Income inequality and mortality in US counties, 1990-2010: A dynamic spatial panel analysis

Tse-Chuan Yang, Ph.D.[†] Stephen Matthews, Ph.D.[‡] Kiwoong Park, Doctoral Student[†]
[†]University at Albany, State University of New York
[‡]Pennsylvania State University

Introduction

One unsettled debate in health research is whether income inequality (hereafter, inequality) positively affects mortality (Deaton 2003; Kawachi and Blakely 2001; Lynch et al. 2004a; Lynch et al. 2004b; Mellor and Milyo 2001). While several recent county-level studies have provided evidence to support the positive inequality-mortality relationship (Cossman et al. 2008; McLaughlin, Stokes and Nonoyama 2001; McLaughlin et al. 2007; Yang et al. 2012), none of these studies adopted a longitudinal perspective. This methodological shortcoming is also shared by studies that suggest that the association between inequality and mortality is spurious (Deaton and Lubotsky 2003; Laporte 2002; Muller 2002). In this paper, we adopt an explicitly longitudinal perspective and analyze county-level panel data (1990-2010) in the United States (U.S.) and our findings will be relevant to ongoing debates regarding inequality and mortality.

Why is a longitudinal perspective important? Briefly, if inequality positively affects mortality, one would expect that the fluctuations in mortality will be reflected in the fluctuations in inequality. Specifically, if inequality increases, mortality should also increase. In the past few decades, mortality in the U.S. has declined from roughly 11 deaths per 1,000 population in 1980 to 7.5 deaths per 1,000 population in 2010 (Hoyert 2012). The downward trend holds for both crude death rate and age-adjusted rate; the latter shows a more conspicuous pattern. However, over the same time period, income inequality has increased remarkably (Piketty and Saez 2004). Using a range of adjusted Gini indices, one recent study (Burkhauser et al. 2011) reported that income inequality has been trending upward since the 1970s, with a significant increase observed around 1990. Though the inequality trend seems to have slowed down in the past decade, the overall upward trend is evident. From a longitudinal perspective the two opposite trends would seem to suggest that income inequality and mortality are either unrelated or negatively related, which challenges the common belief that inequality is bad for mortality.

Also, a longitudinal perspective on inequality and mortality is necessary because of potential lag effects; that is, inequality may not have an immediate impact on mortality. Most, if not all, of the previous research used cross-sectional design and as such causality between inequality and mortality cannot be identified. We argue that, if inequality is a determinant of mortality, then the impact on mortality will need time to unfold. The impact of inequality on mortality is analogous to that of diseases that have long latency periods. Indeed, the theoretical pathways from inequality to mortality imply the potential for latencies. Specifically, on the one hand, a psychosocial pathway refers to the sense of relative deprivation caused by unequal distribution of wealth within a society (Marmot 2004; Wilkinson and Pickett 2006) where individuals who feel deprived are more likely to adopt unhealthy behaviors (e.g., smoking, alcohol intake) and/or suffer from mental health-related issues (e.g., stress and depression). These factors typically expose individuals to higher risks of death. On the other hand, an underinvestment pathway implies that high inequality results in limited public services, poor infrastructure, and undesirable social conditions (e.g., poverty and unemployment). Individuals living in such a society are at high risk of poor health as these social conditions and infrastructure may reflect fundamental causes of disease and death (Link and Phelan 1995; Phelan, Link and Tehranifar 2010). While mental illness may be an immediate threat to life (e.g., suicide), most unhealthy behaviors need time to develop into chronic health conditions that lead to premature death. Similarly, poor social conditions and/or infrastructure are more likely to be the consequences of persistent underinvestment rather than directly associated with short-term change in the level of underinvestment. Thus a longitudinal perspective should incorporate lags or latency and in doing so we can advance understanding of the inequality-mortality relationship.

This study uses 1990 and 2000 Census data and the American Community Surveys (ACS) 2006-2010 five-year estimates to construct a three-wave panel dataset for all counties in the contiguous U.S. We then combine the panel dataset with restricted mortality data maintained by the National Centers for Health Statistics (NCHS) and can then answer the following related questions: (1) In separate cross-sectional analysis does the relationship between inequality and morality remain consistent over time (e.g., in 1990 and 2010, respectively)? (2) What does a longitudinal analysis yield? (3) After accounting for other variables, does a longitudinal perspective find evidence of lag effects of inequality on mortality? These research questions and use of spatial panel data and methods will improve our understanding of the inequality-mortality relationship in the U.S.

Measures

Mortality: Our dependent variable, age-sex standardized mortality rate, was derived from three restricted Compressed Mortality Files (CMF) (NCHS 2000, 2003, 2012). To account for the annual fluctuations, we used five-year average mortality rate in the analysis. For example, we averaged 1988-1992 mortality rates in the 1990 wave. Note that the latest available mortality data from NCHS is for the year 2010 and so we use 2006-2010 average mortality for our third wave; a time period corresponding with the ACS 2006-2010 five-year estimates. Also, the reason why we did not standardize mortality by race/ethnicity is that the CMF did not include this until 1999. To address this issue, we include racial/ethnic composition of a county in the analysis.

Income inequality: The Gini index was used to capture income inequality. This measure gauges the level to which the actual income distribution of a population deviates from the complete equal income distribution. The Gini index ranges from 0 (no income inequality) to 1 (completely unequal income distribution). Individual income data are required to compute the most accurate Gini index; however, publicly available census (and ACS) data only provide income data in multiple ordinal categories so the Gini index used in this study may slightly underestimate income inequality due to the open top coding issue.

Social conditions: Following the fundamental-cause argument (Link and Phelan 1995), ten social condition variables were derived from Census and ACS data which can be further classified into four groups, namely *racial/ethnic composition*, *socioeconomic status*, *disadvantaged groups*, and *residential deprivation*. Percentages of non-Hispanic Black and Hispanics are included in racial/ethnic composition; socioeconomic status comprises percentage of population aged 25 and above who have at least a bachelor degree, the unemployment rate, and poverty rate. Percentages of female-headed families and households receiving public assistance are two disadvantaged groups. Residential deprivation encompasses percentage of household without a car, the housing vacancy rate, and the percentage of renter occupied housing units.

Method

To answer the research questions above, we will first apply cross-sectional spatial econometrics models (Anselin 1988), such as spatial lag and spatial error, to each wave of data, respectively. This will enable us to identify whether cross-sectional analyses generate consistent results. Due to the space constraint and the popularity of these models, we opt not to describe the details of these models. The second stage of analysis will use a dynamic spatial panel modeling approach and the equations below shows the most general dynamic spatial panel model (Elhorst 2014):

$$\boldsymbol{Y}_t = \boldsymbol{\tau}\boldsymbol{Y}_{t-1} + \boldsymbol{\delta}\boldsymbol{W}\boldsymbol{Y}_t + \boldsymbol{\eta}\boldsymbol{W}\boldsymbol{Y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta}_1 + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\beta}_2 + \boldsymbol{X}_{t-1}\boldsymbol{\beta}_3 + \boldsymbol{W}\boldsymbol{X}_{t-1}\boldsymbol{\beta}_4 + \boldsymbol{\upsilon}_{\boldsymbol{t}} \quad \textbf{(Eq.1)}$$

$$\boldsymbol{\upsilon}_{\boldsymbol{t}} = \boldsymbol{\rho}\boldsymbol{\upsilon}_{\boldsymbol{t-1}} + \boldsymbol{\lambda}\boldsymbol{W}\boldsymbol{\upsilon}_{\boldsymbol{t}} + \boldsymbol{\mu} + \boldsymbol{\xi}_t\boldsymbol{1}_N + \boldsymbol{\varepsilon}_{\boldsymbol{t}} \quad \textbf{(Eq.2)}$$

$$\boldsymbol{\mu} = \boldsymbol{\varphi}\boldsymbol{W}\boldsymbol{\mu} + \boldsymbol{\zeta} \quad \textbf{(Eq.3)}$$

$\boldsymbol{Y}_t$ denotes an *N by 1* vector including the mortality rate for each county (1,..,*N*) at time t (1,…,*T*). $\boldsymbol{Y}_{t-1}$ indicates the temporal lag mortality rate from a previous wave and $\boldsymbol{W}$ refers to a *N by N* spatial weight matrix (reflecting the spatial linkages between counties). Multiplying by $\boldsymbol{W}$ generates the spatially lagged mortality rate for each county and τ, δ, and η are corresponding parameters for the temporal and spatial lag mortality, as well as the spatiotemporal lag mortality. $\boldsymbol{X}_t$ denotes a *N by K* matrix including the independent variables and $\boldsymbol{X}_{t-1}$ indicates the temporal lag independent variables. Similarly, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$ are *K by 1* vectors representing the parameters for the contemporaneous, spatial lag, and temporal lag independent variables. $\boldsymbol{\upsilon}_{\boldsymbol{t}}$ is a *N by 1* vector reflecting the error terms specific to the dynamic spatial panel model and $\boldsymbol{\upsilon}_{\boldsymbol{t}}$ is designed to be both spatially ($\boldsymbol{\lambda}\boldsymbol{W}\boldsymbol{\upsilon}_{\boldsymbol{t}}$) and temporally ($\boldsymbol{\rho}\boldsymbol{\upsilon}_{\boldsymbol{t-1}}$) autocorrelated. $\boldsymbol{\mu}$ further captures the spatial errors that may bias parameter estimates and $\boldsymbol{\xi}_{\boldsymbol{t}}$ reflects the temporal errors that are not included in the model. $\boldsymbol{\zeta}$ represents the errors that are normally distributed.

Equations 1-3 includes all possible interactions between mortality rates and the independent variables (i.e., spatial, temporal, and spatiotemporal interactions) and this is why the model is referred to as “dynamic” (Elhorst 2014). Since the dynamic spatial panel model is complex, several scholars have noted that the most general model may encounter the identifiability problem. That is, some of the parameters in Eq. 1 may not be differentiated (Anselin, Le Gallo and Jayet 2008; Elhorst 2012). To counter this, Elhorst (2014) suggests that researchers may need to exclude some parameters. We will use the toolbox for MATLAB developed by Elhorst (2014), which, to our knowledge, is the only toolbox that allows researchers to implement the dynamic spatial panel model. Moreover, a range of diagnostic statistics will be used to compare if the dynamic spatial panel model outperforms the cross-sectional spatial econometrics models, such as Akaike Information Criterion and adjusted R-square.

Timeline

The age-sex standardized mortality rates were calculated using the CMF data and they have been merged with all independent variables discussed above. We will conduct cross-section spatial analysis in October and then implement the dynamic spatial panel model in November. We anticipate to have all results at the end of 2014 and start to draft the manuscript in January 2015 and upload the full manuscript by March 2015.

References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L., J. Le Gallo, and H. Jayet. 2008. "Spatial panel econometrics." Pp. 625-660 in *The econometrics of panel data*: Springer.

Burkhauser, R.V., S. Feng, S.P. Jenkins, and J. Larrimore. 2011. "Estimating trends in US income inequality using the Current Population Survey: the importance of controlling for censoring." *The Journal of Economic Inequality* 9(3):393-415.

Cossman, R.E., J.S. Cossman, A.G. Cosby, and R.M. Reavis. 2008. "Reconsidering the rural–urban continuum in rural health research: a test of stable relationships using mortality as a health measure." *Population Research and Policy Review* 27(4):459-476.

Deaton, A. 2003. "Health, inequality, and economic development." *Journal of Economic Literature* 41:113-158.

Deaton, A.and D. Lubotsky. 2003. "Mortality, inequality and race in American cities and states." *Social Science & Medicine* 56(6):1139-1153.

Elhorst, J.P. 2012. "Dynamic spatial panels: models, methods, and inferences." *Journal of Geographical Systems* 14(1):5-28.

—. 2014. *Spatial econometrics: from cross-sectional data to spatial panels*: Springer.

Hoyert, D.L. 2012. "75 Years of Mortality in the United States, 1935–2010." in *NCHS Data Brief*, edited by NCHS.

Kawachi, I.and T.A. Blakely. 2001. "When Economists and Epidemiologists Disagree." *Journal of Health Politics, Policy and Law* 26(3):533.

Laporte, A. 2002. "A note on the use of a single inequality index in testing the effect of income distribution on mortality." *Social Science & Medicine* 55(9):1561-1570.

Link, B.G.and J. Phelan. 1995. "Social conditions as fundamental causes of disease." *Journal of health and social behavior* 35:80-94.

Lynch, J., G.D. Smith, S. Harper, and M. Hillemeier. 2004a. "Is Income Inequality a Determinant of Population Health? Part 2. US National and Regional Trends in Income Inequality and Age and Cause Specific Mortality." *Milbank Quarterly* 82(2):355-400.

Lynch, J., G.D. Smith, S. Harper, M. Hillemeier, N. Ross, G.A. Kaplan, and M. Wolfson. 2004b. "Is income inequality a determinant of population health? Part 1. A systematic review." *Milbank Quarterly* 82(1):5-99.

Marmot, M.G. 2004. *The status syndrome: How social standing affects our health and longevity*: Times Books.

McLaughlin, D.K., C.S. Stokes, and A. Nonoyama. 2001. "Residence and income inequality: Effects on mortality among US counties." *Rural Sociology* 66(4):579-598.

McLaughlin, D.K., C.S. Stokes, P.J. Smith, and A. Nonoyama. 2007. "Differential mortality across the U.S.: The influence of place-based inequality." Pp. 141-162 in *The sociology of spatial inequality*, edited by Linda M. Lobao, Gregory Hooks, and A.R. Tickamyer. Albany, NY: SUNY Press.

Mellor, J.M.and J. Milyo. 2001. "Reexamining the evidence of an ecological association between income inequality and health." *Journal of Health Politics, Policy and Law* 26(3):487-522.

Muller, A. 2002. "Education, income inequality, and mortality: a multiple regression analysis." *BMJ* 324(7328):23-25.

NCHS. 2000. "Compressed Mortality File, 1968-1988 (machine readable data file and documentation, CD ROM Series 20, No. 2A) ". Hyastville, MD.

—. 2003. "Compressed Mortality File, 1989-1998 (machine readable data file and documentation, CD ROM Series 20, No. 2E) ". Hyattsville, MD.

—. 2012. "Compressed Mortality File, 1999-2010 (machine readable data file and documentation, CD ROM Series 20, No. 2P) ". Hyattsville, MD.

Phelan, J.C., B.G. Link, and P. Tehranifar. 2010. "Social Conditions as Fundamental Causes of Health Inequalities." *Journal of health and social behavior* 51(1 suppl):S28-S40.

Piketty, T.and E. Saez. 2004. *Income inequality in the United States, 1913-2002*: University of California, Berkeley, Department of Economics.
Wilkinson, R.G.and K.E. Pickett. 2006. "Income inequality and population health: a review and explanation of the evidence." *Social Science & Medicine* 62(7):1768-1784.
Yang, T.C., Y.J. Chen, C. Shoff, and S.A. Matthews. 2012. "Using quantile regression to examine the effects of inequality across the mortality distribution in the US." *Social Science & Medicine* 74(12):1900-1910.