

# Simulation and Financial Data Analysis Using Stata



高等院校现代金融系列教材

## 随机模拟与金融数据处理 Stata教程

李春涛 张璇 著



中国金融出版社

7830.49

50

高等院校现代金融系列教材

# 随机模拟与金融 数据处理 Stata 教程

李春涛 张璇 著



中国金融出版社

责任编辑：孔德蕴

责任校对：张志文

责任印制：张 莉

## 图书在版编目 (CIP) 数据

随机模拟与金融数据处理 Stata 教程 (Suiji Moni yu Jinrong Shuju Chuli Sta-  
ta Jiaocheng) / 李春涛 张 璇著. —北京：中国金融出版社，2009. 12

(高等院校现代金融系列教材)

ISBN 978 - 7 - 5049 - 5299 - 8

I. 随… II. ①李…②张… III. 金融—数据管理系统—应用软件, Sta-  
ta—高等学校—教材 IV. F830. 49

中国版本图书馆 CIP 数据核字 (2009) 第 189486 号

出版  
发行 **中国金融出版社**

社址 北京市丰台区益泽路 2 号

市场开发部 (010) 63272190, 66070804 (传真)

网上书店 <http://www.chinafph.com>

(010) 63286832, 63365686 (传真)

读者服务部 (010) 66070833, 82672183

邮编 100071

经销 新华书店

印刷 北京松源印刷有限公司

装订 平阳装订厂

尺寸 170 毫米 × 228 毫米

印张 23

字数 419 千

版次 2009 年 12 月第 1 版

印次 2009 年 12 月第 1 次印刷

印数 1—5060

定价 43.00 元

ISBN 978 - 7 - 5049 - 5299 - 8/F. 4859

如出现印装错误本社负责调换 联系电话 (010) 63263947



统计方法已成为各门类社会科学研究必不可少的工具，国内流行的统计软件也越来越多。作为一个界面友好、编程简单、功能强大的统计软件，Stata 越来越引起国内用户的关注和重视。虽然如此，关于 Stata 编程的教材却很少，目前市面上见到的有关图书主要以命令介绍为主，缺乏系统的编程技巧，读者即使掌握了一些具体的命令，仍然很难有系统编程的能力，况且统计软件都是以数据为处理对象的，如果读者没有相应的具体数据，就不能有效地进行编程训练。本书首先用蒙特卡洛方法构造数据，然后用这些数据进行具体的计算，通过读者自己构造的数据反过来验证常用的计量经济学方法和统计模型的正确性。

本书第一部分介绍 Stata 软件的安装、帮助、基本命令和 Stata 的日期编码。通过这一部分的学习，读者对 Stata 会有一个初步的认识，并掌握利用各种资源得到帮助的途径，同时建立起关于 Stata 日期编码的基本概念，能熟练应用各种日期函数将不同的日期格式数据进行转换，能将日期型数据按照习惯的显示方式显示出来。

第二部分是蒙特卡洛模拟的入门，首先介绍伪随机数的产生机理，然后介绍常用的不同分布随机样本的 Stata 仿真方式。通过伪随机数产生机理的学习，读者能了解到计算机仿真出的所谓随机数事实上是通过程序模拟产生出来的确定性数字序列，只不过看似随机且能够通过大部分随机性检验而已。伪随机数的设计，涉及参数的选择，不恰当的选择，会导致一系列的问题，比如自循环性等。接下来，我们通过蒲丰投针试验，介绍蒙特卡洛在处理确定性问题上强大的功能。鉴于实际中可能遇到的形形色色的随机数问题，我们分两个章节分别介绍了离散型和连续型随机数的生成。这些模拟随机数的过程都附有相应的 Stata 程序，几乎每一行程序都有相应的注释。这样，虽然我们没有刻意介绍单个的命令，但是读者可以通过这些程序了解到具体命令的用法。接下来我们介绍了蒙特卡洛模拟在数值积分和处理复杂性系统中的应用，将复杂的数值积分问题和概率问题，通过蒙特卡洛模拟直观地展示在读者面前，目的是让读者对 Stata 编程和蒙特卡洛试验产生浓厚的兴趣。

第三部分我们从蒙特卡洛模拟转入具体数据的处理方法，我们的重点是中国金融数据的处理，在介绍过程中我们也会使用蒙特卡洛模拟生成某些不易得到的数据作为统计处理的对象。在数据处理方法上，这一部分涉及数据的编码、不同类型和来源数据的 Stata 处理方法，不同来源数据的合并等。特别是在第 14 章，本书介绍了如何计算相关系数矩阵，以及如何通过修改部分代码将结果输出的过程简化。通过这一部分的学习，读者可以掌握到常用的中国金融市场数据的 Stata 处理技巧，以及一些特殊数据的编程处理方法。

第四部分介绍具体模型的估计和计算结果的输出方法。我们介绍了一些常用的回归分析模型，包括线性回归模型和离散选择模型，但是我们把重点放在模型结果的输出上，力求使读者通过这部分的学习，能够在实际研究中将研究结果通过 Stata 编程，将计算结果以期刊文章常用的表格形式几乎自动地输出到 MS Word 中。

最后，本书在第五部分通过几个具体的实证案例介绍几个常用的金融实证分析方法，涉及期权定价的蒙特卡洛模拟、事件研究方法和对照组研究方法等非常实用的知识。有兴趣的读者可以根据本书的介绍，重复有关的研究。

本书把最有用的程序设计方法融合在一本关于蒙特卡洛模拟和金融研究方法的教材里，相信会给读者带来事半功倍的效果。通过对这本书的学习，读者不仅能熟悉 Stata 编程的技巧，而且能独立进行数据处理和熟悉蒙特卡洛模拟的精髓。本书的使用对象包括社会科学的科研人员、高等学校研究生、高年级本科生，既可以作为《计量经济学》、《应用计量经济学》和《金融实证分析》的教学辅助材料，也可以独立成为一门涉及统计软件编程或者蒙特卡洛模拟的教学参考用书。

作者曾经在香港大学长期从事研究工作，多年来我们一直希望能将自己在编程上的心得体会介绍给国内的 Stata 用户，特别是那些刚刚开始学习使用该软件的大学生和研究生。如今，虽然历时多年完成了本书的撰写，但是常常感觉到时间紧迫，没有精力将有关的方法详细地介绍，其中的错误也是在所难免的。我们欢迎广大读者对本书的错漏之处提出建议和批评，也希望有机会能在再版的时候，增加更多的内容和程序介绍。读者如有问题、建议或者批评，烦请发送电子邮件到 leechtcn@gmail.com，我们将非常感谢。

李春涛 张璇  
2009 年 10 月 16 日



## 第一篇 Stata 基础

|    |                         |
|----|-------------------------|
| 3  | 第 1 章 Stata 的安装、升级和界面介绍 |
| 3  | 1.1 Stata10 的安装         |
| 4  | 1.2 Stata 的升级           |
| 5  | 1.3 Stata 的界面介绍         |
| 11 | 1.4 Stata10 窗口介绍        |
| 16 | 第 2 章 帮助系统              |
| 16 | 2.1 系统帮助                |
| 18 | 2.2 网络帮助                |
| 21 | 2.3 专家帮助                |
| 23 | 第 3 章 Stata 的日期和时间      |
| 24 | 3.1 date() 函数           |
| 28 | 3.2 mdy() 函数            |
| 30 | 3.3 td() 函数             |
| 30 | 3.4 从日期到年、月、日等          |
| 31 | 3.5 月度和季节数据             |
| 32 | 3.6 clock() 函数          |

## 第二篇 蒙特卡洛模拟

|    |               |
|----|---------------|
| 39 | 第 4 章 伪随机数的生成 |
| 39 | 4.1 引言        |

|     |                                 |
|-----|---------------------------------|
| 40  | 4.2 线性同余法                       |
| 41  | 4.3 线性同余法在 Stata 中的实现           |
| 48  | <b>第 5 章 蒲丰投针问题</b>             |
| 48  | 5.1 试验概述与数学推导                   |
| 50  | 5.2 计算机模拟                       |
| 52  | 5.3 计算机模拟的改进                    |
| 56  | 5.4 多次试验模拟                      |
| 61  | 5.5 方圆鱼缸试验                      |
| 64  | <b>第 6 章 离散型随机变量的模拟</b>         |
| 64  | 6.1 事件空间有限的离散型随机变量              |
| 70  | 6.2 几何分布离散型随机变量                 |
| 73  | 6.3 泊松分布离散型随机变量                 |
| 75  | 6.4 贝努利离散型随机变量 ( $n$ 次试验成功的次数)  |
| 82  | <b>第 7 章 连续型随机变量的模拟</b>         |
| 82  | 7.1 均匀分布                        |
| 84  | 7.2 指数分布                        |
| 85  | 7.3 指数分布和正态分布 (Box - Muller 方法) |
| 89  | 7.4 Gamma 分布                    |
| 91  | 7.5 正态分布                        |
| 93  | 7.6 多维正态分布                      |
| 99  | 7.7 截断分布抽样方法                    |
| 100 | 7.8 接受—拒绝抽样方法                   |
| 103 | <b>第 8 章 数值积分方法</b>             |
| 103 | 8.1 定积分问题                       |
| 106 | 8.2 广义积分                        |
| 112 | 8.3 积分应用问题                      |
| 114 | <b>第 9 章 蒙特卡洛应用</b>             |
| 114 | 9.1 搭车问题                        |

|     |          |
|-----|----------|
| 116 | 9.2 排队问题 |
|-----|----------|

### 第三篇 金融数据预处理

|     |                            |
|-----|----------------------------|
| 135 | <b>第 10 章 数据读入方法</b>       |
| 135 | 10.1 键盘输入数据                |
| 136 | 10.2 读入 Stata 数据文件         |
| 139 | 10.3 读入制表符分割的“.txt”数据文件    |
| 142 | 10.4 读入固定宽度的“.txt.”数据文件    |
| 145 | 10.5 特殊数据的读取               |
| 154 | 10.6 基金经理变更数据              |
| 160 | <b>第 11 章 交易数据的下载和预处理</b>  |
| 160 | 11.1 Wind 数据库下载交易数据        |
| 165 | 11.2 Wind 交易数据的纵向合并        |
| 168 | 11.3 CSMAR 数据库交易数据的处理技巧    |
| 175 | 11.4 CSMAR 股权变更数据的处理技巧     |
| 181 | <b>第 12 章 财务数据的下载和预处理</b>  |
| 181 | 12.1 Wind 上市公司财务数据的下载      |
| 183 | 12.2 Wind 上市公司财务数据的合并处理    |
| 189 | <b>第 13 章 数据标签</b>         |
| 189 | 13.1 文件标签 (label data)     |
| 191 | 13.2 变量标签 (label variable) |
| 192 | 13.3 赋值标签 (label values)   |
| 198 | <b>第 14 章 相关系数</b>         |
| 198 | 14.1 命令格式与缺陷               |
| 201 | 14.2 手动解决方案                |
| 206 | 14.3 修改 Ado 文件的解决方案        |



## 第四篇 模型估计与结果输出

### 第 15 章 线性回归分析与结果输出

- 215      15.1 引言
- 215      15.2 数据准备
- 217      15.3 数据描述与基本统计量
- 226      15.4 利用图形了解数据
- 230      15.5 线性回归分析
- 234      15.6 线性回归结果的输出
- 241      15.7 预测

### 第 16 章 定性变量回归分析

- 245      16.1 二元选择问题
- 246      16.2 数据介绍
- 249      16.3 线性回归方法
- 251      16.4 Probit 模型
- 256      16.5 Probit 回归结果的输出
- 265      16.6 边际效应
- 266      16.7 Logit 回归
- 272      16.8 顺序选择模型 oprobit (ologit) 回归

## 第五篇 金融实证研究方法专题

### 第 17 章 期权定价问题

- 277      17.1 期权及期权定价模型简介
- 278      17.2 Black – Scholes 期权定价模型
- 280      17.3 用蒙特卡洛模拟计算期权价格
- 284      17.4 修正的 B – S 期权定价模型

### 第 18 章 股票价格与指数的同步性

- 286      18.1 问题概述
- 287      18.2 数据说明

|     |                       |
|-----|-----------------------|
| 289 | 18.3 同步性的计算及程序        |
| 296 | 18.4 完整的程序            |
| 308 | 18.5 模型设定             |
| 310 | 18.6 最小二乘回归结果         |
| 313 | 18.7 工具变量回归           |
| 316 | <b>第 19 章 事件研究方法</b>  |
| 316 | 19.1 数据准备：分析师评级历史记录   |
| 321 | 19.2 累积超额收益率（CAR）的计算  |
| 334 | <b>第 20 章 对照组研究方法</b> |
| 334 | 20.1 方法概述             |
| 335 | 20.2 规模相近             |
| 335 | 20.3 行业相同             |
| 338 | 20.4 数据模拟             |
| 342 | 20.5 对照组研究编程方法        |
| 354 | <b>参考文献</b>           |
| 355 | <b>后记</b>             |

# 第一篇 Stata 基础

工欲善其事，必先利其器。

社会科学研究，包括经济、金融、法律、生物统计、公共卫生、社会学、教育学甚至历史学<sup>①</sup>等各个学科门类，都越来越重视实证研究，越来越依赖于统计软件。

Stata 是一个非常优秀的统计、计量软件，由于界面友好、编程简单、功能强大，它正受到越来越多的关注，其市场占有率也越来越高，Stata 正逐渐成为许多研究者进行各门类社会科学学习和研究的首选软件。

本书第 1 章首先介绍该软件的安装和基本界面，第 2 章介绍如何利用帮助和如何利用网络寻找有用的代码，第 3 章介绍 Stata 的日期和时间函数。本书按照 Stata10 展开介绍，但是大部分内容适用于当前广泛使用的 7.0 以后的各个版本。阅读本章之前，本书强烈建议读者阅读 Stata 手册 Getting Started 部分（关于 Stata 的最基本的详细介绍），或者访问 UCLA 关于 Stata 功能介绍的网站（<http://www.ats.ucla.edu/stat/stata/>）。

---

<sup>①</sup> 1993 年瑞典皇家科学院和瑞典银行将诺贝尔经济奖授予 Robert William Fogel 和 Douglass Cecil North，以表彰他们在新历史研究领域作出的杰出贡献。他们的研究被称做历史计量学（Cliometrics），该学科将历史学和计量经济学有机地结合起来，开创了一个全新的历史学研究领域。





## Stata 的安装、升级和界面介绍

### 1.1 Stata10 的安装

Stata10 的安装与其他软件相比没有特殊的地方，购买 Stata 软件后，打开 readme 或安装说明文件，双击 setup.exe，然后按照安装精灵的提示进行操作。安装以后，用户电脑桌面出现 Stata 图标，点击该图标，在弹出的对话框中输入有关的用户信息、版权信息和序列号，即完成安装。我们可以用 verinst 命令检查安装是否正确。从命令输入窗口键入 verinst 然后按回车键，如果得到类似如下的信息，说明 Stata 安装正确：

```
• verinst
```

```
Stata/SE 10.0 for Windows
```

```
Born 03 Mar 2008
```

```
Copyright (C)1985 -2008
```

```
Total physical memory: 1048040 KB
```

```
Available physical memory: 468952 KB
```

```
20 -student Stata for Windows (network) perpetual license:
```

```
Serial number: 81910554459
```

```
Licensed to: LEE CHUNTAO
```

```
CUFE CHINA
```

## 1.2 Stata 的升级

新安装的 Stata 软件一定要立即升级, 否则某些功能不能正常运行 (比如 `intreg` 命令会返回一个 “Mata Run Time Error” 的错误信息, 但升级以后该问题就立即解决)。从 Stata 公司购买的 Stata10 软件, 可以通过 Internet 实现在线升级, 操作方法是直接在命令窗口输入以下命令:

- `update query`
- `update all`

在线升级完成以后, Stata 会告诉你最新版本的主程序文件已经成功下载, 但是需要你从命令输入窗口键入 `update swap` 来完成整个升级过程。因此, 真正完成在线升级, 用户需要执行如下的三行命令:

- `update query`
- `update all`
- `update swap`

对于无法连接 Internet 的计算机, 可以采用离线升级的方法升级。stata10 的离线更新有以下几个步骤:

(1) 下载最新的执行文件, 注意你使用的 Stata 的版本, SE (Special Edition) 版本用户要下载 `wsestata. bin`, MP (Multiple Processor) 版用户则需要下载 `wmpstata. bin` 文件, 并将下载得到 `. bin` 文件存放于 Stata 的安装目录中<sup>①</sup>。

(2) 将原来的 Stata 执行文件删除或者将扩展名由 `. exe` 更名为 `. old`。

(3) SE 版用户将前面下载到的 `wsestata. bin` 文件更名为 `wsestata. exe`, MP 版本用户将文件 `wmpstata. bin` 更名为 `wmpstata. exe`。

(4) 下载最新 `ado` 文件: 下载文件 `ado. zip`, 解压到某个路径名中不带空格和汉字的目录下 (如 `c:\ temp` 目录), 打开 stata10, 从命令输入窗口中依次执行如下命令。

- `update from c:\ temp`
- `update ado`

---

<sup>①</sup> MP 是 Stata SE 的一个增强版, 由于 2006 年以后的计算机很多支持双核或者多 CPU 模式, Stata MP 针对这种计算机开发, 事实上是能通过多个处理器进行并行计算的 Stata 新版本。Stata 除了 SE 和 MP 之外还有很多其他的版本, 比如网络版本, 基于 Macintosh 操作系统的版本和基于 Unix 操作系统的版本。本书主要介绍基于 Windows 操作系统的 SE 版本, 但是, 据作者所知, 大部分命令适用于基于 Unix 的网络版本。

对应的下载地址分别是:

MP: <http://www.stata.com/updates5/win mp/wmpstata.bin>

SE: <http://www.stata.com/updates5/win se/wsestata.bin>

Ado: <http://www.stata.com/support/updates/stata10/ado/ado.zip>

## 1.3 Stata 的界面介绍

### 1.3.1 主窗口布局

图 1.1 给出 Stata 的窗口布局。右上方为屏幕输出窗口, 屏幕输出窗口下面的白色长条是命令输入窗口, 而我们输入的每一行命令都会滚动显示在左上方的历史命令窗口中。统计软件 Stata 处理的对象是数据, 数据以变量为最主要的存

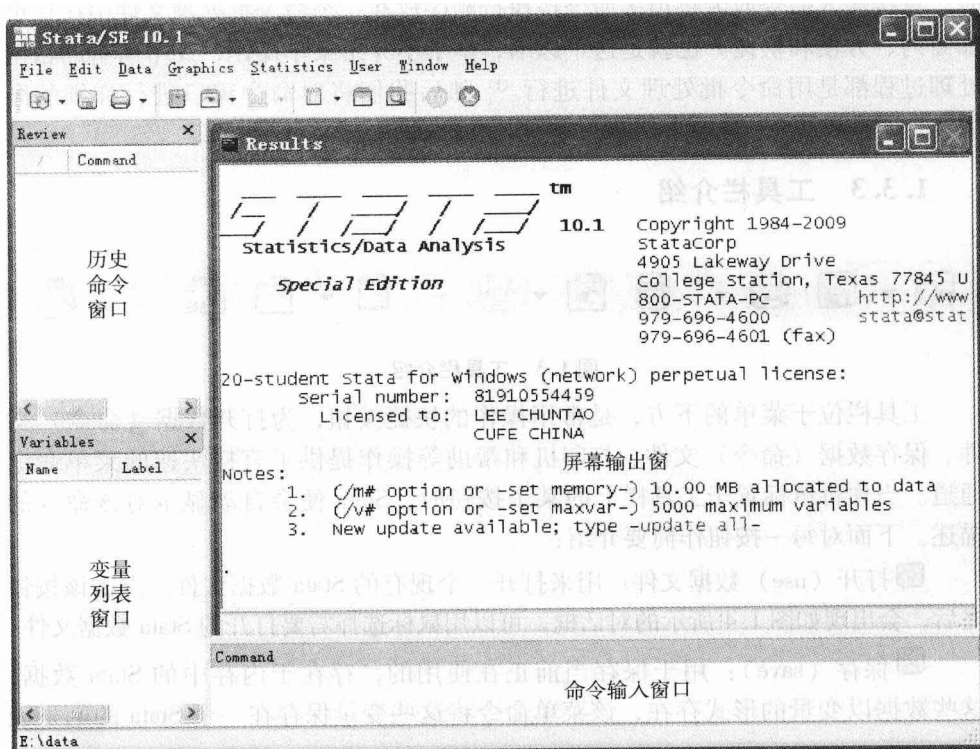


图 1.1 Stata 基本界面介绍

在形式。读入数据后，每一个变量的名称都会显示在左下方的变量列表窗口中。

### 1.3.2 主菜单介绍

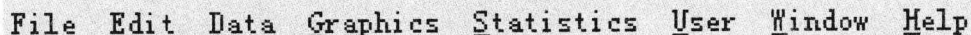


图 1.2 菜单栏介绍

菜单栏位于 Stata 界面的顶端，一共包括 8 个下拉式菜单（某些版本是 9 个甚至 10 个菜单，而且用户也可以增加自己的菜单）。用鼠标点击其中的任何一个菜单，下拉式菜单会显示出该菜单项的一系列命令。有时读者需要使用不同菜单下的多个命令，点击菜单的操作方式显得麻烦而且不利于重复运算，不少用户偏爱直接在命令窗口输入命令的操作方式。特别是当您需要进行大规模的运算时，直接输入命令或者将很多命令按执行顺序放在一个命令批处理文件中往往更加简约、方便和快捷。也就是这个原因，本书不介绍菜单操作，几乎所有的数据处理过程都是用命令批处理文件进行。<sup>①</sup> 随后我们将对命令窗口进行详细的介绍。

### 1.3.3 工具栏介绍



图 1.3 工具栏介绍

工具栏位于菜单的下方，是常用操作的快捷按钮，为打开数据（命令）文件、保存数据（命令）文件、打印机和帮助等操作提供了直接快速的菜单命令通道。当你将鼠标置于工具栏上的某个按钮时，Stata 便会自动显示对该命令的描述。下面对每一按钮作简要介绍：

 打开（use）数据文件：用来打开一个现有的 Stata 数据文件。点击该按钮之后，会出现如图 1.4 所示的对话框，可以用鼠标选择需要打开的 Stata 数据文件。

 保存（save）：用于保存当前正在使用的，存在于内存中的 Stata 数据，这些数据以变量的形式存在，该菜单命令将这些变量保存在一个 Stata 的电子表

<sup>①</sup> 初学者往往偏爱菜单式操作方式，但是大部分研究人员都发现菜单驱动的弊端，往往选择基于命令和程序的统计软件，比如 Stata、Matlab、R 和 SAS 等。



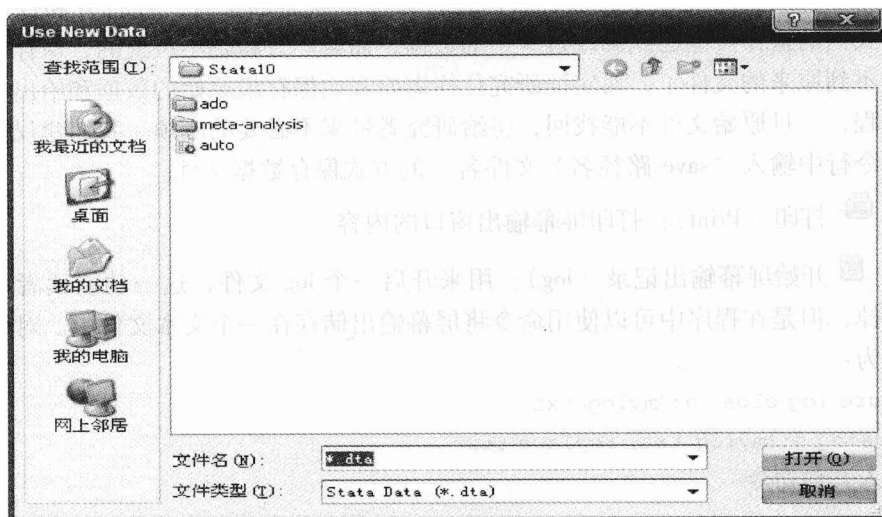


图 1.4 打开数据文件

格中。点击该按钮之后，会出现如图 1.5 所示的对话框，用户需要选择文件的保存路径、保存类型并输入要保存的文件名称。一般情况下不要使用这一按钮，因为你的数据可能来自于一个已经存在的数据文件，点击 save 按钮以后，Stata 将

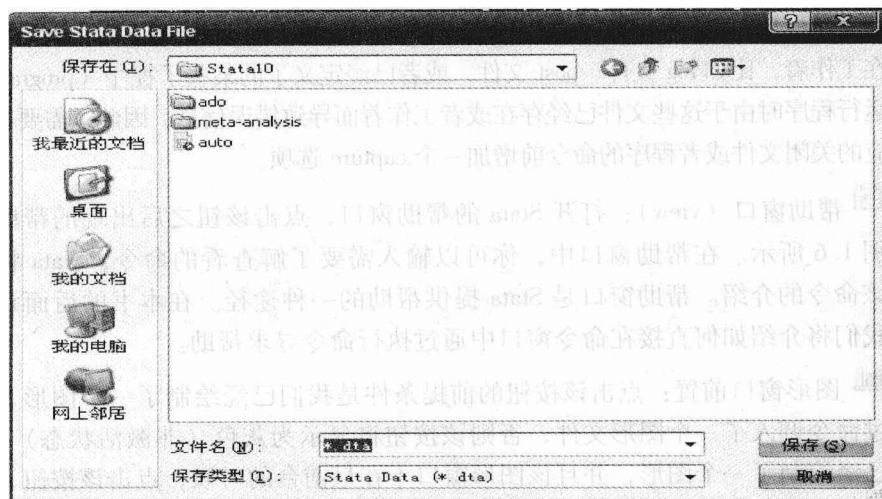


图 1.5 保存数据文件