Probability Theory and Statistical Inference

Econometric Modeling with Observational Data

Aris Spanos

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1999

Printed in the United Kingdom at the University Press, Cambridge

Typeset in MT Times NR 9½/12½ [SE]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Spanos, Aris, 1952–
Probability Theory and Statistical Inference: econometric modeling with observational data / Aris Spanos
p. cm.
Includes bibliographical references (p.) and index.
ISBN 0 521 41354 0
1. Econometrics. 2. Probabilities. I.Title.
HB139.S62 1998
330′.01′5195–dc21

ISBN 0 521 41354 0 hardback
ISBN 0 521 42408 9 paperback

Contents

1 An introduction to empirical modeling

1.1 Introduction

In an attempt to give some idea of what empirical modeling is all about, we begin the discussion with an epigrammatic demarcation of its intended scope:

> **Empirical modeling** is concerned with the parsimonious description of observable stochastic phenomena using statistical models.

The above demarcation is hardly illuminating because it involves the unknown terms *stochastic phenomenon* and *statistical model* which will be explained in what follows. At this stage, however, it suffices to note the following distinguishing features of empirical (as opposed to other forms of) modeling:

(a) the *stochastic* nature of the phenomena amenable to such modeling,
(b) the indispensability of the *observed data*, and
(c) the nature of the description in the form of a *statistical model*.

The primary objective of empirical modeling is to provide an *adequate description* of certain types of observable phenomena of interest in the form of stochastic mechanisms we call *statistical models.* A statistical model purports to capture the *statistical systematic information* (see sections 2–3), which is different from the theory information (see section 4). In contrast to a *theory model*, a statistical model is codified exclusively in terms of probabilistic concepts and it is descriptive and anti-realistic in nature (see chapter 10 for further discussion). The *adequacy* of the description is assessed by how well the postulated statistical model accounts for all the statistical systematic information in the data (see section 5). In section 6 we provide a preliminary discussion of certain important dimensions of the constituent element of empirical modeling, the observed data.

Empirical modeling in this book is considered to involve a wide spectrum of interrelated procedures including:

(i) *specification* (the choice of a statistical model),
(ii) *estimation* (estimation of the parameters of the postulated statistical model),

(iii) *misspecification testing* (assessing the validity of the probabilistic assumptions of the postulated statistical model), and
(iv) *respecification* (an alternative choice of a statistical model).

As argued below, these facets of modeling are particularly involved in the case of **observational data**. In the case of **experimental data** the primary focus is on estimation because facets (i) and (iv) constitute the other side of the *design* coin and (iii) plays a subsidiary role.

A quintessential example of empirical modeling using observational data is considered to be *econometrics*. An important thesis adopted in this book is that econometrics differs from mainstream statistics (dominated by the experimental design and the least-squares traditions), not so much because of the economic theory dimension of modeling, but primarily because of the particular modeling issues that arise due to the *observational nature* of the overwhelming majority of economic data. Hence, we interpret the traditional definition of econometrics "the estimation of relationships as suggested by economic theory" (see Harvey (1990), p. 1), as placing the field within the experimental design modeling framework. In a nutshell, the basic argument is that the traditional econometric textbook approach utilizes the experimental design modeling framework for the analysis of non-experimental data (see Spanos (1995b) for further details).

1.1.1 A bird's eye view of the chapter

The rest of this chapter elaborates on the distinguishing features of empirical modeling (a)–(c). In section 2 we discuss the meaning of **stochastic observable phenomena** and why such phenomena are amenable to empirical modeling. In section 3, we discuss the relationship between stochastic phenomena and **statistical models**. This relationship comes in the form of *statistical systematic information* which is nothing more than the formalization of the chance regularity patterns exhibited by the observed data emanating from stochastic phenomena. In section 4 we discuss the important notion of statistical adequacy: whether the postulated statistical model "captures" all the statistical systematic information in the data. In section 5 we contrast the statistical and theory information. In a nutshell, the theoretical model is formulated in terms of the behavior of economic agents and the statistical model is formulated exclusively in terms of probabilistic concepts; a sizeable part of the book is concerned with the question of: What constitutes statistical systematic information? In section 6 we raise three important issues in relation to **observed data**, their different *measurement scales*, their *nature*, and their *accuracy*, as they relate to the statistical methods used for their modeling.

The main message of this chapter is that, in assessing the validity of a theory, the modeler is required to ensure that the observed data constitute an unprejudiced witness whose testimony can be used to assess the validity of the theory in question. A statistical model purports to provide an adequate summarization of the statistical systematic information in the data in the form of a stochastic mechanism that conceivably gave rise to the observed data in question.

1.2 Stochastic phenomena, a preliminary view

As stated above, the intended scope of empirical modeling is demarcated by the stochastic nature of observable phenomena. In this section we explain intuitively the idea of a stochastic phenomenon and relate it to the notion of a statistical model in the next section.

1.2.1 Stochastic phenomena and chance regularity

A **stochastic phenomenon** is one whose observed data exhibit what we call *chance regularity patterns*. These patterns are usually revealed using a variety of graphical techniques.

The essence of *chance regularity,* as suggested by the term itself, comes in the form of two entwined characteristics:

chance: an inherent uncertainty relating to the occurence of particular outcomes,
regularity: an abiding regularity in relation to the occurence of many such outcomes.

TERMINOLOGY: the term chance regularity is introduced in order to avoid possible confusion and befuddlement which might be caused by the adoption of the more commonly used term known as **randomness**; see chapter 10 for further discussion.

At first sight these two attributes might appear to be contradictory in the sense that *chance* refers to the *absence* of order and "regularity" denotes the *presence* of order. However, there is no contradiction because the disorder exists at the level of individual outcomes and the order at the aggregate level. Indeed, the essence of chance regularity stems from the fact that the disorder at the individual level creates (somehow) order at the aggregate level. The two attributes should be viewed as inseparable for the notion of chance regularity to make sense. When only one of them is present we cannot talk of chance regularity.

Any attempt to define formally what we mean by the term *chance regularity* at this stage will be rather pointless because one needs several mathematical concepts that will be developed in what follows. Instead, we will attempt to give some intuition behind the notion of chance regularity using a simple example and postpone the formal discussion until chapter 10.

Example
Consider the situation of casting two dice and adding the dots on the sides facing up. The *first* crucial feature of this situation is that at each trial (cast of the two dice) the outcome (the sum of the dots of the sides) cannot be guessed with any certainty. The only thing one can say with certainty is that the outcome will be one of the numbers:

$$\{2,3,4,5,6,7,8,9,10,11,12\},$$

we exclude the case where the dice end up standing on one of the edges! All 36 possible combinations behind the outcomes are shown in table 1.1. The *second* crucial feature of the situation is that under certain conditions, such as the dice are symmetric, we know that certain outcomes are more likely to occur than others. For instance, we know that the number 2 can arise as the sum of only one set of faces: {1,1} – each die comes up with 1; the same applies to the number 12 with faces: {6,6}. On the other hand, the number 3 can arise as the sum of two sets of faces: {(1,2), (2,1)}; the same applies to the number 11 with faces: {(6,5),(5,6)}. In the next subsection we will see that this line of combinatorial reasoning will give rise to a *probability distribution* as shown in table 1.3.

Table 1.1. *Outcomes in casting two dice*

	1	**2**	**3**	**4**	**5**	**6**
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

At this stage it is interesting to pause and consider the notions of chance regularity as first developed in the context of such games of chance. This is, indeed, the way probabilities made their first appearance. Historically, probabilities were introduced as a way to understand the differences noticed empirically between the likely occurrence of different betting outcomes, as in table 1.1. Thousands of soldiers during the medieval times could attest to the differences in the empirical relative frequencies of occurrence of different events related to the outcomes in table 1.1. While waiting to attack a certain town, the soldiers had thousands of hours with nothing to do and our historical records suggest that they indulged mainly in games of chance like casting dice. After thousands of trials they knew intuitively that the number 7 occurs more often than any other number and that 6 occurs less often than 7 but more often than 5. Let us see how this intuition was developed into something more systematic that eventually led to probability theory.

Table 1.2 reports 100 actual trials of the random experiment of casting two dice and adding the number of dots turning up on the uppermost faces of the dice. A look at the table confirms only that the numbers range from 2 to 12 but no real patterns are apparent, at least at first sight.

Table 1.2. *Observed data on dice casting*

3	10	11	5	6	7	10	8	5	11	2	9	9	6	8	4	7	6	5	12
7	8	5	4	6	11	7	10	5	8	7	5	9	8	10	2	7	3	8	10
11	8	9	5	7	3	4	9	10	4	7	4	6	9	7	6	12	8	11	9
10	3	6	9	7	5	8	6	2	9	6	4	7	8	10	5	8	7	9	6
5	7	7	6	12	9	10	4	8	6	5	4	7	8	6	7	11	7	8	3

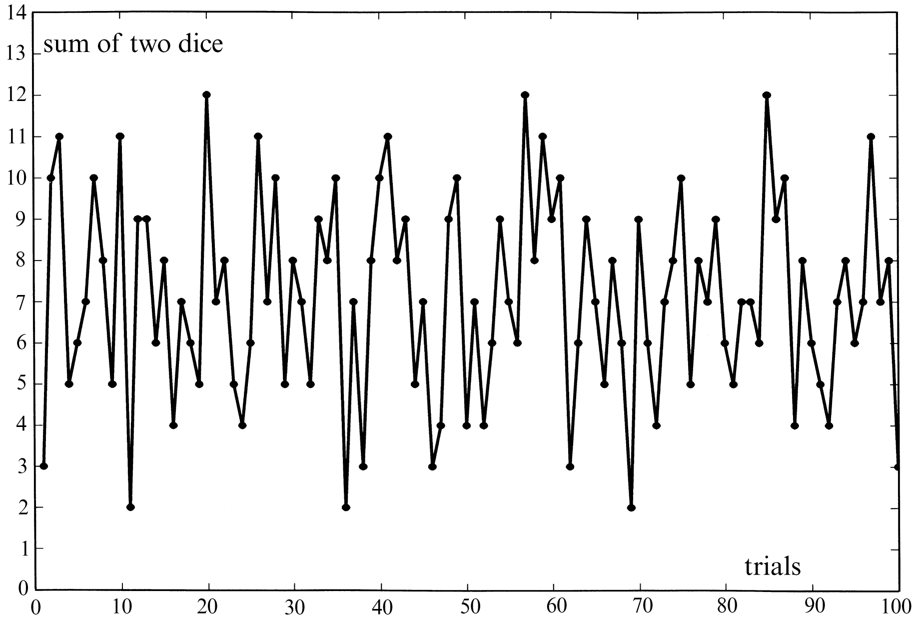

Figure 1.1 A sequence of 100 throws of two dice

In figure 1.1 the data are plotted over the index of the number of the trial. At the first casting of the dice the sum was 3, at the second the sum was 10, at the third the sum of 11 etc. Joining up these outcomes (observations) gives the viewer a better perspective with regard to the sequential nature of the observations. NOTE that the ordering of the observations constitutes an important dimension when discussing the notion of chance regularity.

Historically, the first chance regularity pattern discerned intuitively by the medieval soldiers was that of *a stable law of relative frequencies* as suggested by the histogram in figure 1.2 of the data in table 1.2; without of course the utilization of graphical techniques but after numerous casts of the dice. The question that naturally arises at this stage is:

How is the histogram in figure 1.2 related to the data in figure 1.1?

Today, c*hance regularity* patterns become discernible by performing a number of thought experiments.

Thought experiment 1 Think of the observations as little squares with equal area and rotate the figure 1.1 clockwise by 90° and let the squares representing the observations fall vertically creating a pile on the x-axis. The pile represents the well-known histogram as shown in figure 1.2. This histogram exhibits a clear triangular shape that will be related to a probability distribution derived by using arguments based on combinations and permutations in the next sub-section. For reference purposes we summarize this regularity in the form of the following intuitive notion:

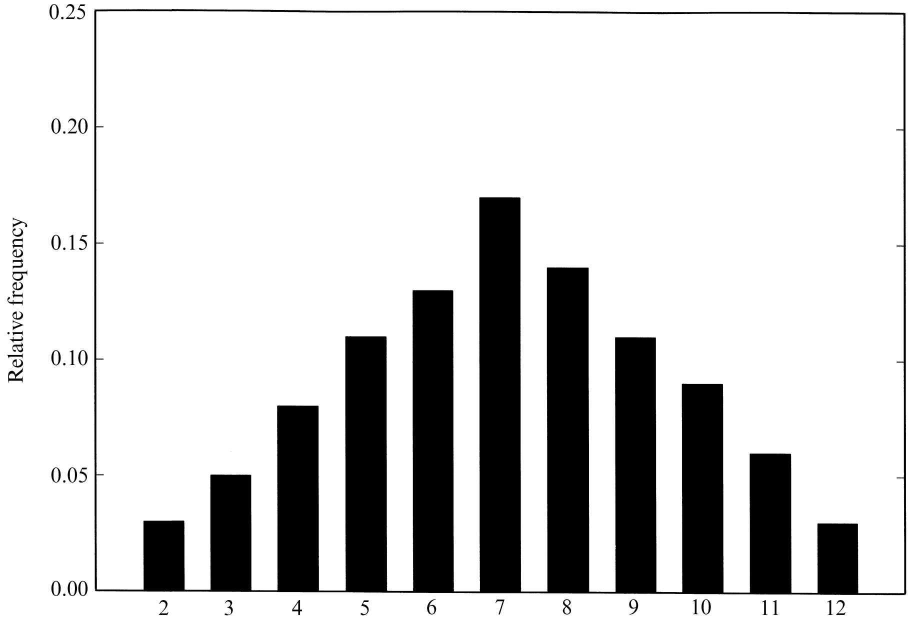

Figure 1.2 Histogram of the sum of two dice data

[1] *Distribution*: after several trials the outcomes form a (seemingly) stable law.

Thought experiment 2 Hide the observations following a certain value of the index, say $t = 40$, and try to guess the next outcome. Repeat this along the observation index axis and if it turns out that it is impossible to use the previous observations to guess the value of the next observation, excluding the extreme cases 2 and 12, then the chance regularity pattern we call *independence* is present. It is important to note that in the case of the extreme outcomes 2 and 12 one is almost sure that after 2 the likelihood of getting a number greater than that is much higher, and after 12 the likelihood of getting a smaller number is close to one. As argued below, this type of predictability is related to the regularity component of chance known as a stable relative frequencies law. Excluding these extreme cases, when looking at the previous observations, one cannot discern a pattern in figure 1.1 which helps narrow down the possible alternative outcomes, enabling the modeler to guess the next observation (within narrow bounds) with any certainty. Intuitively, we can summarize this notion in the form of:

[2] *Independence*: in any sequence of trials the outcome of any one trial does not influence and is not influenced by that of any other.

Thought experiment 3 Take a wide frame (to cover the spread of the fluctuations in a t-plot such as figure 1.1) that is also long enough (roughly less than half the length of the horizontal axis) and let it slide from left to right along the horizontal axis looking at the picture inside the frame as it slides along. In the case where the picture does not change significantly, the data exhibit *homogeneity*, otherwise *heterogeneity* is present; see chapter 5. Another way to view this pattern is in terms of the average and the *variation* around this average of the numbers as we move from left to right. It appears as though this *sequential average* and its *variation* are relatively constant around 7. The *variation* around this constant average value appears to be within constant bands. This chance regularity can be intuitively summarized by the following notion:

[3] *Homogeneity*: the probabilities associated with the various outcomes remain identical for all trials.

NOTE that in the case where the pattern in a t-plot is such so as to enable the modeler to guess the next observation *exactly*, the data do not exhibit any chance pattern, they exhibit what is known as *deterministic* regularity. The easiest way to think about deterministic regularity is to visualize the graphs of mathematical functions from elementary (polynomial, algebraic, transcendental) to more complicated functions such as Bessel functions, differential and integral equations. If we glance at figure 1.1 and try to think of a function that can describe the zig-zag line observed, we will realize that no such mathematical function exists; unless we use a polynomial of order 99 which is the same as listing the actual numbers. The patterns we discern in figure 1.1 are chance regularity patterns.

1.2.2 Chance regularity and probabilistic structure

The step from the observed regularities to their formalization (mathematization) was prompted by the distribution regularity pattern as exemplified in figure 1.2. The formalization itself was initially very slow, taking centuries to materialize, and took the form of simple combinatorial arguments. We can capture the essence of this early formalization if we return to the dice casting example.

Example

In the case of the experiment of casting two dice, we can continue the line of thought that suggested differences in the likelihood of occurrences of the various outcomes in $\{2,3,4,5,6,7,8,9,10,11,12\}$ as follows. We already know that 3 occurs twice as often as 2 or 11. Using the same common sense logic we can argue that since 4 occurs when any one of $\{(1,3), (2,2), (3,1)\}$ occurs, its likelihood of occurrence is three times that of 2. Continuing this line of thought and assuming that the 36 combinations can occur with the same probability, we discover a distribution that relates each outcome with a certain likelihood of occurrence shown below in figure 1.3; first derived by Coordano in the 1550s. As we can see, the outcome most likely to occur is the number 7; it is no coincidence that several games of chance played with two dice involve the number 7. We think of the likelihoods of occurrence as *probabilities* and the overall pattern of such probabilities associated with each outcome as a *probability distribution*; see chapter 3.

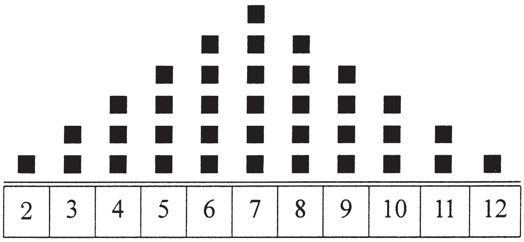

Figure 1.3 Regularity at the aggregate

Table 1.3. *The sum of two dice: a probability distribution*

outcomes	2	3	4	5	6	7	8	9	10	11	12
probabilities	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The probability distribution in table 1.3 represents a probabilistic concept formulated by mathematicians in order to capture the chance regularity in figure 1.1. A direct comparison between figures 1.2 and 1.3 confirms the soldiers' intuition. The empirical relative frequencies in figure 1.2 are close to the theoretical probabilities shown in figure 1.3. Moreover, if we were to repeat the experiment 1000 times, the relative frequencies would have been even closer to the theoretical probabilities; see chapter 10. In this sense we can think of the histogram in figure 1.2 as an empirical realization of the probability distribution in figure 1.3 (see chapter 5 for further discussion).

Example

In the case of the experiment of casting two dice, the medieval soldiers used to gamble on whether the outcome is an odd or an even number (the Greeks introduced these concepts at around 300 BC). That is, soldier A would bet on the outcome being $A=\{3,5,7,9,11\}$ and soldier B on being $B=\{2,4,6,8,10,12\}$. At first sight it looks as though soldier B will be a definite winner because there are more even than odd numbers. The medieval soldiers, however, knew by empirical observation that this was not true! Indeed, if we return to table 1.3 and evaluate the probability of event A occurring, we discover that the soldiers were indeed correct: the probability of both events is $\frac{1}{2}$; the probability distribution is given in table 1.4.

Table 1.4. *The sum of two dice: odd and even*

outcomes	$A=\{3,5,7,9,11\}$	$B=\{2,4,6,8,10,12\}$
probabilities	$\frac{1}{2}$	$\frac{1}{2}$

We conclude this subsection by reiterating that the stochastic phenomenon of casting two dice gave rise to the observed data depicted in figure 1.1, which exhibit the three different forms' chance regularity patterns:

[1] Distribution (triangular), [2] Independence, and [3] Homogeneity.

For reference purposes, it is important to note that the above discernible patterns, constitute particular cases of chance regularity patterns related to three different broad categories of probabilistic assumptions we call **Distribution**, **Dependence**, and **Heterogeneity**, respectively; see chapter 5. The concepts underlying these categories of probabilistic assumptions will be defined formally in chapters 3–4.

A digression – Chevalier de Mere's paradox

Historically, the connection between a stable law of relative frequencies and probabilities was forged in the middle of the 17th century in an exchange of letters between Pascal and Fermat. In order to get a taste of this early formulation, let us consider the following historical example.

The Chevalier de Mere's paradox was raised in a letter from Pascal to Fermat on July 29, 1654 as one of the problems posed to him by de Mere (a French nobleman and a studious gambler). De Mere observed the following empirical regularity:

> the probability of getting at least one 6 in 4 casts of a die is greater than $\frac{1}{2}$, but
>
> the probability of getting a double 6 in 24 casts with *two* dice is less than $\frac{1}{2}$.

De Mere established this empirical regularity and had no doubts about its validity because of the enormous number of times he repeated the game. He was so sure of its empirical validity that he went as far as to question the most fundamental part of mathematics, arithmetic itself. Reasoning by analogy, de Mere argued that the two probabilities should be identical because one 6 in 4 casts of one die is the same as a double 6 in 24 casts of two dice since, according to his way of thinking: 4 is to 6 as 24 is to 36.

The statistical distribution in table 1.4 can be used to explain the empirical regularity observed by de Mere. Being a bit more careful than de Mere, one can argue as follows (the manipulations of probabilities are not important at this stage):

Probability of one double six $=\frac{1}{36}$,

Probability of one double six in n throws $=\left(\frac{1}{36}\right)^n$,

Probability of no double six in n throws $=\left(\frac{35}{36}\right)^n$,

Probability of at least one double six in n throws $=1-\left(\frac{35}{36}\right)^n=p$,

For $n=24, p=1-\left(\frac{35}{36}\right)^{24}=0.4914039$.

It is interesting to note that in the above argument going from the probability of one double six in one trial to that of n trials we use the notion of *independence* to be defined later.

Using a statistical distribution for the case of *one* die, whose probability distribution is given in table 1.5, one can argue analogously as follows:

Table 1.5. *One die probability distribution*

outcomes	1	2	3	4	5	6
probabilities	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Probability of one six $= \left(\frac{1}{6}\right)$,

Probability of one six in n throws $= \left(\frac{1}{6}\right)^n$,

Probability of no six in n throws $= \left(\frac{5}{6}\right)^n$,

Probability of at least one six in n throws $= 1 - \left(\frac{5}{6}\right)^n = q$,

For $n = 4$, $q = 1 - \left(\frac{5}{6}\right)^4 = 0.5177469$.

The two probabilities $p = 0.4914039$ and $q = 0.5177469$ confirm de Mere's empirical regularity and there is no paradox of any kind! This clearly shows that de Mere's empirical frequencies were correct but his reasoning by analogy was faulty.

The chance regularity patterns of *unpredictability*, which we related to the probability concept of [2] *Independence* and that of sameness we related to [3] *homogeneity* using figure 1.1, are implicitly used throughout the exchange between Pascal and Fermat. It is interesting to note that these notions were not formalized explicitly until well into the 20th century. The probabilistic assumptions of Independence and homogeneity (Identical Distribution) underlay most forms of statistical analysis before the 1920s.

At this stage it is important to emphasize that the notion of probability underlying the probability distributions in tables 1.3–1.5, is one of *relative frequency* as used by de Mere to establish his regularity after a huge number of trials. There is nothing controversial about this notion of probability and the use of statistical models to discuss questions relating to games of chance, where the chance mechanism is explicitly an integral part of the phenomenon being modeled. It is not, however, obvious that such a notion of probability can be utilized in modeling other observable phenomena where the chance mechanism is not explicit.

1.2.3 Chance regularity in economic phenomena

In the case of the experiment of casting dice, the chance mechanism is explicit and most people will be willing to accept on faith that if this experiment is actually performed, the chance regularity patterns [1]–[3] noted above, will be present. The question which naturally arises is:

> Is this chance regularity conceivable in stochastic phenomena beyond games of chance?

In the case of stochastic phenomena where the chance mechanism is not explicit, we often:

(a) cannot derive a probability distribution a priori using some physical symmetry argument as in the case of dice or coins, and

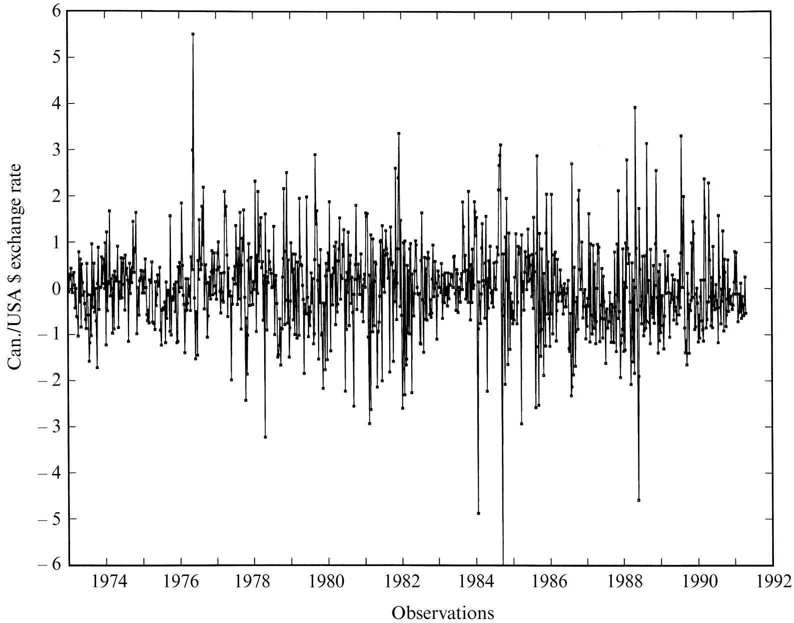

Figure 1.4 Changes in exchange rates data

(b) cannot claim the presence of any explicit chance mechanisms giving rise to the observations.

Using these observations our first task is to decide whether the underlying phenomenon can be profitably viewed as *stochastic* and our second task is to utilize the chance regularity patterns discerned in such data so as to choose an appropriate statistical model. Hence, discerning chance regularity patterns from data plots and relating them to the corresponding probability theory concepts will be a crucial dimension of the discussion that follows.

A number of observable phenomena in econometrics can be profitability viewed as stochastic phenomena and thus amenable to statistical modeling. In an attempt to provide some support for this proposition, consider the time-plot of X-log changes of the Canadian/USA dollar exchange rate, for the period 1973–1992 (weekly observations) shown in figure 1.4. What is interesting about the data is the fact that they do exhibit a number of *chance regularity* patterns very similar to those exhibited by the dice observations in figure 1.1, but some additional patterns are also discernible. The regularity patterns exhibited by both sets of observations are:

(a) the arithmetic average *over the ordering* (*time*) appears to be constant,
(b) the band of variation around the average appears to be relatively constant.

The regularity pattern in relation to a (possibly) stable relative frequencies law exhibited by the exchange rate data, do not suggest a triangular stable law as in figure 1.2. Instead:

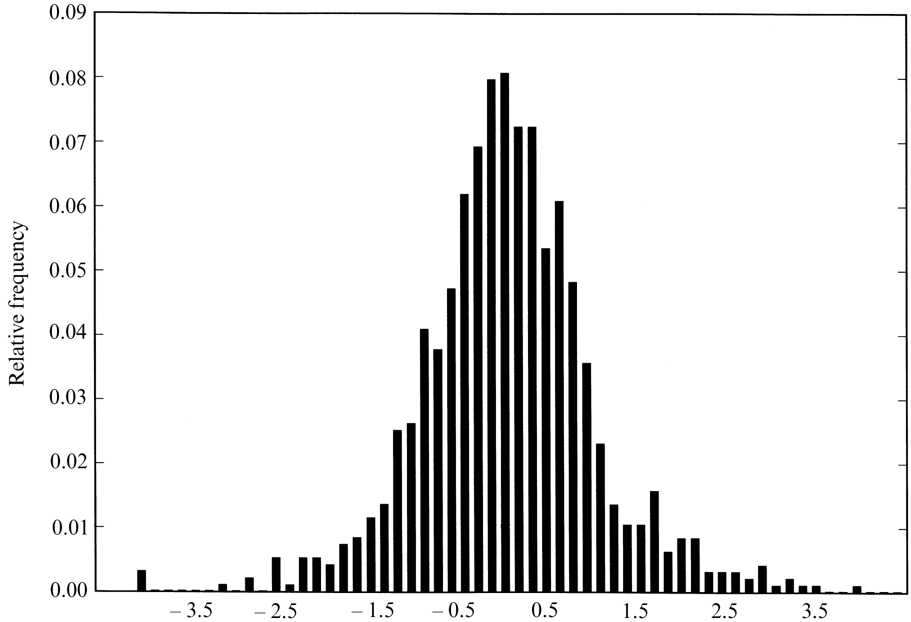

Figure 1.5 Histogram of exchange rates

(c) the data in figure 1.4 exhibit a certain bell-shaped symmetry (there seems to be as many points above the average as there are below but the relative frequencies die out as the value of X moves away from the center to the tails). This regularity can be seen in the graph of the relative frequencies given in figure 1.5.

How the graphs in figures 1.4 and 1.5 are related will be discussed extensively in chapter 5, together with a more detailed account of how one can recognize the patterns (a)–(c) mentioned above.

In addition to the regularity patterns encountered in figure 1.1, it is worth noting that the data in figure 1.4 exhibit the following regularity pattern:

(d) there seems to be a sequence of clusters of small changes and big changes succeeding each other.

At this stage the reader is unlikely to have been convinced that the features noted above are easily discernible from t-plots. However, an important dimension of modeling in this book is indeed how to *read* systematic information in data plots, which will begin in chapter 5.

In conclusion, the view adopted in this book is that **stochastic phenomena** (those exhibiting *chance regularity*) are susceptible to empirical modeling, irrespective of whether the built-in chance mechanism is apparent or not. Indeed, an important task for the modeler is to identify the observable phenomena which can be profitably viewed as stochastic phenomena. The question of whether there exists such a mechanism or not is only of metaphysical interest.

1.3 Chance regularity and statistical models

The discussion so far has identified the presence of chance regularity patterns in stochastic phenomena. Motivated by the desire to utilize the information conveyed by chance regularity patterns, probability theory proceeded to formalize them by developing (inventing) related (mathematical) probabilistic concepts; in the next few chapters we will introduce a number of probability theory concepts. In particular, the stable relative frequencies law regularity pattern will be formally related to the concept of a probability distribution; see tables 1.3–1.5. In the case of the exchange rate data the apparent stable relative frequencies law in figure 1.5 will be related to distributions such as the Normal and the Student's t, which exhibit the *bell-shaped symmetry* (see chapter 5). The unpredictability pattern will be formally related to the concept of Independence ([1]) and the sameness pattern to the Identical Distribution concept ([2]). The regularity patterns (a)–(b), exhibited by the exchange rate data, will be formally related to the concept of *stationarity* (see chapters 5 and 8), and (d) will be related to non-linear *dependence* (see chapter 6). It is important to emphasize that chance regularity patterns, such as those noted above, comprise the lifeblood of statistical modeling because their proper utilization constitutes the essence of empirical modeling.

The bridge between chance regularity patterns and probabilistic concepts, transforms the intuitive cognitive pattern recognition into **statistical (systematic) information**. In an attempt to render the utilization of the statistical systematic information easier for modeling purposes, the probabilistic concepts purporting to formalize the chance regularity patterns are placed into three broad categories:

(D) Distribution, (M) Dependence, and (H) Heterogeneity.

This basic taxonomy is designed to provide a logically coherent way to view and utilize statistical information for modeling purposes. These broad categories can be seen as defining the basic components of a statistical model in the sense that every statistical model can be seen as a smooth blend of ingredients from all three categories. The smoothness of the blend in this context refers to the internal consistency of the assumptions making up a statistical model. The *first* recommendation to keep in mind in empirical modeling is

1 A statistical model is just a set of (internally) compatible probabilistic assumptions from the three broad categories: (D), (M), and (H).

REMARK: to those knowledgeable readers who are not convinced that this is indeed the case, we mention in passing that distribution assumptions are sometimes indirect in the form of smoothness and existence of moments conditions; see chapter 10.

The statistical model chosen represents a description of a tentative chance mechanism with which the modeler attempts to capture the systematic information in the data (the chance regularity patterns). A statistical model differs from other types of models in so far as it specifies a situation, a mechanism or a process in terms of a certain **probabilistic**

structure, which will be formally defined in chapters 2–4. Mathematical concepts such as *a probability distribution*, *independence*, and *identical distribution* constitute forms of probabilistic structure. Indeed, the main objective of the first part of the book is to introduce many additional concepts which enable the modeler to specify a variety of forms of probabilistic structure, rich enough to capture, hopefully all, chance regularity patterns. The statistical model is specified exclusively in terms of such probabilistic assumptions designed to capture the systematic information in observed data.

The examples of casting dice, discussed above, are important not because of their intrinsic interest in empirical modeling but because they represent examples of a simple stochastic phenomenon which will play an important role in the next few chapters. The stochastic phenomenon represented by the above examples is referred to generically as a *random experiment* and will be used in the next three chapters (2–4) to motivate the basic structure of probability theory. The observable phenomenon underlying the exchange rate data plotted in figure 1.4 cannot be considered as a random experiment and thus we need to extend the probabilistic framework in order to be able to model such phenomena as well; this is the subject matter of chapters 6–8.

In view of the above discussion, successful empirical modeling has two important dimensions:

(a) recognize the chance regularity patterns as exhibited by the observed data, and
(b) capture these patterns by postulating appropriate statistical models.

The first requires a skill on behalf of the modeler to detect such patterns using a variety of graphical techniques. Indeed, it will be impossible to overestimate the importance of graphical techniques in empirical modeling. This brings us conveniently to the *second* recommendation in empirical modeling:

2 Graphical techniques constitute an indispensable tool in empirical modeling!

If we return momentarily to the data in table 1.2, there is no doubt that the reader will have a hard time recognizing any chance regularity patterns in the data set. A glance at data plots in figures 1.1 and 1.4 provide an overall picture of the structure of both data sets that would require more than a thousand words to describe. This merely confirms the natural perceptual and cognitive capacities of the human brain; humans are able to recognize, classify, and remember visual patterns much more efficiently than numbers or words. Chapter 5 brings out the interplay between chance regularity patterns and probabilistic concepts using a variety of graphical displays.

Capturing the statistical systematic information in the data presupposes a mathematical framework rich enough to model whatever patterns are detected. It is through probability theory that chance regularity has been charmed into compliance. In this sense the interplay between modeling and probability theory is not a one way street. For example, as late as the early 20th century the pattern of *dependence* was rather nebulous and as a consequence the corresponding mathematical concept was not as yet formalized. In view of this, there are no good reasons to believe that there are no chance regularity patterns which we cannot recognize at present but will be recognized in the future. As more patterns are detected, additional probabilistic assumptions will be devised in order to

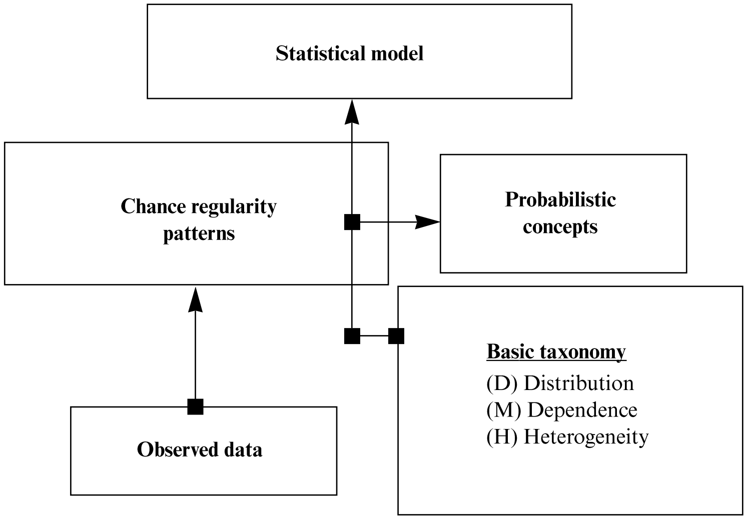

Figure 1.6 Chance regularity patterns, probabilistic assumptions, and a statistical model

formalize them and thus enrich probability theory as a modeling framework. Because of the importance of the interplay between observable patterns and formal probabilistic concepts, in figure 1.6 we present this relationship in a schematic way: chance regularity patterns are formalized in the form of probabilist concepts, these in turn are categorized into the basic taxonomy, and then utilized to postulate statistical models which (hopefully) capture the statistical systematic information; no effort will be spared in relating chance regularity patterns to the corresponding probabilistic concepts throughout this book.

The variety and intended scope of statistical models are constrained only by the scope of probability theory (as a modeling framework) and the training and the imagination of the modeler. There is no such thing as a complete list of statistical models which the modeler tries out in some sequence and chooses the one that looks the least objectionable. Moreover, empirical modeling is not about choosing optimal estimators (from some pre-specified menu), it is about choosing adequate statistical models; models which are *devised* by the modeler in an attempt to capture the systematic information in the data. In the discussion of statistical models in chapters 2–8 particular attention is paid to the relationship between observed data and the choice of statistical models. Some of the issues addressed in the next few chapters are:

(a) What do we mean by a statistical model?
(b) Why should statistical information be coded in a theory-neutral language?
(c) What information do we utilize when choosing a statistical model?
(d) What is the relationship between the statistical model and the features of the data?
(e) How do we recognize the statistical systematic information in the observed data?

We conclude this section by emphasizing the fact that the *statistical systematic information* in the observed data has to be coded in a language which is free from any economic theory concepts. Probability theory offers such a theory-neutral language which will be utilized exclusively in the specification of statistical models. As shown in chapters 6–7, statistical models as specified in this book, do not rely on any theory-based *functional forms* among variables of interest; instead they are specified exclusively in terms of statistical relationships based on purely statistical information. The codification of statistical models exclusively in terms of statistical information is of paramount importance because one of the primary objectives of empirical modeling is to assess the empirical validity of economic theories. This assessment can be thought of as a trial for the theory under appraisal, with the theoretical model as the main witness for the defence and the observed data as the main witness for the prosecution. For the data to be an unprejudiced witness, no judge (modeler) should allow coaching of the main prosecution witness by the defence, before the trial! Statistical information has to be defined exclusively in terms of concepts which are free from any economic-theoretical connotations; only then can observed data be viewed as an independent (and fair) witness for the prosecution. The *third* recommendation in empirical model is:

3 Do not allow the observed data to be coached a priori by the theory to be appraised.

The statistical model is viewed initially as a convenient summarization of the systematic information in the data which exists irrespective of any theory. The *fourth* recommendation in empirical modeling is:

4 Statistical model specification is guided primarily by the nature and structure of the observed data.

1.4 Statistical adequacy

As argued above, the success of empirical modeling is judged by how adequately the postulated statistical model captures the statistical systematic information contained in the data. A central theme of this book is that of **statistical adequacy** and how it can be achieved in practice, by utilizing several methods including graphical displays (see chapters 5–6) and misspecification testing (see chapter 15). Without a statistically adequate model which captures the systematic information in the data, no valid statistical inference is possible, irrespective of the sophistication and/or the potential validity of the theory!

Statistical inference is often viewed as the quintessential *inductive* procedure: using a set of data (specific) to derive conclusions about the stochastic phenomenon (general) that gave rise to the data (see figure 1.7). However, it is often insufficiently recognized that this inductive procedure is embedded in a fundamentally deductive premise. The procedure from the postulated model (the premise) to the inference results (estimation, testing, prediction, simulation) is *deductive*; no data are used to derive results on the optimality of estimators, tests, etc.; estimators and tests are pronounced *optimal* based

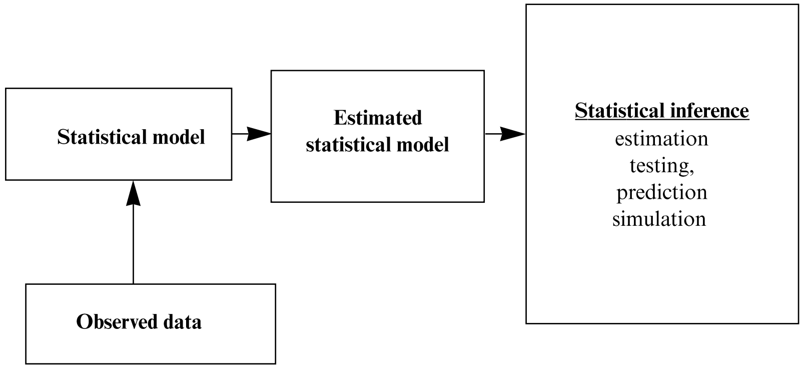

Figure 1.7 Statistical inference

on a purely deductive reasoning. The deductive component of the statistical inference reasoning amounts to:

if certain premises are assumed, certain conclusions necessarily follow.

More formally, if we denote the premises by p and the conclusions by q, then the above form of deductive reasoning takes the form of *modus ponens* (affirming the antecedent):

if p then q.

In this sense, statistical inference depends crucially on the validity of the premises: postulating a statistical model in the context of which the observed data are interpreted as a realization of the postulated stochastic mechanism. On the basis of this premise we proceed to derive statistical inference results using mathematical deduction. Correct deductive arguments show that if their premises are valid, their conclusions are valid. Using the observed data in question, the modeler relies on the validity of this deductive argument in order to draw general inference conclusions from specific data. However, if the premises are invalid the conclusions are generally unwarranted. In view of this, we consider the problem of assessing the validity of the postulated statistical model (misspecification testing) of paramount importance, especially in the case of observational data. The *fifth* recommendation in empirical modeling is:

5 No statistical inference result should be used to draw any conclusions unless the statistical adequacy of the postulated model has been established first.

The first and most crucial step in ensuring statistical adequacy is for the modeler to specify explicitly all the probabilistic assumptions making up the postulated model; without a complete set of probabilistic assumptions the notion of statistical adequacy makes no operational sense. For this reason the next several chapters pay particular attention to the problem of statistical model specification (probability and sampling models) to an extent that might seem unnecessary to a traditional textbook econometrician. It is emphasized at this stage that the notation, the terminology, and the various taxonomies introduced in the next four chapters play an important role in ensuring that

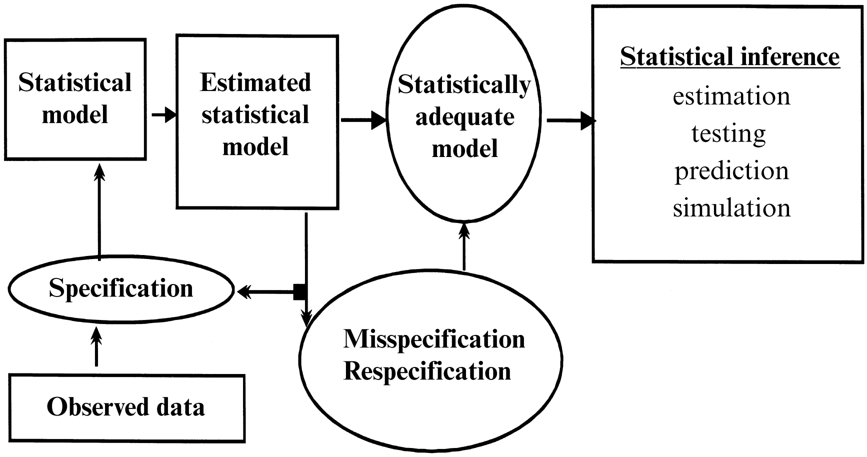

Figure 1.8 Statistical inference with statistical adequacy

the nature and structure of the probabilistic assumptions underlying the postulated model is made explicit and transparent to the modeler.

In the context of the probabilistic reduction approach, departures from the postulated statistical model are viewed as systematic information in the data that the postulated model does not account for. The statistical model needs to be respecified in order to account for the systematic information overlooked by the model postulated initially. Hence, the procedure in figure 1.7 is supplemented with the additional stages of misspecification testing and respecification. Figure 1.8 shows the modified procedure with the notion of a statistically adequate model coming between the estimated model and statistical inference. As shown in figure 1.8, reaching a statistically adequate model involves misspecification testing and respecification.

The notion of statistical adequacy is particularly crucial for empirical modeling because it can provide the basis for establishing *stylized facts* which economic theory will be required to account for. A cursory look at the empirical econometric modeling of the last 50 years or so will convince, even the most avid supporter of the traditional econometric approach, that it does not constitute a progressive research program because it has not led to any real accumulation of empirical evidence. Separating the statistical and theoretical models and ensuring the statistical adequacy of the former, will provide a good starting point for a progressive research strategy where empirical regularities are established by statistically adequate models (proper stylized facts) and theories are required to account for them. It is worth reiterating that in this book statistical and theoretical information are clearly distinguished in order to avoid any charges of circularity in implementing this research strategy.

1.5 Statistical versus theory information*

In an attempt to provide a more balanced view of empirical modeling and avoid any hasty indictments on behalf of traditional econometricians that "the approach adopted in this book ignores economic theory," this section will discuss briefly the role of economic theory in empirical modeling (see also Spanos (1986,1995b)).

Economic data are growing at an exponential rate but at the same time when a modeler attempts to give answers to specific questions he/she often finds that the particular data needed for the analysis do not exist in the form required. This is symptomatic of the absence of an adequate econometric methodology which would have played a coordinating role between economic theory and the appropriate observed data. More often than not, there exists a huge gap between theory-concepts and the data series that are usually available; the available data often measure something very different. As argued above this gap arises primarily because of the differences between the experimental-design circumstances assumed by economic theory, via the *ceteris paribus* clause, and the observational nature of the available data; the result of an on-going process with numerous influencing factors beyond the potential control of the modeler. The *sixth* recommendation in empirical modeling that one should keep in mind is:

6 Never assume that the available data measure the theory concept the modeler has in mind just because the names are very similar (or even coincide)!

A striking example is the theoretical concept *demand* versus the often available data in the form of *quantities transacted*; see Spanos (1995b). As a result of this gap, empirical modeling often attempts to answer theoretical questions of interest by utilizing data which contain no such information.

As argued in the previous three sections, the statistical systematic information is:

(a) related to the chance regularity patterns exhibited by the observed data,
(b) defined exclusively in terms of probabilistic concepts, and
(c) devoid (initially) of any economic theory connotations.

The clear distinction between statistical and theoretical systematic information constitutes one of the basic pillars of the empirical modeling methodology expounded in this book; see also Spanos (1986, 1995b, forthcoming). Theory and statistical models constitute distinct entities built on different information, the behavior of economic agents, and statistical systematic information, respectively. This constitutes a necessary condition for the statistical model to be used as an unprejudiced witness on the basis of whose testimony the empirical adequacy of the theory model can be assessed.

The theory influences the choice of an appropriate statistical model in two ways. First, the theory determines the choice of the observed data of interest. Although the choice of the observed data is theory laden, once chosen, the data acquire an objective existence which is theory free. The only further influence the theory has on the specification of the statistical model is that the latter should be general enough to allow the modeler to pose theoretical questions of interest in its context. Hence, the misspecification testing and respecification facets of empirical modeling have nothing to do with the theory model; they are purely statistical procedures determined by the notion of statistical information. The *seventh* recommendation in empirical modeling is:

7 No theory, however sophisticated, can salvage a misspecified statistical model.

As argued in chapter 7, the statistical and theory viewpoints provide very different viewing angles for modeling purposes. These viewing angles are complementary but they are often used as substitutes with dire consequences; see Spanos (1997a).

A statistically adequate model provides a good summary (description) of the statistical systematic information in the data but does not constitute the ultimate objective of empirical modeling. Ultimately, the modeler wants to assess the theory in terms of a statistically adequate model, as well as to synthesize the statistical and theory models in an attempt to bestow economic-theoretic meaning and explanatory capability to the statistical model. Hence, the *eighth* recommendation to keep in mind in empirical modeling is:

8 The success of empirical modeling is assessed by how skillfully the modeler can synthesize the statistical and theory models, without short-changing either the theoretical or the statistical information!

In order to distinguish between a statistical model, built exclusively in terms of statistical systematic information, and the synthesis of the theory and statistical models we call the latter an **econometric model** (see Spanos (1986)).

1.6 Observed data

In this section we will attempt a preliminary discussion of the constituent element of empirical modeling, the observed data. Certain aspects of the observed data play an important role in the choice of statistical models.

1.6.1 Early data

Numerical data have been collected for one reason or another since the dawn of history. Early data collections, however, were non-systematic and the collected information was not generally available. The systematic collection of economic data can be dated to the 17th century as a by-product of government activities such as tax and customs collection, spending and regulating, as well as the desire to quantify certain aspects of government activity (see Porter (1995)). For instance, earlier data on income distribution were simply a by-product of tax data. Towards the end of the 19th century special censuses were undertaken by (in particular the US) governments in the agricultural and manufacturing sectors in order to consider specific questions of interest (see Christ (1985)) Thus, it should come as no surprise to find out that the data used in the early empirical work in economics (early 20th century) were mostly data on exports, imports, production and price (see Stigler (1954, 1962)). Gradually, however, governments began to appreciate the use of such data in assessing economic performance as well as providing guideposts for