
Introduction to the *JHR's* Special Issue on Cross-National Comparative Research Using Panel Surveys

James P. Smith, Frank Stafford, and James R. Walker

I. Introduction

Until recently, international comparative research had a poor reputation, especially in economics. Comparative international research was seen as overly descriptive and typically based on simple contrasts of aggregate national statistics. When cross-sectional microdata were used, the surveys that did exist in many countries did not contain the range and depth of variables to which the scientific research community had become accustomed. Part of this bad reputation also stemmed from the fact that, using the United States as the benchmark, comparable panel surveys either did not exist in many countries or were viewed as being of much lower quality. The end result was that scholars from all countries would tend to test their models only using the best U.S. data. It was also felt by many foreign scholars that unless they used U.S. panel data they would have great difficulty in publishing in the better-known scientific journals.

We believe that a combination of circumstances is about to radically change this situation. These circumstances start by recognizing that comparative international research has some unique analytical advantages for testing the effects of important policies common to many countries. In addition, the number and quality of international panel surveys are improving rapidly and in many dimensions will exceed those of their U.S. counterparts. Finally, many of the best graduate students trained in U.S. institutions were foreign nationals who are now some of the leading scholars in the best U.S. departments. It is not surprising that these scholars would want to test their ideas using data from their home countries.

With this in mind, a conference on comparative research using international panel surveys was held in Ann Arbor, Michigan, on October 26 and 27, 2000. The conference was intended as a vehicle to encourage researchers in the social sciences to use panel surveys to address critical scientific and policy issues that would be better informed by international comparisons and the variation in policy environments across countries. The topics of interest for this conference spanned substantive and methodological issues relevant to the social sciences and that make use of unique features of specific longitudinal data sets in a cross-national context.

This conference was sponsored by the Board of Overseers of the Panel Study of Income Dynamics (PSID), the Technical Review Committee of the National Longitudinal Surveys (NLS), and the Health and Retirement Study (HRS). Funding was

provided by the National Science Foundation, the Bureau of Labor Statistics, the National Institute on Aging, and the University of Michigan's Institute for Social Research. Fifteen scientific papers were presented at this conference and the papers included in this volume represent a selected subset of those papers.

II. A Rationale for Comparative International Panel Research

Why do international comparative research with panel data? The fundamental value of any national panel must rest in its ability to facilitate first-rate scientific and policy research within its own borders without any cross-national comparative component. That said, we believe that, when used together to make comparisons across nations, these national panel surveys are a currently underutilized but quite valuable analytical tool.¹

Cross-national studies, if done in a common framework with comparable measures, can be a more powerful tool for the analysis of policy impacts than studies of single countries. There are several reasons why this may be so. First, there is often considerable cross-national variation in the parameters of many key policies. The difference in the parameters may not only be "large" in some sense, but members of one country are treated with that parameter while residents of another country are not. This provides a treatment group and a comparison group, both of which may represent large populations since we are dealing with nations. It also may be much easier to sustain the argument that the comparison group is unaffected by the policy (no spillover effects) than it would be using, say, cross-state variation as is commonly done in studies in the United States. When within-country variation is informative, cross-country comparisons can add to the variability in the data, improving the precision of the estimated impact of the intervention.

A few examples of the type of analysis that is possible may help. A dramatic change in the last few decades has been a sharp decline in most Western countries in the age of retirement from the labor force. This phenomenon has raised concerns about whether national government pension programs are financially viable over the long term, especially as the large baby boom cohorts in most of these countries retire. Knowing what role changing financial incentives to retire have played is critical both in understanding why the age of retirement may have declined but also in predicting whether these trends are reversible if the underlying financial incentives are altered. In most countries, however, government-provided pensions programs are national in scope and the parameters may only change slowly over time if at all. This limited "exogenous" variation in national program parameters has made it difficult to estimate the impact of these programs on the incentive to retire. In contrast, there exists considerable across-country variation in many dimensions of these national program parameters involving the incentive to retire.²

A closely related issue concerns whether individuals will take a larger responsibility for their own retirement if current governmental programs are scaled back due

1. For a more complete exposition of the rationale, see National Research Council (2001).

2. For an excellent example of the power of this approach, see Gruber and Wise (1999).

to budgetary pressures. Will households accumulate more private wealth during their working years to finance their retirement years? Once again, a promising research strategy for answering this question is based on international comparisons as there is a great deal of variation in how countries finance the retirement of their older populations. Different countries place different weights on publicly provided pensions, private or occupation pensions, and private savings. These combinations may produce quite different rates of income replacement during retirement, and, therefore, have quite different implications for the incentives for private savings. For example, a public sector benefit that provides almost complete income replacement would reduce and possibly even eliminate incentives for private retirement savings. Two of the papers in this volume investigate this question (see the Banks, Blundell, and Smith article as well as the piece by Klevmarcken, Lupton, and Stafford, this issue).

While these examples used national pension programs to illustrate the point, many other types of programs have relatively little parameter variation within country but considerable variation across nations. The provision of health care is another case in point. Across nations, health care is provided using an amazingly diverse array of private and public support. What is the impact on health outcomes and expenditures of that diversity in the way health care is financed? Hurd and Kapteyn address this issue in this volume.

Although there are clear benefits in our view, international comparative research has its costs as well. It can demand much knowledge of national institutions and data sources, especially as the number of countries included and the complexity of the panels used increases. While understanding the institutions and programs most directly related to the target behavior are absolutely necessary, other programs and policies could have an influence as well. The difficulty is that one requires a deep understanding of all programs to decide which programs can be neglected. As an example, modeling labor supply behavior requires knowledge of tax and income transfer programs. But how to factor in the influence of Sweden's active labor market policies or centralized collective bargaining?

The institutional and contextual differences across countries also may be so large that they prevent a meaningful basis of comparison for politically feasible changes. Finally, one is much more likely to hear the argument that Italian, French, and American people cannot be compared due to "cultural" differences than one would if the comparisons involved Americans living in different states. Although we think that the evidence in this volume and elsewhere suggests the cultural differences (but not the institutional differences) argument is at a minimum overblown, not everyone shares that view.

III. International Panel Surveys

The main purpose of the conference was to advance scientific research on the value and use of panel surveys for international comparative research. While panel surveys such as the PSID and NLS have been ongoing for more than 30 years, an increasing number of important panels were fielded during the last decade in the United States, among them the HRS. Many European, Asian, and South America countries have launched panel surveys during this time, some with the

explicit intention of facilitating international comparisons. International comparative studies often relied mainly on simple aggregate data, as comparable cross-sectional and especially panel surveys did not exist. The situation now is very different from what existed ten or even five years ago and continues to be an important arena of scientific innovation.

Now 30-something, the PSID began in 1968 as an initial five-year project designed to understand transitions into and out of poverty. From this start it has evolved through the years to become one of the most widely used, ongoing panel and data archives. The PSID is generally acknowledged to be the main source of our knowledge for the entire age distribution about income and family dynamics in the United States. Two of the papers in this volume use the PSID (Banks, Blundell, and Smith; Klevmarcken, Lupton, and Stafford).

Soon after the PSID was in operation, news of and use of data from the PSID spread to several European countries and some years later generated interest in launching similar studies. The most ambitious and widely used are the German Socio-Economic Panel, the Swedish Household Panel (Huhallens ekonomiska levnadsförhållan or HUS for short), and the Dutch Socio-Economic Panel (SEP), all of which collected first-wave data in 1984. A bit of a laggard, the British Household Panel Survey (BHPS) collected its first wave in 1990. Klevmarcken, Lupton, and Stafford use the Swedish Panel while the BHPS is used both by Banks, Blundell, and Smith and by Jenkins and Schluter (who also rely on the German panel). The Dutch panel is used in the paper by Hurd and Kapteyn.

Luxembourg, the Netherlands, and the Lorraine region of France ran panels in the 1980s; quite comparable household panels in all European Community countries began in the early 1990s.³ More recently, countries embarking on such projects include Korea (1999), Japan (2000), Taiwan (1999), and Australia (2001); such a study is planned for Israel (2002).⁴

One surprising result from comparative longitudinal analyses of income and wealth data is that the United States is far from alone in its high degree of economic mobility. The Luxembourg Income Study project has documented the much higher rates of poverty prevailing in the United States than in other Western industrialized countries. But around these quite different cross-sectional poverty rates, all countries tend to exhibit considerable amounts of economic mobility.

Many of these international income panel studies were shaped with input from those with PSID experience and initially were tied quite closely to the PSID structure and modules. Subsequently, the panels in each country began to be influenced by the policy and scientific issues that emerged locally as well as by the research interests of the country's principal investigator of the survey. Moreover, the PSID itself was by no means static during these years (for example, it added periodic wealth modules and switched to a two-year periodicity in 1999) so that the international income panels, even when they shared a common parentage, began to diverge.

This divergence illustrates an inevitable but important tension in international data collection efforts: There is a desire for standardized measures so that comparisons

3. For more details on the European Community surveys, see Paracchi (2002).

4. A more detailed description of these surveys (and a route to their websites) can be obtained from the PSID website: www.isr.umich.edu/src/psid/panelstudies

across countries are meaningful; there is a competing desire to capture possibly unique and salient issues and institutions within each individual country. Imposing too strict a common template would be the death knell for innovation. Communication among the new generation of leaders of the international panel surveys had become so irregular and sparse, however, that changes were being made with little knowledge or thought about any consequences for international comparative research.⁵ This seemed to have moved too far in the other direction.

Thus, one purpose of the conference was to convene the principal investigators of the major international panel surveys. In a meeting held immediately after the scientific papers were presented, the assembled Principal Investigators were able to share with others the current state of their surveys, to tout any innovations that were successful, and to warn against those things that may not have worked as well as anticipated or did not translate so easily across international borders.

HRS, which began in the early 1990s, has several unique attributes. It attempted to monitor workforce, income, and health transitions during the pre- and post-retirement years and is representative of cohorts of those born before 1948. Two-year followups were the norm with the intention of continuing the effort until the original cohort died out. Finally, the original design teams were multidisciplinary; expertise spanned economics, demography, health, sociology, psychology, and cognition. The HRS is used in the article by Hurd and Kapteyn.

Given its success in the United States, it is not surprising that HRS also became a model for similar surveys in other countries. But by now the time lags were much shorter. The intellectual and policy impetus for change clearly stemmed from the much more acute demographic changes due to population aging in Europe alongside the reality of far more generous and thus expensive set of state provided retirement, disability, and health subsidies in most European countries.

The first to follow suit was the English Longitudinal Survey of Aging (ELSA). ELSA is actually a followup of those 50 and older who had previously participated in the Health Survey for England, where detailed physical and self-reported measures of their health were obtained. In some respects, ELSA goes far beyond the HRS by including detailed physical measurements including blood samples, blood pressure, height and weight, and waist and hip size. It does parallel HRS in attempting to measure the financial incentives to retire, the income and wealth position of households, the transitions in health and cognitive status between waves, and the systems of financial and social support on which the elderly must rely. The first round of ELSA went into the field in the spring of 2002 and is now completed. Followup rounds are planned at two-year intervals. The research team was multidisciplinary with many of the key researchers drawn from the departments of Epidemiology and Economics at University College-London.

ELSA was but a precursor to a much larger pan European effort— the Survey on Health—Aging, and Retirement in Europe (SHARE). Like HRS and ELSA, the core research questions in SHARE concentrate on the determinants of retirement, the interactions between health and economic outcomes, and the role of kin networks,

5. Several papers in this volume demonstrate that even when specific modules (such as wealth or income) were constructed by closely following the practices of a prior U.S. survey, great care has been exercised to obtain comparable measures across countries.

intergeneration exchanges, and living arrangements. SHARE also shares the main features of those surveys—a deliberately multidisciplinary team across health, economics, sociology, and demography, a longitudinal survey design starting with participants in their early 50s, and a commitment to open data availability to all researchers (an increasing problem in European data collection). Participating countries in the first stage of SHARE include Denmark, France, Germany, Greece, Italy, the Netherlands, Spain, and Sweden. Support has been obtained from the European Union and the National Institute on Aging to develop the questionnaire and to conduct pilot studies in the participant countries. These pilot studies are scheduled to take place in 2003.⁶

The eventual impact of these HRS-like surveys throughout Europe is of course not now known. While there are many pitfalls that remain in the execution, the potential is enormous. Within the next five years, the international research community may have access to a relatively common set of surveys that have attempted to measure across ten or more countries the economic and social determinants of retirement, health transitions, intergenerational exchanges, and cognitive ability. That will be a very different research world indeed.

At the other end of the age spectrum, the most respected panel surveys of American youth are the National Longitudinal Survey of Youth 1979 (NLSY79) and the National Longitudinal Survey of Youth 1997 (NLSY97). The NLSY79 is a random sample of young men and women who were 14–22 years old in 1979 while the NLSY97 is a sample of those who were 12–16 years old in December 1996. Both of these major longitudinal surveys address key questions about the economic, social, and academic experiences of young adolescents and young adult respondents, with a special emphasis on their initial entry into the work force and subsequent transitions in and out of the labor force.

To our knowledge no panel surveys in other countries were directly modeled after the NLS. However, a number of cohort and youth surveys in the United Kingdom adopted several key components of the NLS with the aim of facilitating such international comparisons. For example, working closely with NLS research staff, recent waves of the British National Child Development Study (NCDS) (used by Aughinbaugh and Gittleman in this volume) included assessment tests quite similar to those used in the 1979 NLSY. Similarly, a new Longitudinal Survey of Young People (LSYP) in Britain will follow a group of 13–15 year-olds and has used the NLSY97 as one of its models.

Why are there so few NLS look-alikes in other countries? That's clearly a good question, especially when the substantive need seems so high. The issues of youth demographic and workforce transitions are arguably more salient throughout Europe than they are in the United States. Perhaps the problem of attracting long-term funders when the substantive questions fall between the bureaucratic cracks (too far into the education process, not yet into the labor force) is one cause. Whatever the reason, an international collaborative effort by scholars to create additional comparable panels that focus on youth transitions would have a high payoff.

The growth in high-quality international panel surveys has not been limited to the

6. One can learn more about SHARE and how matters are progressing from their website: www.share-project.org

developed world. Research in developing countries faces several stumbling blocks: It is difficult to find funders willing to make long-term commitments to panel surveys in developing countries. Moreover, there is a misguided belief that it is too difficult to obtain high quality results in developing countries or a misplaced willingness to settle for a lower-quality effort. While by no means the only example, the Indonesian Family Life Survey (IFLS) used in one of the papers in this volume (Frankenberg, Smith, and Thomas) serves as a good counter-argument. Both retrospectively and in panel, IFLS provides detailed information on a broad array of topics including education; migration; assets and wealth; use of health care and health status; marriage; fertility and contraception. On all the standard dimensions of survey quality (baseline response rates, attrition, data quality), the IFLS matches up well with the best of the panel surveys in developed countries.⁷

The IFLS is an ongoing panel survey of individuals, families, and communities in Indonesia that began in 1993, with its first followup taking place in late 1997, with another in 2000. In addition, an originally unplanned re-interview took place in 25 percent of the enumeration areas in late 1998 in order to measure the impact of the severe ongoing financial crisis taking place at that time. This ability to field the “unplanned” round in 1998 illustrates an important benefit of having an ongoing national panel that covers a broad array of behavioral outcomes. One can never fully anticipate when financial or other types of crises may emerge; nor can one predict the precise outcomes that one may want to monitor. Ongoing panels provide the essential “before and after” and often the “control and treatment” groups that are necessary to test the impact of unanticipated shocks.

IV. The Papers

This volume includes revised versions of seven papers that were presented at the Ann Arbor conference. Five of the papers explicitly or implicitly examine international differences in savings behavior and wealth accumulation. In their paper, James Banks, Richard Blundell, and James P. Smith document the size of and investigate the determinants of some vast differences in financial wealth holdings in Great Britain and the United States. These differences are indeed large, especially in the upper fifth of the financial wealth distribution. For example in the mid-1990s among those aged 60 or older, at the 90th percentile U.S. households had about \$561,000 in financial wealth compared with only \$128,000 among British households.

The authors identified some often-mentioned factors that they do not believe played much of a role in producing these differences. The rejected factors include differential financial inheritances, different rates of return to corporate equity or differential employment or income risk. One factor that does appear to be important are differences in initial conditions—in reaction to quite similar stock market booms during the 1980s and 1990s, household wealth expanded far more in the United States than in the United Kingdom because U.S. households were initially far more invested in the equity market. Two other factors may have played a significant role.

7. See Thomas, Frankenberg, and Smith (2001) for details.

Housing price risk is far greater in the United Kingdom, making investments in risky financial assets such as the stock market a questionable strategy. Finally, due to forced and voluntary annuitization of their retirement incomes, British households face much less longevity risk and thus have less need for precautionary wealth holdings in older ages.

It is sometimes thought that the use of financial assets to “smooth out” consumption when economic shocks hit is primarily an issue for developed countries only. In their paper, Elizabeth Frankenberg, James P. Smith, and Duncan Thomas show that this is far from the case. In 1998, Indonesia, like many other South Asian countries, suffered a severe and unanticipated financial crisis that in that year alone reduced real wages in the formal sector by 40 percent while its GNP fell by 15 percent. How can households adjust to an economic shock of that magnitude to mitigate its consequences? Using a unique panel survey in Indonesia, these authors show that households used several strategies. First, households became larger as dependents moved into households living in lower-cost locations and working-age people combined into fewer households. In many households, labor supply also increased, often by entering the self-employed sector. Spending on consumption of semidurables was deferred, lessening the impact on the more temporarily critical food consumption.

Household wealth holdings also played a smoothing role but in a manner quite context specific. One consequence of the crisis was that the real value of many assets fell substantially, lessening their value for consumption smoothing. The exception was gold, a widely held asset in Indonesia whose real price rose significantly. There appears to be clear evidence that rural households in particular sold significant amounts of their gold to mitigate the necessity of reducing their consumption expenditures.

Among Western countries, Sweden and the United States are often thought to lie at opposite ends of the spectrum in their tolerance for dispersion, differentiation, individualism, and the role of the state. In their paper, Anders Klevmar, Joseph Lupton, and Frank Stafford demonstrate that this characterization also is reflected in their respective household wealth distributions with much higher dispersion and positive skewness in the U.S. distribution. The authors also highlight other key differences including a relatively large fraction of American households who are mostly “out of the asset game” and a greater growth over time in wealth inequality in the United States compared with Sweden.

These authors show that both the higher level and secular growth in wealth inequality in America is not due to the African-American households who have very low levels of household wealth. Their principal interest lies however in wealth mobility—the relative tendency to remain in the same decile of the wealth distribution over time. In this evaluation, the metric really matters. A larger fraction of Swedish households move across wealth deciles over time, but these deciles are much closer in Sweden compared with those in the United States. To deal with this problem, the authors develop a matching algorithm to standardize the two wealth distributions for any initial differences that may exist as well as for some of the impacts of the well-documented measurement error in household wealth. After making these adjustments, they conclude that decile mobility is roughly the same in both countries but that dollar wealth mobility is much higher in the United States.

Income, savings, wealth, and consumption are linked by identities, but no data set anywhere claims to have good measures of all concepts. Of all these concepts, consumption comes closest to matching our ultimate variable of interest, but it is widely felt that consumption is too time consuming and difficult to include in multi-purpose household surveys. Erich Battistin, Raffaele Miniaci, and Guglielmo Weber use two Italian data sets—one based on a lengthy time-consuming diary and the other based on much shorter recall of consumption aggregates—to investigate how comparable alternative measures of consumption are. They develop and empirically implement a statistical methodology that shows that after correcting for the rounding and heaping of answers, measures of food expenditures appear to be of comparable quality in the two quite different types of surveys. In contrast, for durable consumption expenditures large differences remain between the two surveys. Based on these findings, they argue that nondurable consumption can best be estimated by using reverse Engel curves including food consumption.

In addition to their methodological contribution, the authors point out a major puzzle that emerges from the Italian data. In contrast to the implications of almost any economic model, savings rates appear to increase with age well into retirement. It remains an open question whether this is an artifact of measurement issues in the data sets or requires a major rethinking of the role of life-cycle savings, at least in the Italian context.

One issue on which international comparisons have been a central part of the recent debate has involved the SES-health gradient—the relationship between several measures of socioeconomic status (SES) and health outcomes. In his provocative book, Wilkinson (1996) cited cross-country differences in levels of income equality and mortality as among the most compelling evidence that unequal societies have negative impacts on individual health outcomes. This question is taken up by Michael Hurd and Arie Kapteyn, who examine the nature of the SES-health gradient in two countries—the Netherlands and the United States. One advantage of selecting these two countries is that the Netherlands has set up institutional arrangements whose goal is to isolate individuals from the economic consequences of poor health in terms of medical expenditures, healthcare access, and earnings interruptions.

Hurd and Kapteyn report results generally consistent with these different institutional arrangements in the two countries. For example, the more generous income maintenance system in the Netherlands appears to mitigate the effect that adverse health changes have on income there compared with the United States. In addition, the smaller impact of wealth and income on subsequent health change in the Netherlands is seen by the authors as a reflection of the more equal access to health care (by income position) there compared with the United States.

The final two papers use international comparisons to assess the status of young children. By contrasting the United States and Great Britain, Alison Aughinbaugh and Maury Gittleman ask whether family income has any differential impact on child development in terms of cognitive and noncognitive test scores. Their expectations were that such differences should emerge in part due to the more extensive welfare system in Britain compared with the United States. In spite of these expectations, the authors find that the effects of income on child development are quite similar in both countries and are also small in absolute value. Their paper suggests that

across these dimensions Great Britain and the United States may actually be too similar and that tests involving other European countries with more extensive child welfare systems may show more systematic differences.

In their paper, Stephen Jenkins and Christian Schluter return to the original theme that begat the PSID in the first place—the dynamics of transitions into and out of poverty—but their comparisons involve Germany and the United Kingdom. During the mid 1990s, child poverty rates were about two and one half times larger in the United Kingdom than in Germany. They conclude that the source of these cross-sectional differences can be best addressed by raising related questions only answerable with panel data: Compared with Germany, why are the transitions into (out of) poverty so much larger (smaller) in the United Kingdom? They argue that the probability of making a labor force or family transition are similar in the two countries and therefore cannot account for the disparities in poverty rates. Instead, conditional on making a labor force or family transition, German households are much less likely to fall into poverty, a reflection of the more generous welfare state in Germany compared with the United Kingdom.

The three editors shared the normal editorial duties with James Walker taking responsibility for papers on which the other two editors were coauthors. Jan Levine Thal did her normal superb job as managing editor for all three of us, a compliment to which a long line of editors of the JHR can attest. We are in her debt.

References

- Gruber, John, and David Wise. 1999. *Social Security and Retirement Around the World*. Chicago: University of Chicago Press.
- National Research Council. 2001. *Preparing for an Aging World: The Case for Cross-National Research*. National Academy of Sciences: National Academy Press
- Paracchi, Franco. 2002. "The European Community Household Panel: A Review." *Empirical Economics* 1:63–90.
- Thomas, Duncan, Elizabeth Frankenberg, and James P. Smith. 2001. "Lost and Forgotten: Attrition and Recall Error in Indonesia." *Journal of Human Resources* 36(3):556–92.
- Wilkinson, Richard. 1996. *Unhealthy Societies. The Afflictions of Inequality*. London: Routledge.