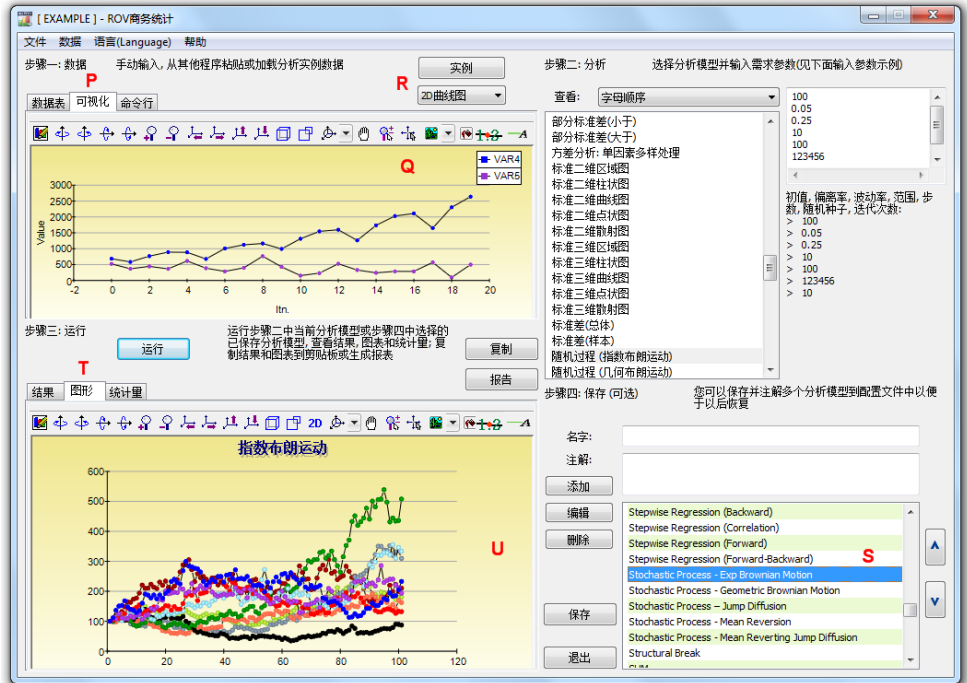


圖 5.53—ROV Bizstats (資料視覺化和結果圖)

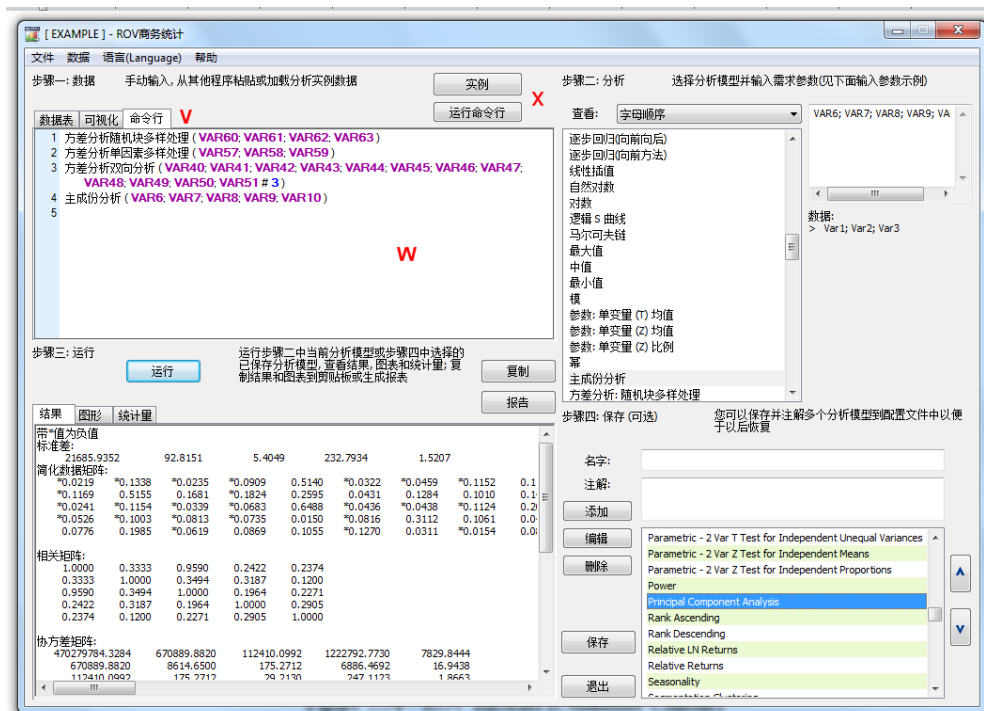


圖 5.54—ROV BizStats (命令列控制台)

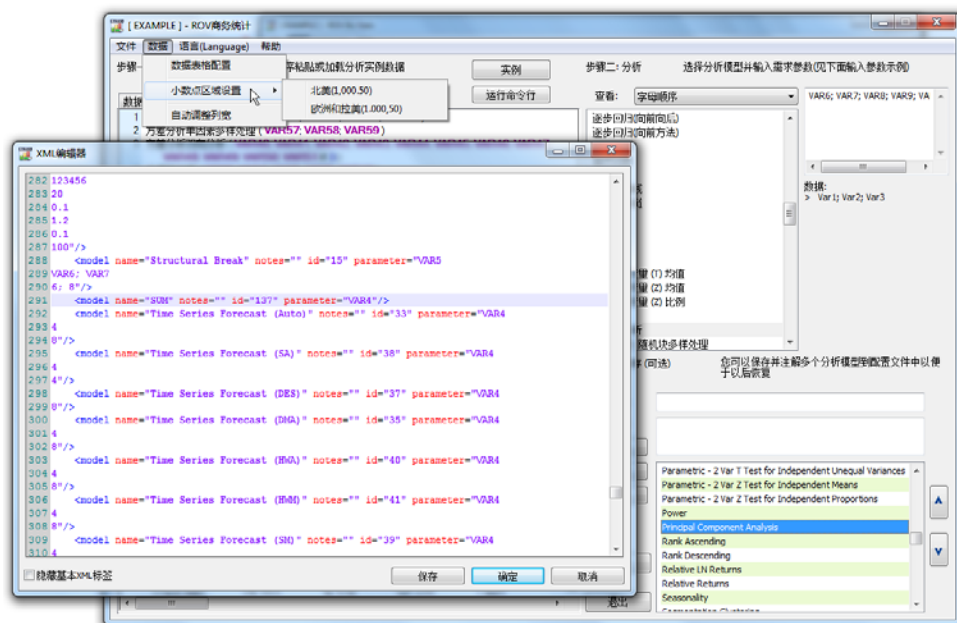


圖 5.55—ROV BizStats (XML 編輯器)

5.23 神經網路和組合模糊邏輯預測方法

神經網路一詞通常用來指一個網路或生物神經元電路，而這個名詞的現代用法通常是指在一個軟體環境中重建的由人工神經元或網路節點組成的人工神經網路。這種方法通過一種思考和辨認模式的方法試圖模仿人類大腦或神經元。在這裡，通過辨認模式來實現預測時間序列資料的目的。你可以在 Risk Simulator 裡面的 [ROV BizStats](#) 模組找到該種方法。位於 [ROV 的 BizStats 神經網路](#)，以及在 [Risk Simulator 預測神經網路](#)。圖 5.56 顯示了神經網路預測方法。

操作流程

- 點擊 Risk Simulator | 預測神經網路
- 通過手動輸入資料或從剪貼板粘貼進行啟動（例如，選擇並複製 Excel 中的一些資料，啟動這項功能，並通過點擊 [粘貼](#) 按鈕來粘貼資料）
- 選擇你希望運行的 [線性](#) 或 [非線性](#) 神經網路模型，輸入你期望的預測期間（如 5），隱藏 [層](#) 的神經網路數（如 3），和 [測試週期](#) 數（如 5）。
- 點擊 [運行](#) 以執行分析和審查計算的結果和圖表。您還可以將結果 [複製](#) 到剪貼板和圖表，並粘貼到另一個軟體應用程式中。

請注意，在網路中隱藏層的數目是一個輸入參數，並需要使用資料來校準。通常情況下，資料模式越複雜，你將需要更多隱藏層數，並且需要更多的時間來計算。建議你在 3 層開始。測試期數僅僅是用於神經網路模型所最後的校準所使用的資料點的數量，我們建議使用至少是你希望預測的期數。

與此相反，模糊邏輯這個詞是從模糊集合理論而來，用於近似而不是精確的推理，與“清脆的邏輯”相反。其中二元集合有二元邏輯，模糊邏輯變數可能有一個其範圍介於 0 和 1 之間的真實數值，而不限於經典命題邏輯的兩個真實數值。這種模糊加權特性與一個組合方法一起使用產生時間序列資料預測結果，如圖 5.57 所示的結果，這種方法應用到具有季節性和趨勢性的時間序列資料時是最適用的。你可以在 Risk Simulator 裡面的 [ROV BizStats](#) 模組來找到這種方法。位於 [ROV BizStats | 組合模糊邏輯](#) 和 [Risk Simulator | 預測 | 組合模糊邏輯](#)。圖 5.57 顯示了模糊邏輯預測方法。

操作流程

- 點擊 Risk Simulator | 預測組合模糊邏輯。
- 通過手動輸入資料或從剪貼板粘貼進行啟動（例如，選擇並複製 Excel 中的一些資料，啟動這項功能，並通過點擊 [粘貼](#) 按鈕來粘貼資料）
- 從下拉清單中選擇你希望運行分析的變數，輸入季節性週期（例如，季度資料是 4，月度資料是 12，等等），以及 [預測期數](#)（如 5）。
- 點擊 [運行](#) 以執行分析和審查計算的結果和圖表。您還可以將結果 [複製](#) 到剪貼板和圖表，並粘貼在另一個軟體應用程式中。

請注意，神經網路和模糊邏輯技術都還沒有被確定為商業預測領域有效和可靠的方法，無論是在戰略，戰術或運營層面。在這些高級預測領域，還需要進行大量持續的研究。然而，Risk Simulator 提供了這兩種技術的基本方法來進行時間序列預測。我們建議你不要孤立地使用這些技術，而與其它的風險模擬預測方法相結合，建立更強大可靠的模型。

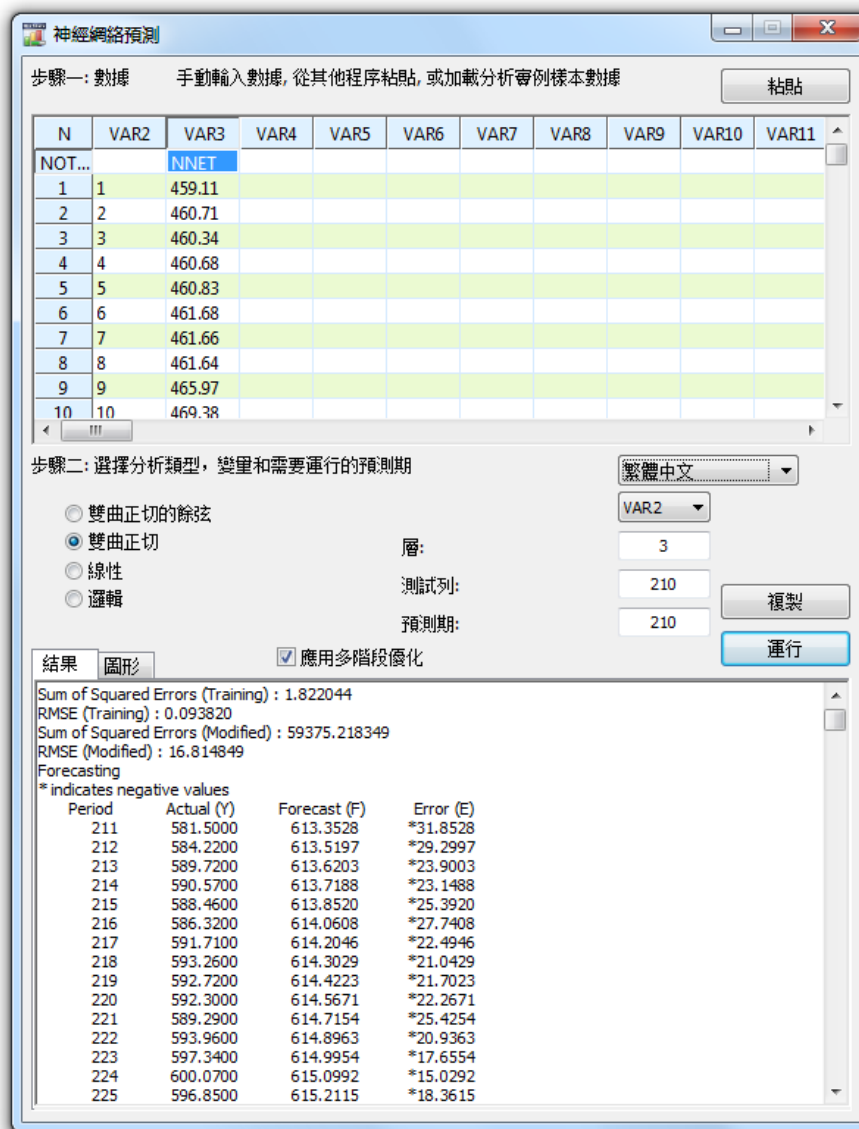


圖 5.56 - 神經網路預測

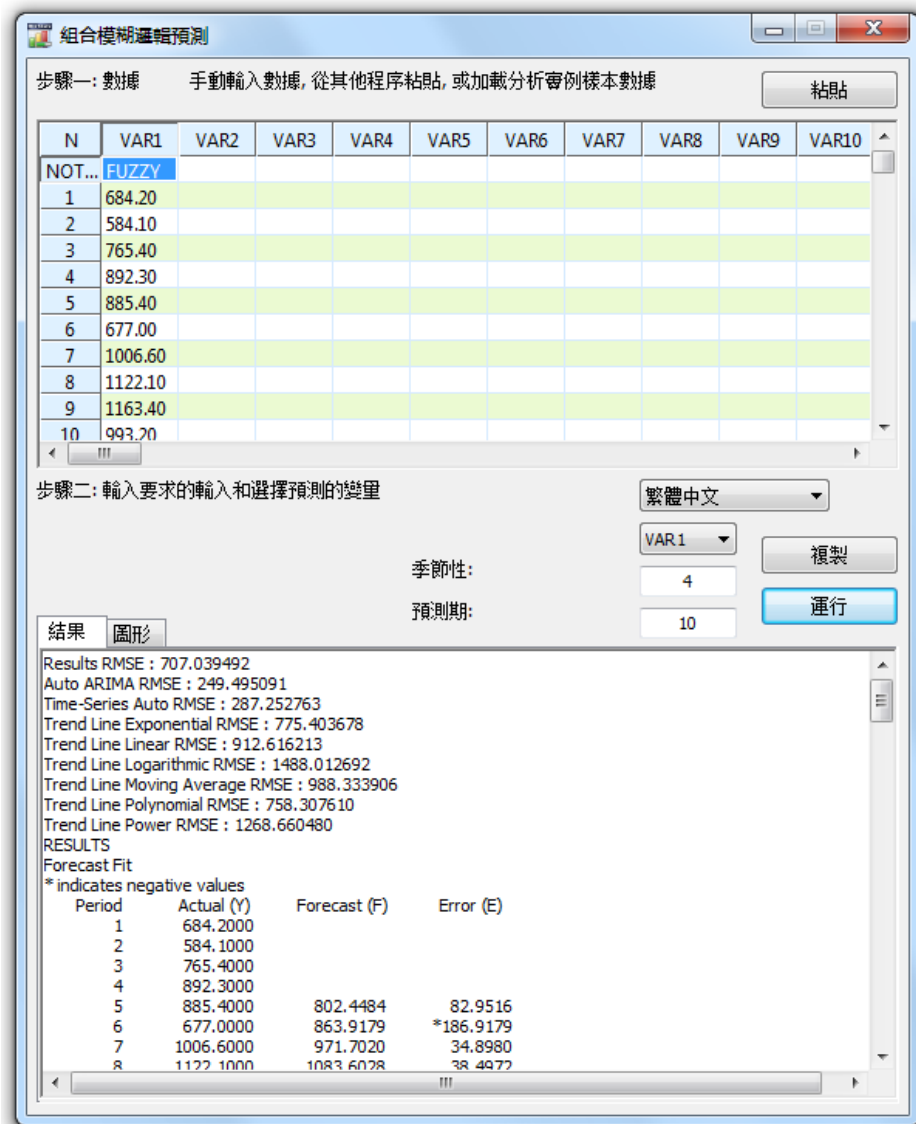


圖 5.57 - 模糊邏輯時間序列預測

5.24 優化器目標搜尋

目標搜尋工具是一個搜索演算法，用於模型來尋找了一個單變數的解。如果你知道一個公式或模型的結果，但不知道什麼樣的輸入值才能得到這一結果，可以使用 **Risk Simulator 工具 | 目標搜尋** 功能。請注意，目標搜尋功能只適用於只有一個變數輸入數值。如果你想接受多個輸入值，使用 Risk Simulator 高級的優化程式。圖 5.58 顯示了一個簡單的模型和如何運行目標搜尋功能。

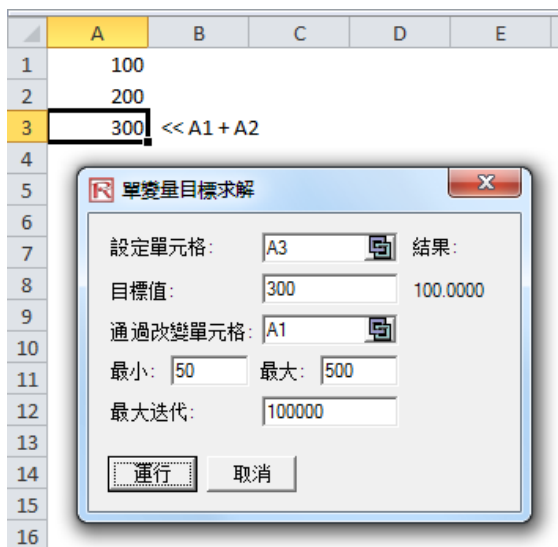


圖 5.58 - 目標搜尋

5.25 單變數優化

單變數優化工具是一個搜索演算法，應用於模型來尋找了一個單變數的解，就像前面討論的目標搜尋方法。如果你想最大化或者最小化一個模型中可能的結果，但不知道公式需要什麼樣的輸入值來得到這一結果，可以使用 [Risk Simulator 工具單變數優化](#) 功能（圖 5.59）。請注意，此單變數優化程式運行速度很快，但只適用於尋找只有一個變數的輸入。如果你想接受多個輸入值，使用 Risk Simulator 高級優化方法。請注意，此工具包括在 Risk Simulator 內，因為有時你可能需要一個快速的優化計算來解決單個決策變數問題。此工具提供了此功能而不需要設置一個模擬文檔，模擬假設，決策變數，目標和約束的優化模型。

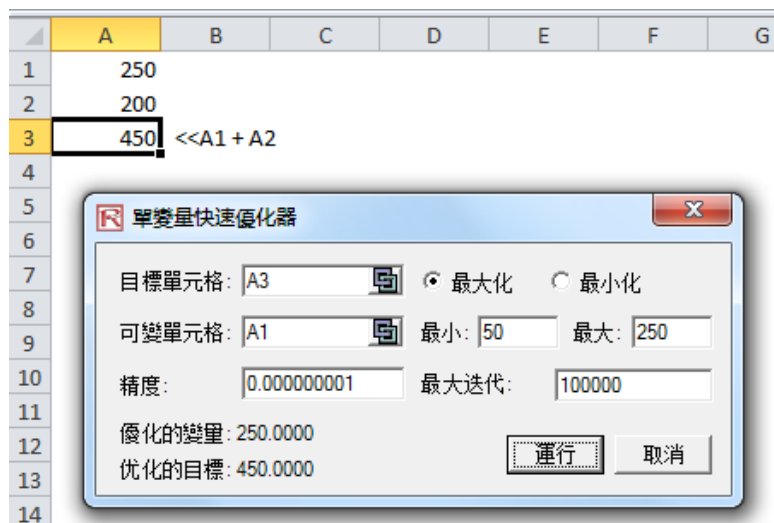


圖 5.59 - 單變數優化

5.26 遺傳演算法優化

遺傳演算法是一種啟發式模仿自然演化過程的方法。此方法一般用於獲得優化和搜索問題的有用的解。遺傳演算法屬於進化演算法的較大級別類，進化演算法為優化問題提供了解決方法，使用的技術包括自然進化，如繼承，變異，選擇和交叉。

遺傳演算法可在 [Risk Simulator工具|遺傳演算法](#) 中找到（圖 5.60）。應注意校正模型的輸入因為其結果對輸入是相當敏感（對於最常見的輸入水準，默認輸入可作為一般指導），並建議將梯度搜索測試作為一個更準確的結果集（可以取消選擇該選項就可以開始，然後選擇該選項，重新運行分析和比較結果）。

注：對於許多問題，遺傳演算法可能有一種向局部最優值收斂的傾向，甚至是任意點而不是要解決的問題的全域最優值。這意味著此方法不知道如何犧牲短期的擬合優度來獲取長期的擬合優度。對於特定的優化問題，其它的優化演算法可能提供比遺傳演算法更好的解（考慮到同樣的計算時間）。因此，建議您首先運行遺傳演算法，然後重新運行，通過檢查選擇 [應用梯度搜索測試](#)（圖 5.60），以檢查模型的穩健性。這種梯度搜索測試將嘗試運行傳統遺傳演算法優化技術與遺傳優化方法的組合，並返回最佳的解。最後，除非有特定的理論需要使用遺傳演算法，我們建議您使用 Risk Simulator 的優化模組以獲取更穩健的結果，它使您可以運行更高級的基於風險的動態和隨機優化方法。

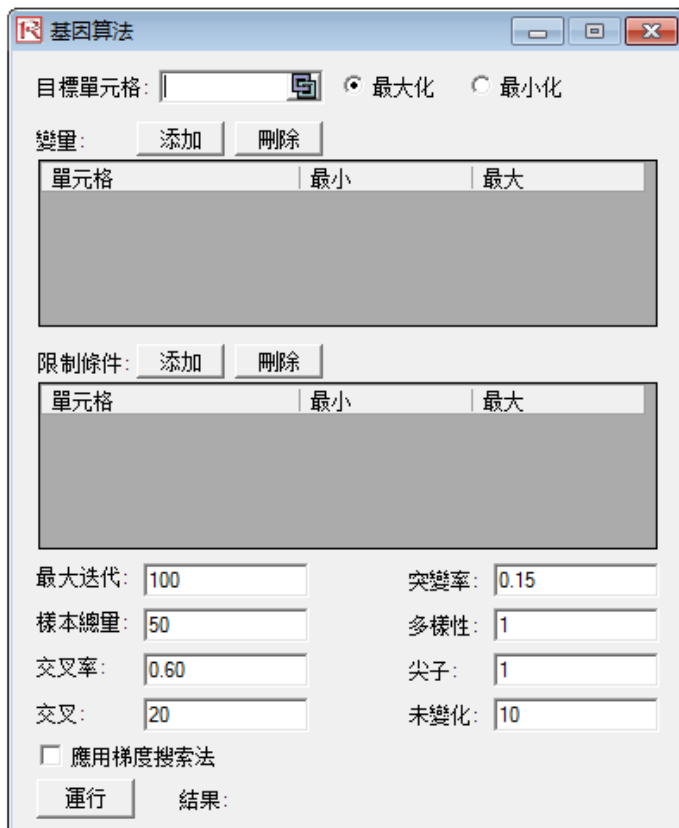


圖 5.60 - 遺傳演算法

5.27 ROV 決策樹模組

5.27.1 決策樹

ROV 決策樹軟體用於創建和評估決策樹模型。高級方法和分析包括:

- 決策樹模型
- 蒙特拉羅模擬
- 敏感性分析
- 情景分析
- 貝葉斯定理(連接及後可能性更新)
- 期望數值資訊
- 極大極小
- 極大極小
- 風險文件

決策樹模組有與戰略樹模組相同的功能(詳細參照戰略樹模組部分),具有以下其他功能(圖 5.61):

- 模組共有十一種語言，可以通過語言功能表改變。
- 選擇任意節點，*插入選項點*或者*插入終止點*，點擊選項節點圖示(方形)或者終止節點圖示(三角形)或使用功能*插入功能表*。
- 按兩下節點，修改單個*選項節點*或者*終止節點*。有時候當您點擊一個節點，所有下屬子節點也會被選擇(這樣可以移動從所選節點下的所有樹)。如果您希望僅僅選擇那個節點，您可以點擊空白背景，單獨選擇節點。此外，您可以移動單個節點或者整個樹(右鍵，或者在*編輯功能表*中，選擇*單個移動節點*或者*一起移動節點*)。
- 以下是描述可以並配置到節點屬性中。可以嘗試屬性的不同設置來判斷戰略樹中的效果:
 - *名稱*. 節點上方顯示名稱
 - *數值*. 節點下方顯示數值
 - *Excel 連接*. 從 Excel 儲存格中連接資料
 - *注釋*. 可以在節點上方或下方插入節點
 - *模型展示*. 名稱，數值，注釋的組合
 - *本地顏色與全部顏色*. 節點顏色可以改變

- *標籤內形狀*. 文本可以放置於節點中(需要將節點擴大些)
- *子事件名*. 文本可以放置在引導節點提示事件。
- *選擇實物期權*. 實物期權模型可以指派到當前節點。指派節點可以生成一系列輸入變數。
- 全部元素是可定制化的，包括戰略樹的背景，連接線，選擇節點，終止節點，文本盒。 具有以下設置項:
 - *字體* 名字，數值，注釋，標籤，項目名
 - *節點大小*(最小和最大的高度和寬度)
 - *邊界*(線類型，寬度，顏色)
 - *陰影* (顏色和是否應用陰影)
 - 全部顏色
 - 全部形狀
- *編輯* 功能表中的 *查找資料要求視窗*能夠在戰略樹右側打開一個視窗，當一個選擇節點或者終止節點被選擇時候，能夠展示節點的屬性，並直接更新。這個功能能夠每次在節點上按兩下。
- *案例檔*能夠通過 *檔*功能表開始構建決策樹。
- *檔 保護檔*可以將戰略樹加密到 256 個位元組密碼。在加密的時候，請牢記密碼。
- *檔 截屏*可以截屏現有模型。圖片可以在其他軟體中使用。
- 添加, 複製, 重命名, 刪除決策樹 可以通過右鍵決策樹或者編輯 功能表。
- 您也可以任意選項或者終止節點*插入檔連結* 或者 *插入評論*，在背景處*插入文本* 或者 *插入圖片*。
- 您也可以改變現有類型，*管理和創建自訂的戰略樹*包括尺寸，形狀，顏色，字體大小，整個決策樹的顏色)。
- *插入決定因素*, *插入不確定性*, *插入終止*，選擇已有節點，點擊決策節點圖示(方形)，不確定性節點圖示(圓形)，終止節點圖示(三角形)，在*插入*功能表中這些功能。
- 按兩下節點來修改單個決策，不確定性，終止節點屬性。以下是決策樹模組中的一些獨特項目，可以定制化並在節點屬性使用者介面中配置。
 - 決策節點: *定制傭金*或者*自動計算*節點的值。默認為自動計算。當在完整決策樹模型上點擊*運行*時，決策樹節點會隨著結果更新。
 - 不確定性節點: *事件名*, *可能性*, *設置模擬假設*。您可以在不確定分支創建之後，添加可能性事件名，可能性，模擬假設。

- 終結點: 人工輸入, *Excel 連接*, 設置模擬假設。最終事件結果可以手動輸入或者連接到一個 Excel 儲存格(例如, 如果有 Excel 模型計算結果, 可以將模型與 Excel 模型輸出儲存格連接), 或者設運行模擬來 設置可能性分佈假設。
- *編輯 流覽節點屬性視窗* 節點被選擇時。節點屬性會更新
- 決策樹模組有以下高級功能:
 - 決策樹下的蒙特拉羅模型
 - 得到次可能性的貝葉斯分析
 - 完全資訊的預測值, 最大最小分析, 風險檔, 不完整資訊的值
 - 敏感性分析
 - 情景分析
 - 效用函數分析

5.27.2 模擬模型

這個工具可以在決策樹上運行蒙特卡洛分析 (圖 5.62)。可以設置可能性分佈作為輸入假設來運行模擬。您可以為選擇的節點設置假設, 也可以用新假設來設置假設, (或者使用過去創建的假設) 在一個數量等式或者公式中。例如, 您可以設置一個新假設叫做正態分佈(平均值為 100, 標準差為 10), 在決策樹中運行假設, 在等式(100*正態+15.25)中使用假設。在一個數字表達盒中創建自己的模型。您也可以使用基本運算或者在等式中添加變數, 按兩下變數清單。新的變數可以添加進清單, 可以是數位運算式, 也可以是假設。

5.27.3 貝葉斯分析

貝葉斯定理分析工具(圖 DT3)可以用於任意連在一條路徑上的兩個不確定事件, 在右邊的案例中(圖 5.63), 不確定的 A 和 B 連在一起, 事件 A 先發生, 事件 B 後發生。事件 A 是市場研究, 有兩個輸出(有利的或者不利的)。事件 B 是市場條件, 同樣有兩個輸出 (強的和弱的)。工具用於計算連接, 臨界, 通過可能性和依賴條件來確定貝葉斯後可能性, 依賴可能性可以通過兩個後更新條件可能來計算。選擇如下相關分析點擊載入案例, 瞭解案例輸入及哪些結果用於決策樹中的輸入。

快速流程

- 第一步: 輸入第一個和第二個未確定事件的名字, 選擇每個事件的可能性數量。
- 第二步: 輸入可能性事件的名稱或輸出
- 第三步: 輸入第二個事件的前期可能性, 每個事件或者輸出的條件, 可能性的和為 100%

5.27.4 精確資訊的預計數值，最大最小分析，風險檔，不精確資訊的數值。

工具(圖 5.64)計算精確資訊(EVPI)的的預計數值，最大最小分析，風險檔，不精確資訊的數值。開始前，輸入決策分支的數量或決策(例如：構建大型的，中型的，小型)，不確定事件的數量，自然輸出的狀態 (例如., 好市場，差市場), 在每種情景下的預測結果。

精確資訊的預測數值 (EVPI), 假設您有預見性，知道需要做什麼(通過市場研究或其他方法來更好地識別概率結果), 計算資訊中是否有增值(例如，市場研究是否會增值) 來比較更多簡單的概率狀態預測。輸入決策分支或者決策(例如:構建大型的，中型的，小型) 以及未知事件或者輸出狀態的數量 (例如好市場，差市場), 在每種情景下的預測結果。

最大最小 (最小化最大遺憾) 以及最大最小(最大化最小結果) 是找到優化決策路徑的兩種方法。這兩種方法並不常用，但依然可以提供為決策制定流程提供參考。輸入決策分支的數量或者存在的路徑 (例如: 構建大型的，中型的，小型的), 或者不確定性的事件或者每個路徑的狀態 (例如：好經濟，壞經濟). 然後，完成不同情景下結果表並計算最大及最小結果.您也可以點擊登錄案例來查找一個計算。

5.27.5 敏感性分析

敏感性分析 (圖 5.65)用來決定輸入對於不同決策路徑數值的影響。首先，選擇一個決策節點進行分析，然後從列表中選擇一個可能性事件來測試。當不同可能性事件具有同一可能性時，能夠分析出是獨立還是同時發生。

敏感性圖展示了在不同可能性程度下，決策路徑的價值。結果表中展示了數值。交叉線的位置，代表了概率事件一個決策路徑變成主導。

5.27.6 情景表

情景表 (圖 5.66) 能夠在輸入變動時，生成決定輸出數值。您可以選擇一個或多個決策路徑來分析(每個選擇路徑的結果將表示成一個單獨的表或者圖)，一二個不確定或者終止節點作為情景表的輸入變數。

快速流程

- 從下列表中選擇一個或多個決策路徑來分析
- 選擇一個或者兩個不確定事件或終止結果來模型
- 決定是否改變事情的可能性，在同一可能性事件下
- 輸入情景區間

5.27.7 生成效用函數

效用函數(圖 5.67), 或者 $U(x)$, 用於代替決策樹中預期終止結果的預測值。 $U(x)$ 可以有兩種方法產生: 使用繁瑣和詳細的實驗的每一個可能的結果或指數外推方法(用於這裡)。他們可以被模擬為一個風險厭惡的決策者(負面影響比潛在下降更具有災難性), 風險中性(兩面有平等的吸引力), 或風險喜好(更希望具有上漲潛力)。輸入最大或最小值的終端結果, 一些資料點之間計算完整曲線表。

如果你有一個 50:50 的賭博, 你可以賺 X 美元或失去 $X/2$ 美元, 或者不玩得得到 0 美元的報酬, X 應該是什麼? 舉個例子, 如果打賭, 你或者可以贏得 100 美元或失去 50 美元, 或者完全不玩, 那麼 X 就是 100 美元。在正收益的下面輸入 X 。請注意, X 越大, 風險平均的可能性就越小, 一個小的 X 代表你更偏風險性。

輸入所需的輸入, 選擇 $U(x)$ 類型, 然後按一下計算函數獲得結果。你也可以使用計算出的 (x) 值來重新運算, 或恢復樹到使用預期價值的成果。

Risk Simulator

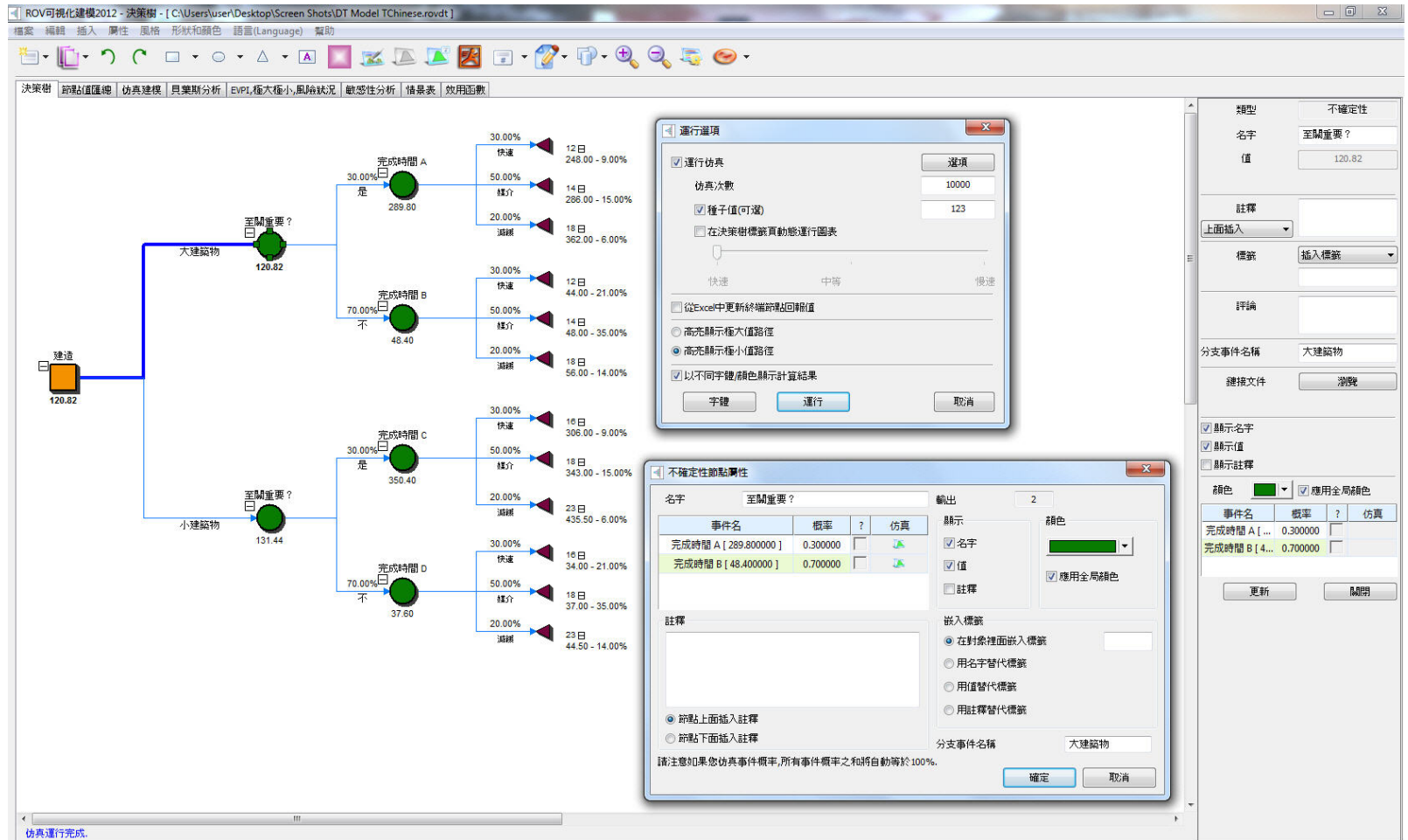


圖 5.61 - ROV 決策樹(決策樹)

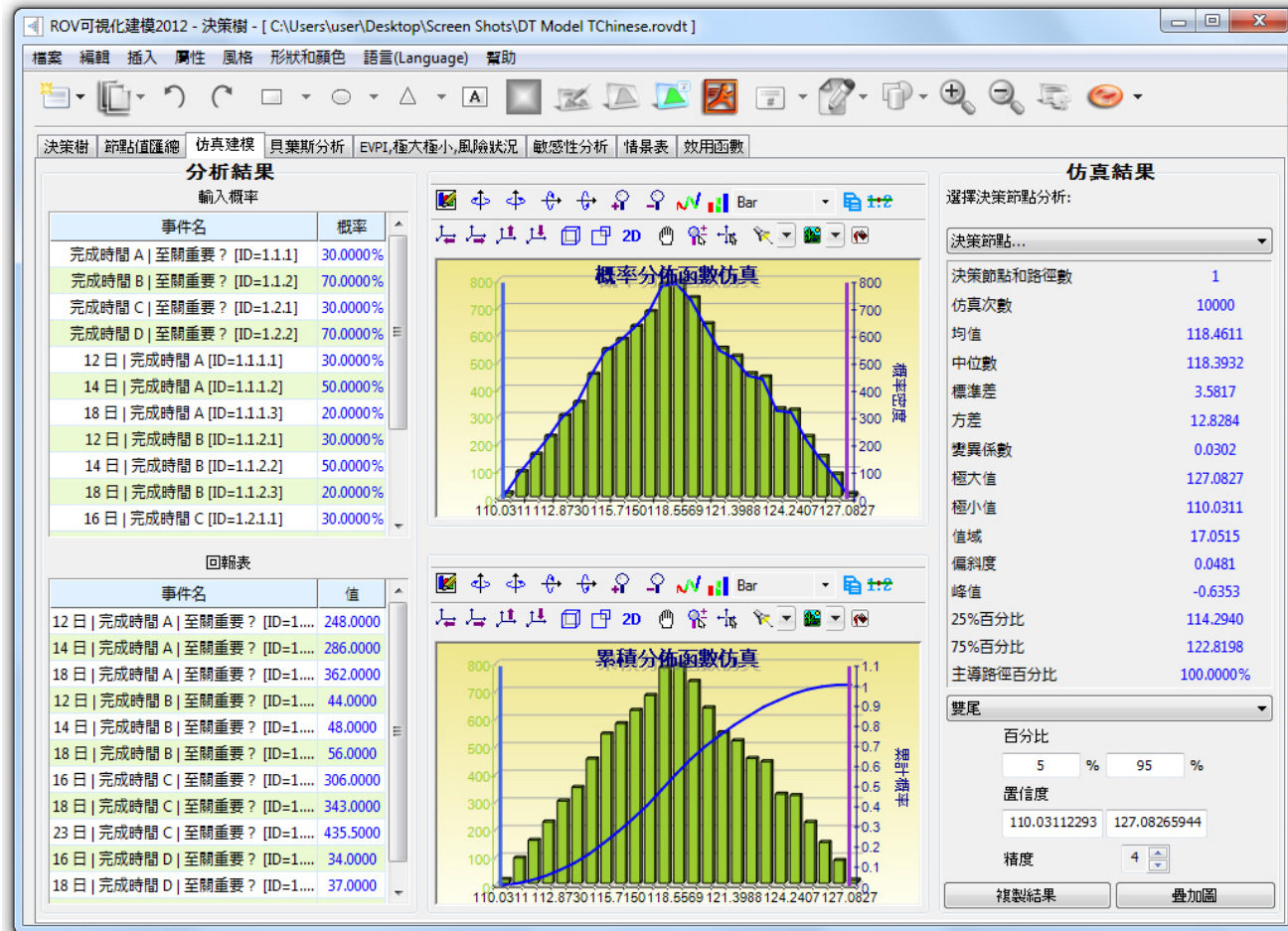


圖 5.62 - ROV 決策樹 (模擬結果)

ROV可視化建模2012 - 決策樹 - [C:\Users\user\Desktop\Screen Shots\DT Model Korean.rovdt]

檔案 編輯 插入 屬性 風格 形狀和顏色 語言(Language) 幫助

決策樹 節點值匯總 仿真建模 貝葉斯分析 EVPI,極大極小,風險狀況 敏感性分析 情景表 效用函數

貝葉斯分析工具可以執行任何兩個沿路徑連接的不確定性事件。例如，在右邊的例子中，不確定性 A 和 B 連接，時間序列上事件 A 先發生，事件 B 後發生。第一事件 A 是 Market Research (兩個輸出 Favorable 或 Unfavorable)。第二事件 B 是市場條件 (也是兩個輸出 Strong 和 Weak)。該工具通過輸入先驗概率和可靠條件概率來計算聯合、邊緣、和貝葉斯後驗更新概率；或輸入後驗條件概率計算可靠性概率。選擇以下相關分析並點擊加載示例來查看相對應的樣例輸入和右邊表格中顯示的結果，還有哪個結果是用來作為在決策樹中的輸入。

給定先驗和可靠條件概率來計算貝葉斯更新後驗概率 (更多使用)

給定先驗和後驗概率來計算可靠聯合概率 (較少使用)

步驟一：輸入第一和第二個不確定性事件的名稱和選擇每個事件有多少概率事件 (自然狀態或結果)。

第一事件名稱: Market Research 概率事件或狀態: 2

第二事件名稱: Market Conditions 概率事件或狀態: 2

步驟二：輸入每個概率事件或結果的名字。

狀態	Market Research	Market Conditions
1	Favorable	Strong
2	Unfavorable	Weak

加載示例 計算

步驟三：輸入第二事件的先驗概率和每個事件或結果的條件概率。概率之和應該等於 100%。

條件概率 (可靠性)				
事件	先驗 P(x)	Favorable	Unfavorable	總計
Strong	45.00%	60.00%	40.00%	100.00%
Weak	55.00%	30.00%	70.00%	100.00%
總計	100.00%			

保存的模型 名字: [] 添加 刪除

市場研究可靠性:
60% (Favorable given Strong)
40% (Unfavorable given Strong)
30% (Favorable given Weak)
70% (Unfavorable given Weak)

先驗概率和可靠條件概率

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

聯合和邊緣概率

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

後驗或更新概率

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

貝葉斯分析結果

圖 5.63 - ROV 決策樹(貝葉斯分析)

ROV可視化建模2012 - 決策樹 - [C:\Users\user\Desktop\Screen Shots\DT Model TChinese.rodvt]

檔案 編輯 插入 屬性 風格 形狀和顏色 語言(Language) 幫助

決策樹 節點值匯總 仿真建模 貝葉斯分析 EVPI,極大極小,風險狀況 敏感性分析 情景表 效用函數

完善信息的預期值, 最小極大和最大極小分析, 風險配置, 和不完全信息值

該工具計算完善信息的預期值(EVPI), 最小極大和最大極小分析, 和風險配置以及不完全信息值。輸入決策分支或考慮的戰略數(如, 構建一個大, 中, 小設施)和不確定性事件數或結果的自然狀態(如, 暢銷, 滯銷), 和輸入各種情況下的預期回報開始。

輸入假設

決策分支: 3

不確定事件或狀態: 2

概率	狀態 1	狀態 2	總計
	80%	20%	100%

回報	狀態 1	狀態 2	均值
決策路徑 1	8	7	7.80
決策路徑 2	14	5	12.20
決策路徑 3	20	-9	14.20
極大值	20.00	7.00	

加載示例

計算

最最小極大和最大極小分析

最小極大(極大值中的最小值)和最大極小(極小回報值中的最大值)是兩種交替的找出最佳決策路徑的方法。這兩種方法不經常使用但是仍然提供以更深入理解決策過程。輸入決策分支數或存在的路徑(如, 構建一個大, 中, 小設施), 和每條路徑裡的不確定性事件數或結果的自然狀態(如, 經濟良好 vs. 經濟不景氣), 完成不同情況的回報表和計算最小極大和最大極小值。您還可以點擊加載示例來查看樣例計算。

回報	狀態 1	狀態 2	極小值
決策路徑 1	8	7	7.00
決策路徑 2	14	5	5.00
決策路徑 3	20	-9	-9.00

遺憾	狀態 1	狀態 2	極大值
決策路徑 1	12.00	0.00	12.00
決策路徑 2	6.00	2.00	6.00
決策路徑 3	0.00	16.00	16.00

最小極大: 6.00 路徑 2 是最優的

最大極小: 7.00 路徑 1 是最優的

風險配置

戰略 1 風險配置

回報	概率
248.00	9.00%
286.00	15.00%
362.00	6.00%
44.00	21.00%
48.00	35.00%
56.00	14.00%
概率總計	100.00%
預期值	120.82

戰略 2 風險配置

回報	概率
306.00	9.00%
343.00	15.00%
425.00	6.00%

不完善信息的預期值: 10.62

保存的模型

名字:

添加

刪除

完善信息的預期值

完善信息的預期值(EVPI), 如, 假設您有玩美遠見和精確制知道做什麼(通過市場調研或其它更好的理解概率結果), 相對於更不成熟的概率自然狀態的估計, EVPI計算如果在這種信息裡有添加值(0, 如果市場調研將添加值), 輸入決策分支或考慮的戰略數(如, 建立大, 中, 小設施)和不確定性事件數或結果的自然狀態(如, 暢銷, 滯銷), 和輸入各種情況下的預期回報開始。

自然狀態的完善信息的預期值: 17.40

無自然狀態完善信息的預期值: 14.20

完善信息的預期值: 3.20

圖 5.64 - ROV 決策樹 (EVPI, 最大最小, 風險文件)

ROV可視化建模2012 - 決策樹 - [C:\Users\user\Desktop\Screen Shots\DT Model TChinese.rovdt]

檔案 編輯 插入 屬性 風格 形狀和顏色 語言(Language) 幫助

決策樹 節點值匯總 仿真建模 貝葉斯分析 EVPI,極大極小,風險狀況 敏感性分析 情景表 效用函數

輸入概率敏感性分析是確定決策路徑值的影響。首先,選擇決策節點分析,其次從列表中選擇概率事件測試。如果多個不確定事件有相同概率,它們可以獨立或共同分析。 敏感性圖表顯示在不同概率點的決策路徑值。結果表顯示數值。如果有交叉線,交叉點表示在該概率事件某一決策路徑相對其他路徑成為主導。

步驟 1: 從列表中選擇一個或多個決策路徑分析。
 步驟 2: 選擇一個不確定事件(自然狀態)或一個終端節點回報值建模。
 步驟 3: 決定是否要改變概率或回報值自身或同時改變所有相同概率/回報值。
 ● 組合分析概率/回報值
 ○ 分析單個概率/回報值
 如果執行組合分析,檢查組成員,選擇其他組成員,或取消其他事件。
 自動選擇組成員

不確定節點和終端節點

節點和編號	概率
至關重要? [1.1]	
<input type="checkbox"/> 完成時間 A	30.00%
<input type="checkbox"/> 完成時間 B	70.00%
至關重要? [1.2]	
<input type="checkbox"/> 完成時間 C	30.00%
<input type="checkbox"/> 完成時間 D	70.00%
完成時間 A [1.1.1]	
<input type="checkbox"/> 12 日	30.00%
<input type="checkbox"/> 14 日	50.00%
<input checked="" type="checkbox"/> 18 日	20.00%
完成時間 B [1.1.2]	
<input type="checkbox"/> 12 日	30.00%
<input type="checkbox"/> 14 日	50.00%
<input checked="" type="checkbox"/> 18 日	20.00%
完成時間 C [1.2.1]	

保存的模型
 名字 Model 1
 添加
 刪除

步驟 4: 輸入敏感性範圍
 不確定性概率 從 0.00% 到 100.00% 步長 5.00%
 終端回報值 從 [] 到 [] 步長 []
 計算

概率	0.00%	5.00%	10.00%	15.00%	20.00%
至關重要?	122.20	124.18	126.16	128.14	130.12

敏感性圖表

圖 5.65 - ROV 決策樹 (敏感性分析)

ROV可視化建模2012 - 決策樹 - [C:\Users\user\Desktop\Screen Shots\DT Model TChinese.rovdt]

檔案 編輯 插入 屬性 風格 形狀和顏色 語言(Language) 幫助

決策樹 節點值匯總 仿真建模 貝葉斯分析 EVPI,極大極小,風險狀況 敏感性分析 情景表 效用函數

給定輸入質的改變決定輸入值來生成情景表。您可以選擇一個或多個決策路徑(選擇的每條路徑的結果代表一個單獨的表格和圖表),和一個或多個不確定或終端節點作為輸入變量來分析生產情景表。

步驟 1: 從下列表中選擇一個或多個決策路徑來分析。
 步驟 2: 選擇一個或多個不確定事件或終端回報值來建模。
 步驟 3: 決定是否改變事件概率或同時改變所有相同概率事件。

組合分析輸入量
 分析單個輸入量

如果執行組合分析, 檢查組成員, 選擇其他組成員, 或取消其他事件。

自動選擇組成員

不確定節點和終端節點

節點和編號	概率
至關重要? [1.1]	
<input type="checkbox"/> 完成時間 A	30.00%
<input type="checkbox"/> 完成時間 B	70.00%
至關重要? [1.2]	
<input type="checkbox"/> 完成時間 C	30.00%
<input type="checkbox"/> 完成時間 D	70.00%
完成時間 A [1.1.1]	
<input type="checkbox"/> 12 日	30.00%
<input type="checkbox"/> 14 日	50.00%
<input checked="" type="checkbox"/> 18 日	20.00%
完成時間 B [1.1.2]	
<input type="checkbox"/> 12 日	30.00%
<input type="checkbox"/> 14 日	50.00%
<input checked="" type="checkbox"/> 18 日	20.00%
完成時間 C [1.2.1]	

決策節點

節點和編號	值
建造 [1]	
<input type="checkbox"/> 至關重要?	120.82
<input checked="" type="checkbox"/> 至關重要?	131.44

Probability [20.00%] 從 5.00% 到 50.00% 步長 5.00%
 Payoff [44.50] 從 35.00 到 50.00 步長 2.50

計算 以行建圖 以列建圖 至關重要? [131.44]

	35.00	37.50	40.00	42.50	45.00	47.50	50.00
5.00%	123.85	123.93	124.02	124.11	124.20	124.28	124.37
10.00%	125.50	125.67	125.85	126.02	126.20	126.37	126.55
15.00%	127.14	127.40	127.67	127.93	128.19	128.45	128.72
20.00%	128.79	129.14	129.49	129.84	130.19	130.54	130.89
25.00%	130.44	130.88	131.31	131.75	132.19	132.63	133.06
30.00%	132.08	132.61	133.13	133.66	134.18	134.71	135.23
35.00%	133.73	134.35	134.96	135.57	136.18	136.80	137.41
40.00%	135.38	136.08	136.78	137.48	138.18	138.88	139.58
45.00%	137.03	137.82	138.60	139.39	140.18	140.97	141.75
50.00%	138.68	139.55	140.43	141.30	142.18	143.05	143.93

至關重要? [131.44]

保存的模型

名字 Model 1

Model 1	添加	刪除
Model 1		

圖 5.66 - ROV 決策樹 (情景表格)

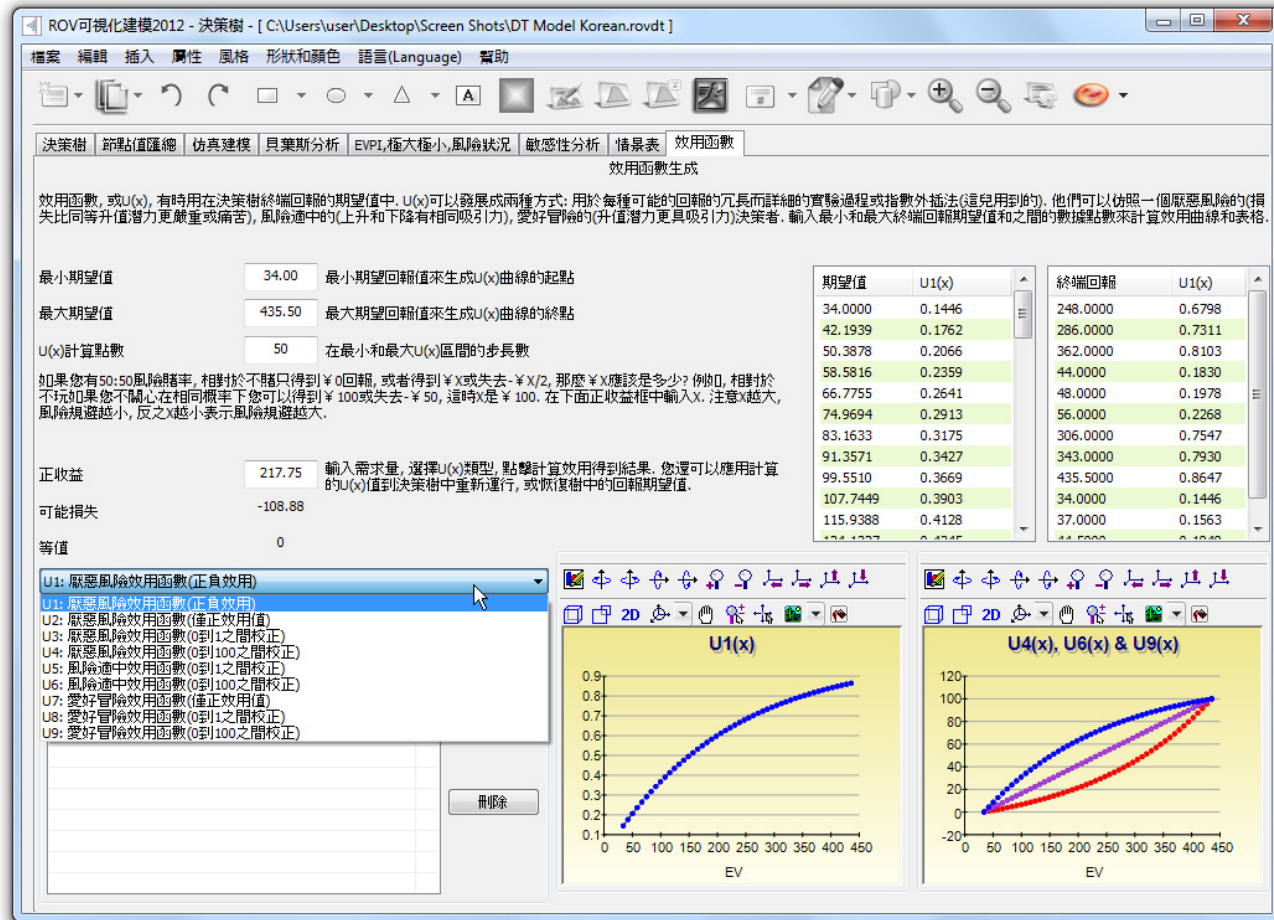


圖 5.67 - ROV 決策樹 (效用功能)



6. 幫助貼士和技巧

下面是對於Risk Simulator高級使用者的一些快速指南和簡要的技巧。對於某些特定工具的詳細資訊，請參考使用者手冊的相關部分。

貼士：假設（設定輸入假設使用者介面）

- 快速跳躍—輸入任何字母選擇某個分佈，將會直接跳到以該字幕開頭的分佈中（例如，點擊正態分佈按 W 鍵，就會直接跳到 Weibull 分佈）。
- 右擊查看—選擇任何分佈，右擊選擇不同的分佈形態（大圖示，小圖示，清單形式）。
- 更新圖表欄—再輸入新的輸入參數（例如，輸入了新的均值或者標準差），點擊 Tab 鍵或者點擊使用者介面任意位置自動查看圖表更新。
- 輸入相關性—可以直接輸入兩兩相關性（列會根據需要重新調整大小），使用多重分佈擬合工具，以及計算兩兩相關係數，或者設定假設，使用相關性編輯器輸入相關矩陣。
- 假設儲存格函數—只有空儲存格或者包含靜態值的儲存格可以設定為假設；但是，有時候函數或者公式要求作為假設儲存格，可以首先在儲存格輸入輸入假設，然後輸入公式或者函數（當模擬運行的時候，模擬的值將取代函數，在模擬完成之後，公式和函數會再次展示）。

貼士：複製和粘貼

- 使用 Esc 進行複製和粘貼—當選擇一個儲存格，使用 Risk Simulator 的複製功能，就會將所有的資訊複製到 Windows 的剪貼板，包括儲存格的數值，函數，方程，顏色，字體，大小，包括 Risk Simulator 的假設，預測，或者決策變數。當應用 Risk Simulator 的粘貼功能，會有兩個選項。第一個選項直接應用 Risk Simulator 的粘貼功能，所有的儲存格值，顏色，字體，函數和公式以及參數都將粘貼到新儲存格中。第二個選項首先點擊 Esc 鍵，然後應用 Risk Simulator 的粘貼，Esc 會告訴 Risk Simulator 只想粘貼假設，預測或者決策變數，而非儲存格的值，顏色，公式，函數，字體等等。在點擊 Esc 之前，可以保持目標儲存格的值和計算，Risk Simulator 的參數只被粘貼過去。

- 在多重儲存格中複製和粘貼—選擇多個需要複製和粘貼的儲存格（包括連續和非連續的假設）。

貼士：相關性

- 設置假設—使用輸入假設對話方塊設置相關性（理想情況只可以輸入定量的相關性）。
- 編輯相關性—手動輸入或者從 Windows 剪貼板粘貼輸入相關性矩陣。
- 多分佈擬合—自動計算和輸入配對相關性（當進行多個變數擬合時，自動計算相關性，並且決定哪些具有顯著相關性）。

貼士：資料診斷和統計分析

- 隨機參數估計—在統計分析和資料爭端報告中，有一欄關於隨機參數估計，基於歷史資料可以計算波動率，漂移率，均值恢復率，跳躍擴散率。注意，這些參數結果都是基於歷史資料的，參數會隨著時間改變而改變。而且，分析的結果並不表明哪個模型是擬合最好的（例如，布朗運動，均值恢復，跳躍擴散，或者混合過程）。這個由使用者自己決定根據時間變數進行預測。只允許客戶自己決定使用什麼過程（例如，布朗運動用於模擬股價波動，但是分析工具並不知道變數的性質）。最後，在這裡提示，某個參數在正常範圍之外，該過程要求的輸入參數可能並不是正確的過程（例如，如果均值恢復率是 110%，均值恢復過程就不是適合的過程，如此類推）。

貼士：分佈分析，圖表

- 分佈分析—快速計算 Risk Simulator 中 42 種分佈的 PDF, CDF 和 ICDF，並且以表格形式返回這些值。
- 分佈圖表—用於比較相同分佈的不同參數（例如，具有不同 Alpha，Beta 參數[2,2]，[3,5]，[3.5,8]的 Weibull 分佈的 PDF，CDF，ICDF 的形狀，並且將彼此重疊到一起）。
- 重疊圖—用於比較不同的分佈（理論輸入假設和試驗模擬輸出預測）並且將彼此重疊比較。

貼士：有效邊際

- 有限邊際變數—可以查看有效邊際變數，首先在設定有效邊際邊際設定模型限制。

貼士：預測儲存格

- 不包含公式的預測儲存格—可以在不包含任何公式或者值的儲存格中設置預測（只需要跳過警告資訊）但是預測圖表將不包含任何資訊。空值的輸出預測儲存格一般用於包含宏的儲存格。

貼士：預測圖表

- *Tab* 和 *空格*—輸入一些參數後敲擊 *Tab* 更新預測圖表或者百分位元數和置信值，敲擊空白鍵在預測圖表的不同欄目中切換。
- 常規視圖和全域視圖—點擊這些視圖在常規視圖和全域視圖直接切換，全域視圖可以一次性顯示所有的圖表。
- 複製—複製的預測圖表取決於當前是常規視圖還是全域視圖。

貼士：預測

- 儲存格連結位址—如果選擇電子報個中的資料，然後運行預測工具，所選資料的儲存格位址會自動輸入到使用者介面，否則的話需要手動輸入儲存格位址或者使用連結圖示連結相關資料位址。
- 預測 RMSE—對多個預測模型使用誤差估計，比較每個模型的精度。

貼士：預測：ARIMA

- 預測期—外生變數資料必須至少超過預測期數（例如，如果想用 100 個時間序列點預測未來 5 期，那外生變數至少需要 105 個），否則的話運行不包含外生變數的 ARIMA，就可以預測沒有任何限制的期間。

貼士：預測：基本計量經濟學

- 變數通過分號隔開—使用分號隔開引數。

貼士：預測：Logit，Probit 以及 Tobit

- 資料要求—運行 logit 和 probit 檢驗的因變數必須是二元的（0 和 1），而 tobit 模型可以接受二元和其他浮點數值。引數對於任何三種模型可取任何值。

貼士：預測：隨機過程

- 預設樣本輸入—使用預設樣本輸入作為模型的起始點。
- 用於參數估計的統計分析工具—使用該工具通過估計原始資料校正

隨機模型的輸入參數。

- 隨機過程模型—有時候如果隨機過程的使用者介面顯示了較長的時間，可能是輸入不正確或者對應模型指定不對（例如，如果均值恢復率是 110%，可能均值恢復過程不是正確的過程，等等）。請使用其他不同的輸入或者使用不同的模型。

貼士：預測：趨勢線

- 預測結果—拖卷報告的結尾部分查看預測的值。

貼士：函式呼叫

- RS 函數—可以讀取假設變數得到預測統計函數。得到這些函數，首先安裝 RS 函數（開始功能表，所有程式，Real Options Valuation，Risk Simulator，工具，安裝函數）然後在設定 RS 函數之前運行模擬。引用示例模型 24 作為案例指導如何使用這些函數。

貼士：入門練習和視頻

- 入門練習—在開始功能表，所有程式，Real Options Valuation，Risk Simulator 快捷方式，包含有多個入門練習和示例結果解釋。這些示例可以讓用戶快速入門 Risk Simulator。
- 入門視頻—這些視頻都可以在我們的官方網站查看
www.realoptionsvaluation.com/download.html
www.rovdownloads.com/download.html

貼士：硬體 ID

- 右鍵 HWID 複製—在安裝授權使用者介面，選擇或者在 HWID 上按兩下選擇該值，右鍵複製或者點擊 Email HWID 連結生成 包含 HWID 的郵件。
- .Troubleshooter—在開始功能表，所有程式，Real Options Valuation，Risk Simulator 資料夾運行 Troubleshooter，以及運行 Get HWID 獲得電腦的 HWID。

貼士：拉丁超立方抽樣（LHS）vs Monte Carlo 模擬

- 相關性—設定相互相關假設之後，建議在 Risk Simulator，選項功能表使用 Monte Carlo 設置，拉丁超立方方法與 copula 模擬法不相容。
- LHS 間距—間距越多，當進行模擬的時候會降低模擬的速度。

- 隨機化—所有在選項功能表的隨機模擬技術都經過了檢驗，方法和引擎的隨機化程度相當高。

貼士：線上資源

- 書籍，入門視頻，模型，白皮書—都可以免費在

www.realoptionsvaluation.com/download.html 或者

www.rovdownloads.com/download.html獲得。

貼士：優化

- 次優解—如果優化運行返回次優解，嘗試將限制條件從等於 (=) 改變為不等於 (>= 或者 <=) 再試一次。該操作對與運行有效邊際分析相同適用。

貼士：文檔

- 多重文檔—創建並且在一個模型的多個文檔間切換。該操作可以進行模擬情景分析，並且可以更改輸入參數或者分佈類型來查看結果。
- 文檔要求—假設，預測，或者決策變數在無文檔的情況下無法被創建。但是，一旦創建了文檔，就沒必要每次都創建新文檔。實際上，如果想運行添加更多假設或者預測的模擬模型，可以在相同的文檔中更改。
- 啟動文檔—最後適用的文檔將隨著 Excel 的保存而保存，打開而打開。
- 多重 Excel 文檔—當在多個打開的 Excel 模型中切換，啟動的文檔也將在不同 Excel 模型中切換。
- 跨工作簿文檔—當有多個 Excel 文檔被打開或者只有一個 Excel 檔和啟動文檔的時候，請注意，突然切換到另外一個 Excel 文檔以及文檔包含設定假設和預測，假設和預測將不會運行。
- 刪除文檔—可以複製當前的文檔或者刪除當前文檔但是注意刪除文檔的時候必須保證還存在有一個文檔最後留下。
- 文檔位置—創建的文檔（包含假設，越策，和決策變數，目標，限制，等等）都保存在壓縮隱藏的工作簿中。這就是每次保存 excel，文檔也相應保存的原因。

貼士：右鍵快速鍵或者其他快速鍵

- 右鍵—可以在 Excel 儲存格中右鍵打開 Risk Simulator 快顯功能表。

貼士：保存

- 保存 Excel 文檔—該操作保存了文檔的設定，假設，預測，決策變數，以及 Excel 模型（包括 Risk Simulator 報告，圖表，和輸出的資料）。
- 保存圖表設置—該操作保存了預測圖表設定例如可以被回復的相同設置，可用于未來的預測圖表（在預測圖表中使用保存和打開圖示）。
- 在 Excel 中保存和輸出模擬資料—該操作可以在 Excel 中輸出模擬的假設和預測，Excel 本身也將被保存以保存這些資料。
- 在 Risk Simulator 中保存模擬資料和圖表—使用 Risk Simulator，資料輸出，和保存為*.RiskSim 格式可以再次打開包含相同資料的動態和預測圖表，而不需要再次運行模擬。
- 保存和生成報告—模擬報告和其他分析報告都以單獨的工作簿輸出，整個 Excel 檔都會保存以保存資料用於未來的研究。

貼士：抽樣和模擬技術

- 亂數產生器—包含 6 中亂數產生引擎（查看使用者手冊瞭解詳情），*ROV Risk Simulator* 默認的方法和*高級負隨機移動法*是推薦是同一方法。不要使用其他的方法，除非模型或者分析會調用這些方法，建議對這兩種方法進行結果測試。

貼士：軟體卡發工具（SDK）和動態連結程式庫

- SDK，DLL 和 OEM—所有 Risk Simulator 中的分析都可以被外部的軟體調用，並且和其他產權的軟體相相容。請聯繫 admin@realoptionsvaluation.com 獲得更多詳細的內容。

貼士：在 Excel 中使用 Risk Simulator

- ROV Troubleshooter—運行 ROV Troubleshooter 獲得自己電腦的 HWID 以獲得授權，查看電腦的設定，以及解決 Risk Simulator 是否被阻止。
- 和 Excel 一起啟動 Risk Simulator—可以讓 Risk Simulator 和 Excel 一起自動啟動，或者點擊開始功能表，所有程式，Real Options valuation，Risk Simulator 快捷方式。這些都可以在 Risk Simulator 的選項中設定。

貼士：極速模擬

- 模型開發—如果在模型中想要進行極速模擬，當模型構建完成後，完全可以運行急速模擬。
- 常規速度—毫無疑問，常規速度也是可以進行的。

貼士：颶風圖分析

- 颶風圖分析—颶風圖分析不能單單只運行一次。它可以作為模型診斷工具，可以在同一模型中運行多次。例如，在較大的模型中，颶風圖可以使用默認設置和所有的引數（選擇顯示所有變數）運行一次。否則的話，該工具提供了起始點用來確定哪些變數是關鍵成功因素（例如，颶風圖可以顯示對輸出影響最大的頭 5 個變數，剩下的 200 個可能就具有較小的影響），第二次颶風圖分析可以運行較少的變數（例如，如果頭 5 個影響較大，就選擇前 10 個變數，這樣就可以創建包含關鍵和非關鍵因素的圖表）。
- 預設值—默認的測試點可以增加 $\pm 10\%$ 到某個更大的值來檢驗非線性（蜘蛛圖將顯示非線性以及颶風圖可以偏向一邊如果變數是非線性的話）。
- 零值和整型值—包含零和整數的輸入需要在颶風圖取消選擇，才能進行運算。否則，進行擾動就沒有效果（例如，如果模型使用者使用 lookup 表格，當 Jan = 1, Feb = 2, Mar = 3 等等，圍繞 1 擾動 $\pm 10\%$ 可以得到 0.9 和 1.1，這對於模型沒有意義）。
- 圖表選項—使用多個圖表選項獲得最優的打開或者關閉模型的選項。

貼士：Troubleshooter

- ROV Troubleshooter—運行 Troubleshooter 或者電腦的 HWID 用於授權，查看電腦的設置，以及重新啟動 Risk Simulator 如果在偶然條件下被阻止。