The Population Research Center

NORC & The University of Chicago

PRC Newsletter | Volume 4, Issue 2 | Winter Quarter 2014

HIGHLIGHTS IN THIS ISSUE

Data Harmonization *by Louise Hawkley* Page 2
Data Science for Social Good Summer 2013 Fellowship Page 3
Research with a Global Focus Page 5

PRC Associate Lars Peter Hansen Wins Nobel Prize

In October 2013 the Royal Swedish Academy of Sciences awarded Population Research Center associate Lars Peter Hansen the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. This honor is shared with Eugene F. Fama, the Robert R. McCormick Distinguished Service Professor of Finance at the University of Chicago Booth School of Business, and Robert J. Shiller of Yale University.

Hansen's research looks at ways to bridge the gap between economic models and economic and financial data. A major obstacle in building economic models is that imperfect data can lead to a number of imprecise equations. Hansen showed that researchers can multiply these equations by a set of weights - putting more weight on equations that they have better information about - and sum them in order to derive better conclusions. His 1982 *Econometrica* paper, "Large Sample Properties of Generalized-Methods of Moments Estimators," fundamentally altered the way that empirical research is done in finance and macroeconomics.

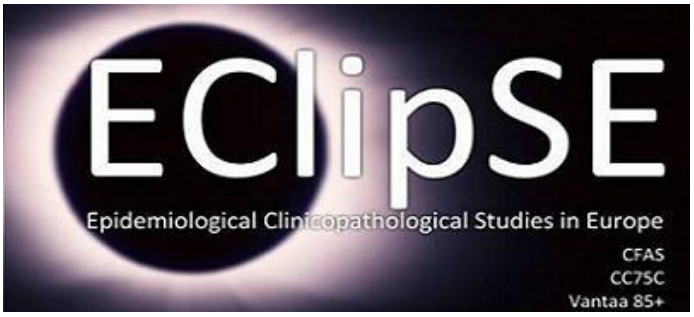
He has applied these methods to study the determinants of consumption, savings, and security market prices. His recent work focuses on models that incorporate ambiguities, beliefs, and skepticism of consumers and investors; specifically, he is exploring how these models can explain economic and financial data to understand the consequences of policy options. He is also principal investigator on a research project that has assembled a group of elite economists to develop macroeconomic models with enhanced linkages to the financial sector. These models will provide more powerful policy tools for measuring and monitoring systemic risks to the economy arising from financial markets.

Nobel Prize recipients Lars Hansen (center) and Eugene Fama (left) stand with Provost Thomas Rosenbaum (right). The prize is shared with Robert Schiller (Yale, not pictured)

The PRC congratulates Hansen for his accomplishments and is proud to support his work.

For more information about Hansen see http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/press.html and http://www.larspeterhansen.org/

Data Harmonization



By Louise Hawkley

What is data harmonization?

Taking a cue from the musical allusion, data harmonization efforts seek to find a common key in which instruments (variables) from various data sources can sing in harmony. Empirically, the goal of data harmonization is to "create data that measure the same conceptual variable and that are measured in the same units" (Birkhauser & Lillard, 2005, p. 317). When the survey process is centralized across different surveys, "input" harmonization (i.e., prior to data collection) can be implemented by ensuring that each survey team applies the same survey procedures and a common questionnaire. When the survey process is determined by individual survey teams, "output" harmonization (i.e., after data collection) can be implemented (a) ex ante to the extent that the concepts of interest were taken into account in survey planning, and/or (b) ex post by a conversion process that conforms individual survey data to the common concept.

Why is data harmonization conducted?

Efforts to harmonize data initially arose out of an interest in cross-national and cross-cultural comparative research. This remains an important motivation for combining data from different countries or different cultures within a country.

Data harmonization is also valuable for studies that require larger sample sizes than any one survey provides, whether these surveys come from within a single country or from different countries or cultures.

Data harmonization is becoming increasingly important to federal funding agencies. For instance, the National Institute on Drug Abuse (NIDA) recently solicited proposals to "fund a single interdisciplinary Coordinating Center to formalize and centralize support of the 'Seek, Test, Treat and Retain' (STTR) data harmonization initiative." NIDA anticipates that the integrated dataset of harmonized measures will include data for approximately 70,000 individuals, thus permitting researchers to go beyond the capacity of any single HIV intervention study to address population-level research questions regarding HIV and drug abuse, including the efficacy of the STTR model in reducing HIV incidence, infectivity, morbidity, and mortality at the systems, community, and population levels.

Similarly, the growing availability of DNA collected in diverse population-based studies has prompted federal support for data harmonization efforts to identify, develop, and implement comparable measures of phenotypic information across studies that can be combined with genetic information and generate the very large sample sizes called for in genome-wide association studies that seek to detect, for example, SNP-disease associations.

How are harmonized data used?

- The Epidemiological Clinicopathological Studies in Europe (EClipSE) collaboration harmonized data from three longitudinal studies of aging and dementia across Europe with brain donation to compare the relationship between cognitive function and neuropathology in countries with different education systems to better understand the effects of education in childhood, and other factors, on dementia in old age (Brayne et al., Brain, 2010).
- The Gallup Organization conducted a World Poll that asked individuals to evaluate their national satisfaction (satisfaction with one's country) and individual life satisfaction. The resulting sample consisted of 132,516 individuals from 128 countries in Western and non-Western regions of the world. Researchers theorized that people would rate their life satisfaction higher when they perceived that their country was doing well (i.e., higher national satisfaction), and that national and life satisfaction

would be more highly related in collectivist than individualist cultures (Morrison, Tay, & Diener, Psychological Science, 2011).

- A research team combined data from 24 studies of the Breast Cancer Association Consortium, harmonized environmental risk factor variables across studies, and tested whether breast cancer risk associated with genetic variation in 23 single nucleotide polymorphisms was modified by environmental risk. The final sample consisted of 34,793 cases of invasive breast cancer and 41,099 controls, and the data provided strong evidence that breast cancer risk is susceptible to gene-environment interactions (e.g., parity modified the association of the LSP1-rs3817198 SNP with cancer risk) (Nickels et al., PLoS Genetics, 2013).

Who can harmonize data?

YOU can! Individual researchers or research teams testing theories are in the best position to conduct output data harmonization across diverse surveys because theoretically motivated substantive research questions guide how data are harmonized. The Population Research Center (PRC) website provides links to publicly available data from large national and international surveys (http://popcenter.uchicago.edu/data/). In addition, a link is provided to the website for the "Cross-National Equivalent File," a resource that provides a set of constructed variables (e.g., income) that are formulated to be equivalent across panel surveys obtained in 8 countries as of this writing (Australia, USA, Britain, Korea, Russia, Switzerland, Canada, and Germany). RAND has created and made available harmonized datasets for the "HRS family of surveys," and has a good search tool for other studies in their repository (https://mmicdata.rand.org/meta/?section=search&modify=y&keywords=loneliness&domains=&subdomains=&topics=&limitSurveys=1&showRestricted=0&studies=&surveys=&modules=1077).

Resources for those seeking to conduct data harmonization can be found in published papers in this growing area, and online (e.g., http://vimeo.com/21036173). NORC at the University of Chicago contributes local expertise in the area of data harmonization. For instance, Felicia LeClere (http://www.norc.org/Experts/Pages/felicia-leclere.aspx) and Michael Davern (http://www.norc.org/Experts/Pages/michael-davern.aspx) have diverse experience with data harmonization that includes projects in the fields of health (e.g., HIV Open Data Project) and education (e.g., National Children's Study). Data harmonization projects typically engage interdisciplinary research teams, and they are an excellent opportunity to expand and deepen a field of research theoretically and empirically.

Data Science for Social Good Summer 2013 Fellowship

Last summer the University of Chicago hosted the Eric and Wendy Schmidt Data Science for Social Good (DSSG) summer fellowship. The fellowship brought together 36 graduate and undergraduate students to work with government and nonprofit agencies to work on data science projects with real world social impact. Several projects are described below:

Predictive Analytics for Smarter City Services

Working with Chapin Hall at the University of Chicago, DSSG fellows analyzed data from the City's 311 System to build predictive models for when and where graffiti, potholes, and other problems are likely to occur. Such models would help the City provide faster service to its citizens. Exploratory analyses uncovered correlations between service requests and temporal and sociodemographic factors. For example, pothole requests are highest in the spring, and graffiti removal is most commonly reported in predominantly Hispanic neighborhoods while weed removal is most commonly reported in Black neighborhoods. Based on these patterns, the team built a Poisson Generalized Linear Model that best explains the relationship between demographic indicators and volume of 311 requests on a monthly basis. Future directions include using more data, such as data from the

Chicago Data Portal, to fit the models, and creating alternative statistical models.

Predicting When Divvy Bike Share Stations Will Be Empty or Full

The City's new bike share system Divvy faces a fundamental problem: because of commuting patterns, bikes tend to pile up downtown in morning and in the outlying neighborhoods in the afternoon. This imbalance makes using the bike share system difficult because people cannot take bikes out from empty stations, or park their bikes at full stations. The City of Chicago's Department of Transportation teamed up with DSSG fellows to predict how many bikes would likely be at each Divvy bike share station at specific times of the day. The team used a Poisson model to simulate the number of bikes at a station in a 2-hour window accounting for factors such as day of the week, weather conditions, and whether special events are happening nearby a certain station. Results from this project helped to create this live web app of Washington, D.C.'s bikeshare system: bikeshare.dssg.io.

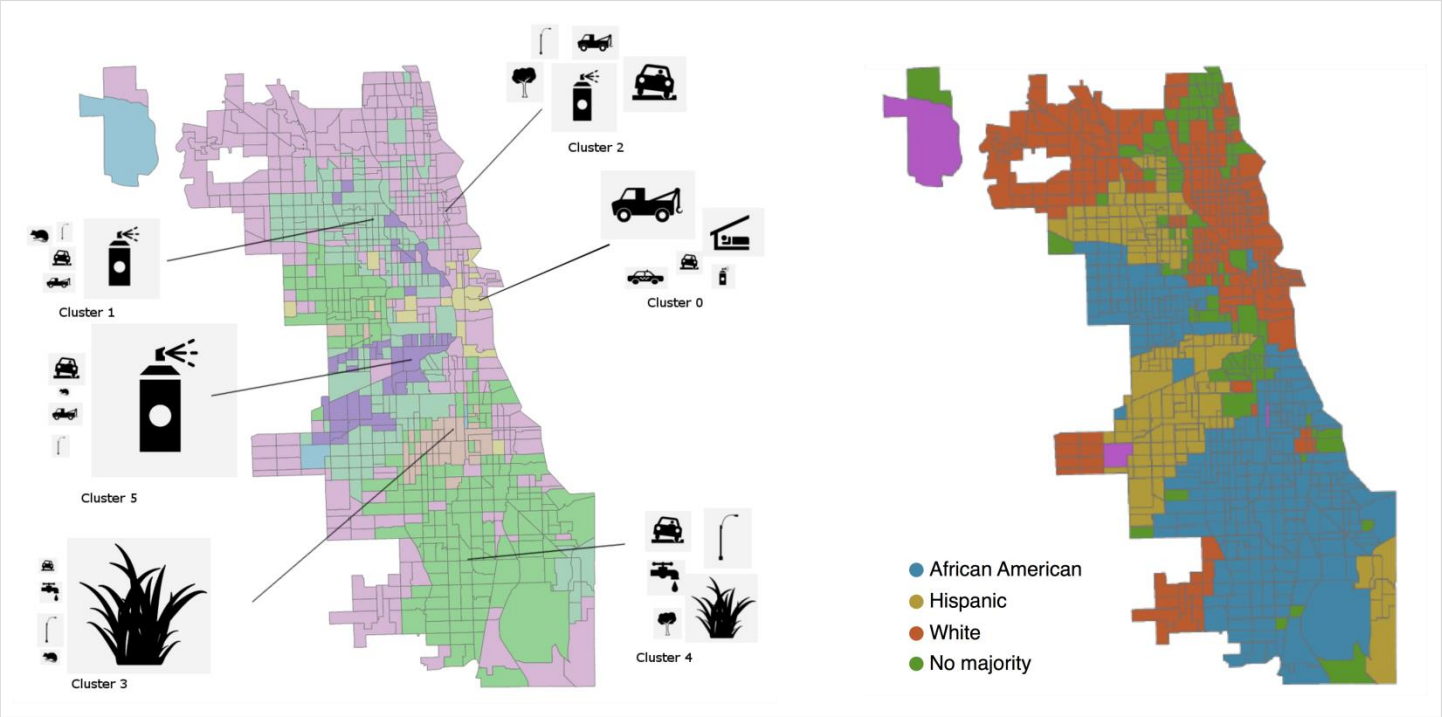
Using Electronic Medical Record Data to Predict Health

A team of DSSG fellows created an analytics tool for pediatricians to better detect kids who are likely to become obese as they age. They tested whether adiposity rebound, a phenomenon in which a child's BMI dips and then rebounds around age 5 to 6, was associated with adult obesity. Using electronic medical record (EMR) data from NorthShore University Hospital, the team confirmed that adiposity rebound could be used as a predictor of childhood obesity. This tool could potentially be built into future EMR systems.

The Data Science for Social Good summer program was a huge success. Not only did it provide hands-on training to the group of data scientists, it also increased their interest in working on social problems. The participating non-profits and government organizations became better-informed about working on data and analytics-related problems. Most notably, the fellowship helped create software, analytical tools, and prototypes useful to the partner organizations as well as to other similar organizations around the world.

To learn more about the projects from the 2013 DSSG Summer Fellowship, visit http://dssg.io/projects/.

Prospective fellows, mentors, partners, and funders can learn about the 2014 DSSG Summer Fellowship here: http://dssg.io/contact/.



This map shows that different service requests tend to cluster according to the racial composition of Chicago's neighborhoods. Models that predict service requests can help cities provide faster service to its citizens.

Recent Peer-Reviewed Publications by PRC Associates

Dai, X. and **Heckman, J. J.** (2013). Older siblings' contributions to young child's cognitive skills. *Economic Modelling*, 35, 235-248.

Guryan, J. and **Charles, K. K.** (Forthcoming). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal.*

Abramsohn, E. M…**Lindau, S. T.** (2013). "I'm not just a heart, I'm a whole person here": A qualitative study to improve sexual outcomes in women with myocardial infarction. *Journal of the American Heart Association*, 2013(2)), 1-11. PMID: 23885024.

Sciandra, M…**Ludwig, J.** (2013). Long-term effect of the Moving to Opportunity residential mobility experiment on crime and delinquency. *Journal of Experimental Criminology*, 9(4), 451-489.

Bilger, M. and **Manning, W. G.** (2013). Measuring overfitting in nonlinear models: a new method and an application to health expenditures. *Health Economics*, published online October 9, 2013.

Kaburu, C. W., **Mojola, S. A.**, et al. (2013). Growing up at the "margins": Concerns, aspirations, and expectations of young people living in Nairobi's slums. *Journal of Research on Adolescence*, 23(1), 81-94.

Neal, D. (2013). The consequences of using one assessment system to pursue two objectives. *The Journal of Economic Education*, 44:4, 339-352.

O'Muircheartaigh,C. and **Hedges, L. V**. (2013). Generalizing from unrepresentative experiments: a stratified propensity score approach. *Journal of the Royal Statistical Society: Series C*, published online November 12, 2013.

Oster, E., Shoulson, I., and Dorsey, E. R. (2013). Limited life expectancy, human capital and health investments. *American Economic Review*, 103(5), 1977-2002.

Phillips, D. J., Turco, C. J., and Zuckerman, E. W. (2013). Betrayal as market barrier: Identity-based limits to diversification among high-status corporate law firms. *American Journal of Sociology*, 118(4), 1023-1054.

Basu, A…**Philipson, T. J.**, et al. (2013). Heterogeneity in action: The role of passive personalization in comparative effectiveness research. *Health Economics*, published online October 9, 2013.

LeBarton, E. S., Goldin-Meadow, S. and **Raudenbush, S.** (2013). Experimentally-induced increases in early gesture lead to increases in spoken vocabulary. *Journal of Cognition and Development*, published online November 12, 2013.

Small, M. L. Weak ties and the core discussion network: Why people discuss important matters with unimportant alters." *Social Networks*, 35, 470-83

Yamaguchi, K. (2013). New regression models with egocentric social network data: An analysis of political party preference in Japan. *Sociological Methodology*, 43, 312-345.

Research with a Global Focus

Research at the Pop Center spans many topics such as education, the family, and health and aging. While many of our associates and affiliates conduct research on a regional and national level, several PRC members have directed their research questions to contexts outside of the United States. A few examples of this global research appear below:

Ofer Malamud – Self-Selection and International Migration: New Evidence from Mexico

Working with the Mexican Family Life Survey (MxFLS), PRC associate Ofer Malamud and colleague Robert Kastner (UIC and NBER) compare characteristics of Mexicans who migrate to the United States with non-migrants who remain in Mexico. Their results show that compared to non-migrants, both male and female Mexican migrants were more likely to be young, unmarried, to come from the middle of the education distribution (i.e., 4 to 9 years of schooling), and to come from rural areas. However, they were similar to non-migrants in cognitive ability and health. Male Mexican migrants are negatively selected on earnings, and this result is largely explained by differential returns to labor market skill between the U.S. and Mexico rather than differential costs of migration. These findings have social and economic implications for Mexico and has implications for U.S. labor market and immigration policy. The complete article is forthcoming in the *Review of Economics and Statistics*.

Alicia Menendez - Fertility Timing and Women's Economic Outcomes in South Africa

This ongoing project led by Murray Leibbrandt (University of Cape Town) and David Lam, (University of Michigan) uses rich longitudinal data from the Cape Area Panel Study (CAPS), the Africa Centre Demographic Information System (ACDIS), and the National Income Dynamics Study (NIDS) to analyze the relationship between fertility and women's economic outcomes in South Africa. A major focus is integrating data on the timing and placement of family planning clinics, particularly the rollout of the National Adolescent-Friendly Clinic Initiative (NAFCI), with the nationally representative data in NIDS. The temporal and spatial variation in access to these clinics is associated with the timing of first births, and provide an exogenous source of variation in fertility timing that can be linked to details on later economic outcomes. This work is supported by a Population, Reproductive Health, and Economic Development grant.

Emily Oster - Do IT Service Centers Promote School Enrollment? Evidence from India

PRC Associate Emily Oster and Bryce Steinberg (Harvard) investigate how globalization has changed job opportunities in India. Combining data from a large administrative dataset on primary school enrollment in India called the District Information System for Education

(DISE) with data on Information Technology Enabled Services (ITES) centers, the researchers find that introducing a new ITES center causes a 5% increase in the number of children enrolled in primary school. They show that changes in returns to schooling, rather than changes in population or income resulting from the ITES center, is the driving mechanism of this increased school enrollment. However, they find that the effect is localized to within several kilometers because of limited information diffusion. The full article has been released in the *Journal of Development Economics*.

Upcoming Events at the Demography Workshop

Co-sponsored by the Population Research Center and the Center on Aging, the Demography Workshop is held on Thursdays from noon to 1:20 p.m. at 1155 East 60th Street, Seminar Room 232/233 during the fall, winter, and spring quarters. The workshop organizer for Fall Quarter was Dan Bennett. Presentation topics included White residential segregation in the U.S. (John Iceland, The Pennsylvania State University), and religion, economic growth, and happiness (David Yanagizawa-Drott, Harvard Kennedy School). Papers and abstracts, if available, are archived at: http://popcenter.uchicago.edu/ events/papers.shtml.

Amy Claessens is the workshop organizer for the Winter Quarter. The schedule appears on the right. For information about the speakers or their talks, please visit http://popcenter.uchicago.edu/events/demography_workshop.shtml or email Program Coordinator Jaclyn Wong at wong-jaclyn@norc.org.

The Spring Quarter workshop will be organized by Heather Hill. Speakers invited to present at the Spring 2014 Demography Workshop include Professor Narayan Sastry and Alex Murphy, Ph.D., from the University of Michigan, Professor Rachel Dunifon from Cornell University, and Professor Rachel Gordon from the University of Illinois at Chicago. In addition to hosting these speakers, the PRC will co-host the annual seminar on the Protection of Human Subjects in Social Science Research (PHSR) with the Center on the Demography and Economics of Aging (CoA) and the NORC Institutional Review Board (IRB) during the spring quarter. The goal of the seminar is to familiarize students and scholars with responsible research practices in the social sciences. Last year's PHSR seminar was led by Professor Laura Stark (Vanderbilt University) and focused on federally mandated Institutional Review Boards and the debate that ensued over writing the rules for the treatment of human subjects. More details about the various historical contingencies that had a lasting impact in determining the treatment of human subjects both after the Second World War and in the modern era can be found in her book *Behind Closed Doors: IRBs and the Making of Ethical Research*. This year's PHSR seminar will be held on April 3, 2014. More information about the upcoming PHSR seminar speaker and topic will be posted to the PRC website when it becomes available.

WINTER QUARTER 2014 DEMOGRAPHY WORKSHOP SCHEDULE

January 9	CANCELLED	
January 16	John Casterline, Professor of Sociology, Ohio State University	The Contours of Contemporary Fertility Declines: A Fresh Assessment
January 23	Asim Khawaja, Professor of Public Policy, Harvard Kennedy School	Tax Farming Redux: Experimental Evidence on Incentive Pay for Tax Collectors
January 30	Christine Percheski, Assistant Professor of Sociology, Northwestern University	Economic Conditions and Pregnancy Rates in the United States During the Great Recession
February 6	Diane Schanzenbach, Associate Professor of Human Development and Social Policy, Northwestern University	Long-Run Impacts of Childhood Access to the Safety Net
February 13	Sarah Mustillo, Associate Professor of Sociology, Purdue University	Women's Health and Poverty over the Life Course: The Ghost of Generations Past?
February 20	Rupa Datta, Vice President and Senior Fellow, Center for Excellence in Survey Research at NORC	Initial Findings from the National Survey of Early Care and Education: Providers of Care to Children Age 5 and Under
February 27	Jason Grissom, Assistant Professor of Public Policy and Education, Vanderbilt Peabody College	Strategic Staffing: How Accountability Pressures Affect the Distribution of Teachers within Schools and Resulting Student Achievement
March 6	TBA	TBA
March 13	Quincy Stewart, Associate Professor of Sociology, Northwestern University	TBA

at the UNIVERSITY *of* CHICAGO

1155 East 60th Street
Chicago, Illinois 60637
Telephone: (773) 256-6315
Fax: (773) 256-6313