1
Introduction

> The increased availability of panel data from household surveys has been one of the most important developments in applied social research in the last thirty years.
>
> Fitzgerald, Gottschalk and Moffitt (1998, p. 252)

1.1 PANEL DATA: SOME EXAMPLES

In this book, the term "panel data" refers to the pooling of observations on a cross-section of households, countries, firms, etc. over several time periods. This can be achieved by surveying a number of households or individuals and following them over time. The data collected from surveying individuals are known as *micro panels* and are collected for a large number N of individuals (usually in the hundreds or thousands) over a short time period T (which may vary from a minimum of two years to a maximum rarely exceeding 10 or 20 years). In contrast, *macro panels* usually involve data collected for a number of countries over time. A macro panel may have a moderately sized N (varying from, say, seven countries of the G7 to a larger set of, say, 20 OECD or European Union countries, or a mix of developed and developing countries, usually between 100 and 200 of them), but the data are usually observed annually over 20 to 60 years. Micro and macro panels require different econometric treatment. For example, the asymptotics for micro panels have to be for large N and fixed T, whereas the asymptotics for macro panels can be for large N and large T. Also, with a long time series for macro panels, one has to deal with issues of nonstationarity in the time series, such as unit roots, structural breaks and cointegration; see Chapter 12. In contrast, for micro panels one does not need to be concerned with nonstationarity issues, since T is short for each individual or household surveyed. Another example of the different treatment required is that in macro panels one has to deal with cross-country dependence, but this is not usually an issue with micro panels where the individuals or households are randomly sampled and hence not likely to be correlated. Chapter 13 studies spatial dependence in panel data as a simple way to model externalities and spillovers across cross-sectional units.

1.1.1 Examples of Micro Panels

Two well-known examples of US micro panel data are the Panel Study of Income Dynamics (PSID) collected by the Institute for Social Research at the University of Michigan (http://psidonline.isr.umich.edu) and the National Longitudinal Surveys (NLS), a set of surveys sponsored by the Bureau of Labor Statistics (http://www.bls.gov/nls/home.htm).

The PSID began in 1968 with 4800 families and grew to encompass more than 7000 families in 2001. By 2003, the PSID had collected information on more than 65 000 individuals spanning as much as 36 years of their lives. Annual interviews were conducted from 1968 to 1996. In 1997, this survey was redesigned for biennial data collection. In addition, the core sample was reduced and a refresher sample of post-1968 immigrant families and their

adult children was introduced. The central focus of the data is economic and demographic. The list of variables includes income, poverty status, public assistance in the form of food or housing, other financial matters (e.g. taxes, interhousehold transfers), family structure and demographic measures, labour market work, housework time, housing, geographic mobility, socioeconomic background, and health. Other supplemental topics include housing and neighbourhood characteristics, achievement motivation, child care, child support and child development, job training and job acquisition, retirement plans, health, kinship, wealth, education, military combat experience, risk tolerance, immigration history, and time use.

The NLS, on the other hand, consists of surveys designed to gather information at multiple points in time on the labour market activities and other significant life events of several groups of men and women:

(1) The NLSY97 consists of a nationally representative sample of approximately 9000 youths who were 12–17 years old as of 1997.
(2) The NLSY79 consists of a nationally representative sample of 12 686 young men and women who were 14–22 years old in 1979. These individuals were interviewed annually through 1994 and are currently interviewed on a biennial basis.
(3) The NLSY79 children and young adults survey includes the biological children born to women in the NLSY79.

The list of variables includes information on schooling and career transitions, marriage and fertility, training investments, child care usage, and drug and alcohol use. A large number of studies have made use of the NLS and PSID data sets. The PSID applications cover a wide range of topics, including: intertemporal models of labour supply; wages and employment over the business cycle; unemployment, job turnover and labour mobility; consumption, income and balance sheet dynamics; extended family behaviour; poverty, welfare and income dynamics; intergenerational transmission of economic status; and antecedents of economic and demographic events.

Panels can also be constructed from the Current Population Survey (CPS), a monthly national survey of about 50 000 households in the USA conducted by the Bureau of Census for the Bureau of Labor Statistics (http://www.census.gov/cps). This survey has been ongoing for more than 50 years. The CPS is the primary source of information on the labour force characteristics of the US population. Compared with the NLS and PSID data, the CPS contains fewer variables, spans a shorter time period and does not follow movers. However, it covers a much larger sample and is representative of all demographic groups. The CPS provides estimates of employment, unemployment, earnings, hours of work, and other indicators. These estimates are available by a variety of demographic characteristics, including age, sex, race, marital status, and educational attainment; they are also available by occupation, industry, and class of worker.

Another important source of household survey data for developing countries is the World Bank's Living Standards Measurement Study (LSMS, http://www.worldbank.org/LSMS), which was established in 1980. Since 1985, LSMS has conducted surveys in about 20 developing countries from Albania to Vietnam. These tend to involve small samples, in the order of 2000 to 5000 households. In some countries it could be one survey or multiple surveys. In other countries it could be a two- to four-year panel. Three types of questionnaires were used: a household, a community, and a price questionnaire. In some cases, a school or health facility questionnaire was added. The LSMS data have focused mostly on documenting regularities concerning the nature of poverty. Repeated surveys like the LSMS, even if they

mic and demographic.
ce in the form of food
s), family structure and
z, geographic mobility,
ude housing and neigh-
ipport and child devel-
hip, wealth, education,
me use.

information at multiple
vents of several groups

oximately 9000 youths

12 686 young men and
e interviewed annually

ogical children born to

ansitions, marriage and
use. A large number of
plications cover a wide
vages and employment
ility; consumption, in-
ty, welfare and income
itecedents of economic

y (CPS), a monthly na-
e Bureau of Census for
y has been ongoing for
i the labour force char-
data, the CPS contains
overs. However, it cov-
ups. The CPS provides
l other indicators. These
including age, sex, race,
ccupation, industry, and

; countries is the World
worldbank.org/LSMS),
urveys in about 20 de-
: small samples, in the
survey or multiple sur-
types of questionnaires
some cases, a school or
mostly on documenting
the LSMS, even if they

might not constitute a genuine panel, can be used to construct a *pseudo panel*, as we will see in Chapter 10.

Although the US panels started in the 1960s, it was not until the 1980s that the European panels began setting up. In 1989, a special section of the *European Economic Review* published papers based on the German Socio-Economic Panel, the Swedish study of household market and nonmarket activities, and the Intomart Dutch panel of households. The first wave of the German Socio-Economic Panel (GSOEP, http://www.diw.de/soep) data was collected by the DIW (German Institute for Economic Research, Berlin) in 1984 and included 5921 West German households, involving 12 290 respondents. Standard demographic variables and information on wages, income, benefit payments, level of satisfaction with various aspects of life, hopes and fears, political involvement, etc. were collected. In 1990, after German reunification, 4453 adult respondents in 2179 households from East Germany were also included in the GSOEP. The attrition rate has been relatively low in GSOEP. Wagner, Burkhauser and Behringer (1993) reported that through eight waves of the GSOEP, 54.9% of the original panel respondents have records without missing years. The British Household Panel Survey (BHPS), an annual survey of private households in Britain first collected in 1991 by the Institute for Social and Economic Research at the University of Essex (https://www.iser.essex.ac.uk/bhps), comprises a national representative sample of some 5500 households and 10 300 individuals drawn from 250 areas of Great Britain. Additional samples of 1500 households each from Scotland and Wales were added to the main sample in 1999, and in 2001 a sample of 2000 households from Northern Ireland was added as well. The data collected include demographic and household characteristics, as well as information on household organization, labour market activity, health, education, housing, consumption, income, and social and political values. The Swedish Panel Study of Market and Non-market Activities (HUS) was conducted in 1984, 1986, 1988, 1991, 1993, 1996 and 1998 (http://www.nek.uu.se/faculty/klevmark/hus.htm), collecting data from 2619 individuals on child care, housing, market work, income and wealth, tax reform (1993), willingness to pay for a good environment (1996), local taxes, public services, and activities in the black economy (1998).

The European Community Household Panel (ECHP) is centrally designed and coordinated by the Statistical Office of the European Communities (EuroStat); see Peracchi (2002). The first wave was conducted in 1994 and included all members of the EU at that time except Austria, Finland and Sweden. Austria joined in 1995, Finland in 1996, and data for Sweden has been obtained from the Swedish Living Conditions Survey. The project was launched to obtain comparable information across member countries on income, work and employment, poverty and social exclusion, housing, health, and many other diverse social indicators of living conditions of private households and persons. The ECHP was linked from the beginning to existing national panels (e.g. those of Belgium and Holland) or ran parallel to existing panels with similar content, such as GSOEP, PSELL (Luxembourg's Panel Socio-Economique Liewen zu Lëtzebuerg) and the BHPS.

Other panel studies include:

- the Canadian Survey of Labor Income Dynamics (SLID), collected by Statistics Canada (www.statcan.gc.ca), which includes a sample of approximately 37 000 households located throughout all ten provinces. Years available are 1993–2000.
- the Japanese Panel Survey on Consumers (JPSC), collected in 1994 by the Institute for Research on Household Economics (www.kakeiken.or.jp). This consists of a nationally

representative sample of 1500 women aged between 24 and 34 years in 1993 (cohort A). In 1997, 500 women were added whose ages were between 24 and 27 (cohort B). Years available are 1994–2000. Information gathered includes family composition, labour market behaviour, income, consumption, savings, assets, liabilities, housing, consumer durables, household management, time use, and satisfaction.

- the Russian Longitudinal Monitoring Survey (RLMS), collected in 1992 by the Carolina Population Center at the University of North Carolina (http://www.cpc.unc.edu/projects/rlms-hse). This is a nationally representative household survey designed to measure the effects of Russian reforms on economic well-being. Variables include individual health and dietary intake, measurement of expenditures and service utilization, and community-level data such as region-specific prices and information on community infrastructure.
- the Korea Labor and Income Panel Study (KLIPS), available for 1998–2001. This is a survey of 5000 households and their members from seven metropolitan cities and urban areas in eight provinces (http://www.kli.re.kr/klips/en/about/introduce.jsp).
- the Household, Income and Labour Dynamics in Australia (HILDA), a household panel survey whose first wave was conducted by the Melbourne Institute of Applied Economic and Social Research in 2001 (http://www.melbourneinstitute.com/hilda/). This survey includes 7682 households with 19 914 individuals from 488 different neighbouring regions across Australia.
- the Indonesia Family Life Survey (IFLS), available for 1993/94, 1997/98 and 2000. The sample is representative of about 83% of the Indonesian population and contains over 30 000 individuals living in 13 of the 26 provinces. In 1993, 7224 households were interviewed, and over 7700 households were interviewed in 2000 (http://www.rand.org/labor/FLS/IFLS.html).

The above list of panel data sets, though by no means exhaustive, provides a good selection of panel data sets readily accessible for economic research.

1.1.2 Examples of Macro Panels

Several sources of macro panel data commonly utilized by economists are the following:

(i) The Penn World Table (PWT), available at www.nber.org, provides purchasing power parity and national income accounts converted to international prices for 188 countries for some or all of the years from 1950 to 2004. In addition, the European Union and the OECD provide detailed purchasing power and real product estimates for their countries, and the World Bank gives current price estimates for most PWT countries at the GDP level.

(ii) The World Bank is a great source of macro panels (http://data.worldbank.org), such as the World Development Indicators (WDI). For example, the 2007 WDI includes more than 900 indicators for 152 economies with populations of more than 1 million.

(iii) The International Monetary Fund (www.imf.org) provides several sources of macro panel data. These include: the World Economic Outlook Databases, which provide time-series data on GDP growth, inflation, unemployment, payments balances, exports, imports, external debt, capital flows, commodity prices, and so on; the International Financial Statistics, which provide approximately 32 000 time series covering more than 200 countries starting in 1948 and containing information about exchange rates, fund

rs in 1993 (cohort A).
l 27 (cohort B). Years
osition, labour market
g, consumer durables,

1992 by the Carolina
.cpc.unc.edu/projects/
signed to measure the
lude individual health
tion, and community-
ity infrastructure.
1998–2001. This is a
litan cities and urban
ce.jsp).
A), a household panel
of Applied Economic
ilda/). This survey in-
neighbouring regions

l, 1997/98 and 2000.
population and con-
ces. In 1993, 7224
interviewed in 2000

vides a good selection

are the following:

ides purchasing power
rices for 188 countries
uropean Union and the
ates for their countries,
countries at the GDP

orldbank.org), such as
l7 WDI includes more
than 1 million.
eral sources of macro
abases, which provide
ents balances, exports,
on; the International Fi-
ies covering more than
t exchange rates, fund

---

accounts, and the main global and country economic indicators; the Direction of Trade Statistics year book, which provides seven years of trade data for about 186 countries, with quarterly data covering the most recent six quarters and the latest year for about 156 countries. Data are also available on balance of payments and international investment positions, as well as indices of primary commodity prices. In addition, the International Monetary Fund provides member countries' data on international reserves and foreign currency liquidity, as well as Financial Soundness Indicators.

(iv) The United Nations provides a wealth of macro country panel data (http://unstats.un.org/unsd/economic_main.htm), including information on national accounts, trade, and industry statistics.

(v) The OECD provides data on its website at www.oecd.org.

(vi) The European Central Bank (ECB) provides data on European Union member countries at www.ecb.int.

(vii) The Central Intelligence Agency's World Factbook, available at https://www.cia.gov/library/publications/the-world-factbook/, is another rich source of country data.

These are but a few of the agencies that provide macro data on individual countries over time, which can be pooled and used in panel studies. We shall study several special types of panel data encountered in practice, such as unbalanced panels in Chapter 9, nested panels in Section 9.6, unequally spaced panels in Section 5.2.5, rotating panels in Section 10.2, pseudo panels in Section 10.3, spatial panels in Chapter 13, count panels in Section 10.6, and heterogeneous panels in Section 10.5.

1.1.3 Some Basic References

Virtually every graduate text in econometrics contains a chapter or a major section on the econometrics of panel data. Recommended readings on this subject include the Econometric Society monograph *Analysis of Panel Data* by Hsiao (2003), along with two chapters in the *Handbook of Econometrics*: Chapter 22 by Chamberlain (1984) and Chapter 53 by Arellano and Honoré (2001). Maddala (1993) edited two volumes collecting some of the classic articles on the subject. This collection was updated with two more volumes covering the period 1992–2002, edited by Baltagi (2002). Other books on the subject include Arellano (2003), Wooldridge (2010), and Mátyás and Sevestre (2008). Some survey papers on non-stationary panel models are Baltagi and Kao (2000), Choi (2006), and Breitung and Pesaran (2008). Special issues of journals dedicated to panel data include a 1995 special issue of the *Journal of Econometrics* edited by Baltagi, two volumes of the *Annales d'Économie et de Statistique* edited by Sevestre (1999), a special issue of the *Oxford Bulletin of Economics and Statistics* edited by Banerjee (1999), two special issues (year 2000, volume 19, issues 3 and 4) of *Econometric Reviews* edited by Maasoumi and Heshmati, a special issue of *Advances in Econometrics* edited by Baltagi, Fomby and Hill (2000), two special issues of *Empirical Economics* edited by Baltagi (2004) and by Baltagi and Breitung (2011), a special issue of the *Journal of Applied Econometrics* edited by Baltagi and Pesaran (2007) and, more recently, a special issue of *Econometric Reviews* edited by Baltagi and Maasoumi (2013).

The objective of this book is to provide a simple introduction to some of the basic issues of panel data analysis. It is intended for economists and social scientists with the usual background in statistics and econometrics. Panel data methods have been used in political science