(e.g. Beck and Katz, 1995) and in sociology, finance and marketing (e.g. Keane, 1997). While restricting the focus of the book to basic topics may not do justice to the rapidly growing literature, it is nevertheless unavoidable in view of space limitations. Topics not covered in this book include duration models and hazard functions (see Heckman and Singer, 1985), as well as the frontier production function literature using panel data (e.g. Kumbhakar and Lovell, 2000; Koop and Steel, 2001), the literature on time-varying parameters, random coefficients and Bayesian models (e.g. Swamy and Tavlas, 2001; Hsiao, 2003), and the literature on nonparametric and semiparametric panels (e.g. Li and Racine, 2007).

1.2 WHY SHOULD WE USE PANEL DATA? THEIR BENEFITS AND LIMITATIONS

Hsiao (2003) lists several benefits of using panel data. These include the following:

(1) The use of panel data enables us to *control for individual heterogeneity*. Panel data suggest that individuals, firms, states or countries are heterogeneous. Time-series and cross-section studies that do not control for such heterogeneity run the risk of obtaining biased results; see, e.g., Moulton (1986, 1987). Let us demonstrate this with an empirical example. Baltagi and Levin (1986) considered panel data estimation of cigarette demand across 46 American states. Consumption is modelled as a function of lagged consumption, price and income; these variables vary across states and over time. There are, however, many other variables that affect consumption and which may be state-invariant or time-invariant. Let us call the state-invariant variables W_t and the time-invariant variables Z_i. Examples of Z_i are religion and education. For the religion variable, one might not be able to know the percentage of each state's population that is, say, Mormon for every year, but nor does one expect the percentage to change much over time. The same holds true for the percentage of each state's population that has completed high school or holds a college degree. Examples of W_t include advertising on TV and radio; such advertising is typically nationwide and does not vary across states. Some variables may be difficult to measure or hard to obtain, so not all possible Z_i and W_t variables will be available for inclusion in the consumption equation. Omission of such variables will lead to bias in the resulting estimates. By using panel data, one is better able to control for such state- or time-invariant variables, whereas a time-series study or cross-section study cannot. In fact, the data on cigarette demand show that Utah has less than half the average per capita consumption of cigarettes in the USA. This is because the population of Utah is mostly Mormon, and Mormonism prohibits smoking. Controlling for Utah in a cross-section regression can be done with a dummy variable which has the effect of removing that state's observation from the regression. This would not be the case for panel data, as we will shortly discover. In fact, with panel data, one might first difference the data to get rid of all Z_i-type variables and hence effectively control for all state-specific characteristics. This holds whether the Z_i variables are observable or not. Alternatively, the dummy variable for Utah controls for every state-specific effect that is distinctive of Utah without omitting the observations for Utah.

Another example is given by Hajivassiliou (1987), who studied the external debt repayments problem using a panel of 79 developing countries observed over the period 1970–1982. The countries in the study differ in terms of their colonial history, financial institutions, religious affiliations and political regimes. All of these country-specific variables affect the attitudes of the countries with regard to borrowing and defaulting and the way they are treated by the lenders. Not accounting for this country heterogeneity would cause serious misspecification.

Deaton (1995) gives another example, arising from agricultural economics. This pertains to the question of whether small farms are more productive than large ones. Ordinary least squares (OLS) regressions of yield per hectare on inputs such as land, labour, fertilizer use, farmer's education, etc. usually find that the sign of the estimate of the land coefficient is negative, implying that smaller farms are more productive. Some explanations from economic theory have argued that higher output per head is an optimal response to uncertainty by small farmers, or that hired labour requires more monitoring than family labour. Deaton (1995) offers an alternative explanation: the regression analysis suffers from the omission of unobserved heterogeneity, in this case "land quality", and the omitted variable is systematically correlated with the explanatory variable (farm size). In fact, farms in low-quality marginal areas (semi-desert) are typically large, while farms in high-quality land areas are often small. Deaton argued that while gardens generate more value-added per hectare than a sheep station, this does not imply that sheep stations should be organized as gardens. In this case, differencing may not resolve the "small farms are productive" question, since farm sizes will usually show little or no change over short periods.

(2) Panels give *more informative data, more variability, less collinearity among the variables, more degrees of freedom, and more efficiency*. Time-series studies are plagued with multicollinearity; for instance, in the above example about cigarette demand, there is high collinearity between price and income in the aggregate time series for the USA; this is less likely with a panel across American states, since the cross-section dimension adds a lot of variability, yielding more informative data on price and income. In fact, the variation in the data can be decomposed into variation between states of different sizes and characteristics, and variation within states. The former variation is usually larger. With additional, more informative data, one can obtain more reliable parameter estimates. Of course, the same relationship has to hold for each state, i.e. the data have to be poolable. This is a testable assumption and one that we shall tackle in due course.

(3) With panel data, one is better able to study the *dynamics of adjustment*. Cross-sectional distributions that look relatively stable can hide a multitude of changes. Spells of unemployment, job turnover, or residential and income mobility are better studied with panels. Panel data are also well suited for studying the duration of economic states such as unemployment and poverty, and if the panels are long enough, they can shed light on the speed of adjustments to economic policy changes. For example, in measuring unemployment, cross-sectional data can be used to estimate what proportion of the population is unemployed at a given point in time; repeated cross-sections can then show how this proportion changes over time. However, only panel data can provide estimates of the proportion unemployed in one period who remain unemployed in another period. Important policy questions, such as determining whether families' experiences of poverty, unemployment and welfare dependence are transitory or chronic, necessitate the use of panels. Deaton (1995) argued that, unlike cross-sections, panel surveys yield data on *changes* for individuals or households. Panel data allow us to observe *how* individual living standards change during the development process, and enable us to determine *who* is benefiting from development. Panel surveys also allow us to observe whether poverty and deprivation are transitory or long-lived, i.e. the income-dynamics question. Furthermore, panels are necessary for the estimation of intertemporal relations and the construction of life-cycle and intergenerational models. In fact, panels can relate the individual's experiences and behaviour at one point in time to other experiences and behaviour at another point in time. For

example, in evaluating training programs, a group of participants and non-participants are observed before and after implementation of the training program. Such a panel involving at least two time periods forms the basis for the "difference in differences" estimator; see Chapter 2.

(4) Panel data are more suitable for *identifying and measuring effects that are simply not detectable in pure cross-section or pure time-series data.* Suppose that we have a cross-section of women with a 50% average yearly labour force participation rate. This might be due to (a) each woman having a 50% chance of being in the labour force in any given year, or (b) 50% of the women working all the time and 50% not at all. Case (a) has high turnover, while case (b) has no turnover. Only panel data could discriminate between the two cases. Another example is the determination of whether union membership increases or decreases wages. To answer this question, it is better to observe individual workers moving from union to non-union jobs or vice versa. Holding the individual's characteristics constant, we would be better equipped to determine whether union membership affects wages and, if so, by how much. This kind of analysis extends to the estimation of other types of wage differentials, holding individuals' characteristics constant – for example, the estimation of wage premiums paid in dangerous or unpleasant jobs.

Economists studying workers' level of satisfaction run into the problem of anchoring in a cross-section study; see Chapter 11 of Winkelmann and Winkelmann (1998). Such a survey usually asks the question "How satisfied are you with your life?", with responses scored on a scale from 0, meaning completely dissatisfied, to 10, meaning completely satisfied. The problem is that each individual anchors their scale at a different level, rendering interpersonal comparisons of responses meaningless. However, in a panel study, where the metric used by each individual is time-invariant over the period of observation, one can avoid this problem by using a difference (or fixed effects) estimator, which will make inference based only on intra- rather than interpersonal comparisons of satisfaction.

(5) Panel data models allow us to *construct and test more complicated behavioural models than do pure cross-section or time-series data.* For example, technical efficiency is better studied and modelled using panels (see Kumbhakar and Lovell, 2000; Koop and Steel, 2001).

(6) Micro panel data gathered on individuals, firms and households can be measured more accurately than similar variables measured at the macro level. Biases resulting from aggregation over firms or individuals may be reduced or eliminated.

(7) Macro panel data, on the other hand, have longer time series and, as we shall see in Chapter 12, panel unit root tests have standard asymptotic distributions and do not suffer from the problem of nonstandard distributions encountered with unit root tests in time-series analysis.

Limitations of panel data include:

(1) *Design and data collection problems.* For an extensive discussion of problems that arise in designing panel surveys, as well as data collection and data management issues, see Kasprzyk *et al.* (1989). These include problems of coverage (incomplete account of the population of interest), nonresponse (due to lack of cooperation of the respondent or interviewer error), recall (respondent not remembering correctly), frequency of interviewing, interview spacing, reference period, the use of bounding, and time-in-sample bias.[1]

(2) *Distortions of measurement errors.* Measurement errors may arise because of faulty responses due to unclear questions, memory errors, deliberate distortion of responses (e.g. prestige bias), inappropriate informants, misrecording of responses, and interviewer effects (see Kalton, Kasprzyk and McMillen, 1989). The validation study by Duncan and Hill (1985) on the PSID illustrates the significance of the measurement error problem. They compared

the responses of employees of a large firm with the records of the employer, and found small response biases except in the case of work hours, which are overestimated. The ratio of measurement error variance to the true variance was found to be 15% for annual earnings, 37% for annual work hours, and 184% for average hourly earnings. These figures are for a one-year recall, i.e. 1983 for 1982, and become more than doubled with two years' recall. Brown and Light (1992) investigated the inconsistency in job tenure responses in the PSID and NLS. Cross-section data users have little choice but to believe the reported values of tenure (unless they have external information), while users of panel data can check for inconsistencies in tenure responses with elapsed time between interviews; for example, a respondent may claim to have three years of tenure in one interview and a year later claim six years. This should alert the user of the panel to the presence of measurement error. Brown and Light (1992) showed that failure to use internally consistent tenure sequences can lead to misleading conclusions about the slope of wage–tenure profiles. Section 10.1 deals with measurement error in panel data.

(3) *Selectivity problems*. These include:

(a) *Self-selectivity*. If people choose not to work because the reservation wage is higher than the offered wage, in this case we would observe the characteristics of the individuals but not their wage. Since only their wage is missing, the sample is censored. However, if we do not observe all data on these people, this would be a truncated sample. An example of truncation is the New Jersey negative income tax experiment: we are only interested in poverty, and people with income higher than 1.5 times the poverty level are dropped from the sample. Inference from this truncated sample introduces bias that is not helped by more data, because of the truncation (Hausman and Wise, 1979). Chapter 11 deals with selectivity problems in panel data.

(b) *Nonresponse*. This can occur at the initial wave of the panel due to refusal to participate, nobody at home, untraced sample unit, and other reasons. Item (or partial) nonresponse occurs when one or more questions are left unanswered or are found not to provide a useful response. Complete nonresponse occurs when no information is available from the sampled household. Besides the efficiency loss due to missing data, such nonresponse can cause serious identification problems for the population parameters. The seriousness of the problem is directly proportional to the amount of nonresponse. Nonresponse rates in the first wave of the European panels varied from 10% in Greece and Italy, where participation was compulsory, to 52% in Germany and 60% in Luxembourg. The overall nonresponse rate was 28%; see Peracchi (2002). The comparable nonresponse rate for the first wave of the PSID was 24%, for the BHPS 26%, for the GSOEP 38%, and for PSELL 35%.

(c) *Attrition*. While nonresponse occurs also in cross-section studies, it is more of a serious problem in panels, because subsequent waves of the panel are still subject to nonresponse. Respondents may die, move, or find that the cost of responding is too high; see Chapter 11 for a discussion of the consequences of attrition in panels. The degree of attrition varies depending on the panel studied; see Kalton, Kasprzyk and McMillen (1989) for several examples. In general, the overall rates of attrition increase from one wave to the next, but the rate of increase declines over time. Becketti *et al.* (1988) studied the representativeness of the PSID 14 years after its start. They found that only 40% of those originally in the sample in 1968 remained in the sample in 1981. Nevertheless, they did find that, as far as the dynamics of entry and exit are concerned, the PSID is still representative. The potentially most damaging threat to the value of panel data is the presence of biasing attrition. Fitzgerald, Gottschalk and Moffit (1998) reported 51% attrition of the original PSID sample by 1989, with the major reasons being family unit nonresponse, death, or a residential move. Attritors were found to have lower earnings, lower education levels, and lower marriage propensities. But despite the large amount of attrition, Fitzgerald, Gottschalk and Moffit (1998) found no strong evidence that it had seriously distorted the representativeness of the PSID through 1989. In the same vein of research, Lillard and Panis (1998) found evidence of significant selectivity in attrition for the PSID; for example, they found that less educated individuals and older people are more likely to drop out, whereas married people are more likely to continue. Propensity to participate in the survey diminishes with increasing duration of the respondent in the sample. Despite this, the effects of ignoring such selective attrition on household income dynamics, marriage formation and dissolution, and adult mortality risk are mild. In Europe, the comparable attrition rates (between the first and second waves) vary from 6% in Italy to 24% in the UK; the average attrition rate was about 10%. For the BHPS, attrition from the first to the second wave was 12%; for PSELL it was 15%. For the GSOEP, attrition was 12.4% for the West German sample and 8.9% for the East German sample; see Peracchi (2002). In order to counter the effects of attrition, rotating panels are sometimes used, where a fixed percentage of respondents are replaced in every wave to replenish the sample. More details on rotating and pseudo panels can be found in Chapter 10. A special issue of the *Journal of Human Resources*, Spring 1998, is dedicated to attrition in longitudinal surveys.

(4) *Short time-series dimension.* Typical micro panels involve annual data covering a short time span for each individual. This means that asymptotic arguments rely crucially on the number of individuals tending to infinity. Increasing the time span of the panel is not without cost either. In fact, this increases the chances of attrition and increases the computational difficulty for limited dependent variable panel data models (see Chapter 11).

(5) *Cross-section dependence.* Macro panels on countries or regions with long time series that do not account for cross-country dependence may lead to misleading inference. In Chapter 12 it is shown that several panel unit root tests suggested in the literature assume cross-section independence. Accounting for cross-section dependence turns out to be important and affects inference. Alternative panel unit root tests have been proposed that account for such dependence. Chapter 13 surveys tests for cross-sectional dependence in panels.

Panel data is not a panacea and will not solve all the problems that a time-series or cross-section study could not handle. Examples are given in Chapter 12 where we cite econometric studies arguing that panel data will yield more powerful unit root tests than individual time-series. This, in turn, should help shed more light on the purchasing power parity (PPP) and growth convergence questions. In fact, this led to a flurry of empirical applications, along with objections from some sceptics who argued that panel data did not really solve the PPP or growth convergence problem; see Maddala (1999), Maddala, Wu and Liu (2000), and Banerjee, Marcellino and Osbat (2004, 2005). Collecting panel data is quite costly, and there is always the question of how often one should interview respondents. Deaton (1995) argues that economic development is far from instantaneous, and so changes from one year to the next are probably too noisy and too short-term to be really useful. He concludes that the payoff for panel data is over long time periods, such as five years, ten years, or even longer. In contrast, for health and nutrition issues, especially those of children, one could argue the opposite case, i.e. panels with a shorter time span are necessary for monitoring the health and development of children.

Moffit (1998) reported 51%
or reasons being family unit
und to have lower earnings,
ut despite the large amount
d no strong evidence that it
through 1989. In the same
of significant selectivity in
ucated individuals and older
are more likely to continue.
creasing duration of the re-
g such selective attrition on
olution, and adult mortality
etween the first and second
rage attrition rate was about
vave was 12%; for PSELL it
st German sample and 8.9%
der to counter the effects of
d percentage of respondents
etails on rotating and pseudo
ournal of Human Resources,
ys.
annual data covering a short
uments rely crucially on the
an of the panel is not without
increases the computational
Chapter 11).
or regions with long time
lead to misleading inference.
ested in the literature assume
ndence turns out to be impor-
e been proposed that account
al dependence in panels.
ns that a time-series or cross-
12 where we cite econometric
ot tests than individual time-
asing power parity (PPP) and
empirical applications, along
did not really solve the PPP or
/u and Liu (2000), and Baner-
a is quite costly, and there is
ndents. Deaton (1995) argues
changes from one year to the
ul. He concludes that the pay-
rs, ten years, or even longer.
children, one could argue the
for monitoring the health and

This book will make the case that panel data offer several advantages worth their cost. However, as Zvi Griliches argued about economic data in general, the more we have of it, the more we demand of it. The economist using panel data, or any data for that matter, has to know their limitations.

NOTE

1. Bounding is used to prevent the shifting of events from outside the recall period into the recall period. Time-in-sample bias is observed when a significantly different level for a characteristic occurs in the first interview than in later interviews, when one would expect the same level.