Methodological Appendix

Data and Sample

The primary data source for this study is the Panel Study of Income Dynamics (PSID), a longitudinal dataset that has followed families from 1968 to the present. The PSID has been conducted continuously since 1968, annually from 1968 to 1997, and biannually from 1997 to 2009. This study specifically uses data from the 2007 and 2009 survey waves. The PSID collects comprehensive socio-economic information on individuals and families, including employment, income, wealth, and homeownership, as well as demographic information. In addition, information on mortgage distress was collected in the 2009 wave, including whether individuals were behind on their mortgage payments or received a mortgage modification.

The main study sample includes all families observed in both 2007 and 2009. Wage, income, and wealth are inflated to 2009 dollars using the Consumer Price Index research series (CPI-U-RS). All descriptive and regression analyses use weights to account for sample attrition and the oversampling of low-income families.

One particularly useful feature of the PSID for this study is the geocoded data that contain identifiers of census tracts in which sample members have lived in each survey wave. Census tracts are designed by the Census Bureau to be relatively homogeneous and to have an average of about 1,500 housing units and 4,000 residents. They are commonly used boundaries for defining neighborhoods.

Data on neighborhood poverty come from the American Community Survey (ACS) five-year summary file 2005-2009. This file provides a wide range of statistics on the demographic composition of residents in a census tract, as well as average or median income and poverty rate in the census tract. Neighborhood poverty is measured by the poverty rate in each census tract that a family lived in at the time of the 2007 interview from the ACS 2005-2009 census tract poverty rate. This measure is an average over the period of 2005 to 2009; thus it is a mixture of economic booms and downturns. To examine whether the results are robust to the measure of neighborhood poverty, analyses using the census tract poverty rate in 2000 from the Census 2000

summary file were also conducted. The results were similar across datasets; therefore, the more current data is used for these analyses.

In descriptive analyses, neighborhoods are divided into four categories based on the level of family poverty: less than 10 percent poor, 10-19.9 percent poor, 20-29.9 percent poor, and 30 percent or more poor. This division ensures sufficient sample size in the highest-poverty neighborhoods and is consistent with the literature.[1] The main family sample in the descriptive analyses contains 2,040 families living in areas less than 10 percent poor, 1,670 families living in areas of 10 to 19.9 percent poor, 858 families living in areas of 20 to 29.9 percent poor, and 730 families living in areas with 30 percent or more poor.

The data on home price trends come from the Federal Housing Finance Agency Quarterly Housing Price Index (HPI). The HPI provides all-transactions housing price indexes in Metropolitan Statistical Areas (MSAs) each calendar quarter. Area home price changes are measured from percentage changes of HPI between 2007 and 2009 using the relevant quarters before each survey was fielded.

Neighborhood Stabilization Program (NSP) foreclosure need data are used to measure neighborhood foreclosure risks. The U.S. Department of Housing and Urban Development NSP developed scores for census tracts that estimate the number and percent of foreclosures started over the past 18 months through June 2008.[2] Estimated foreclosure risk is used as an explanatory variable measuring neighborhood housing market conditions.

Variable Descriptions

Losses in Income and Wealth

Four measures of losses in income and wealth during the Great Recession are examined. The first measure is whether the household head experienced wage losses of more than 20 percent between 2007 and 2009. The PSID collects information on whether the individual is paid by the hour or by salary on the current main job. For hourly workers, the hourly wage rate is reported. For salary workers, the amount of salary and pay unit are reported. Hours worked is not collected in either case. Hourly wage rates for salaried workers are imputed based on the assumption that they work full time. For example, the wage rates of workers paid per week are calculated by dividing their salary by 40 for those who reported their pay unit as "weekly."

The second measure is whether family income losses exceeded 20 percent between 2006 and 2008. The advantage of this measure is that it captures change among all families—not only those who had employment earnings, but also those who had self-employment earnings and those who did not work. A disadvantage

of this measure is the time period, 2006-2008, which means it may not fully capture the impact of the recession.

The third and fourth measures are based on the share of families with wealth losses between 2007 and 2009. The magnitude of change is measured by the median dollar change and percentage change in total net worth, which includes home equity, among families with wealth losses between 2007 and 2009.

Housing and Home Equity

The percent who owned or rented in both time periods is reported. In addition, among homeowners, the percent who experienced a home equity loss is computed as well as the median percent change among those who both owned a home and experienced a home equity loss.

Mortgage Distress

In 2009, the PSID started collecting information on home mortgages. The questions included whether homeowners with a mortgage were currently behind, how likely they were to fall behind in the next 12 months, and whether they had a mortgage modification.

Analytic Methods

Multivariate regressions are used to estimate how economic outcomes during the Great Recession vary with neighborhood and metropolitan characteristics, as well as individual and family characteristics. The main regression model is as follows:

$$y_i = X_i \, \beta + N_i \, \gamma + \varepsilon_i$$

where y_i is the economic outcome of interest, including changes in employment and income. X_i includes a set of initial individual and family background characteristics. N_i contains a set of neighborhood and metropolitan characteristics, such as initial neighborhood poverty prior to the recession or area housing or labor market changes between 2007 and 2009 to capture the disparate impact of the Great Recession on different areas.[3] ε_i is an error term that incorporates unobserved characteristics of individual i.

The individual and family background control variables include: age dummies (30 to 39, 40 to 49, 50 to 59, 60 and above, with age below 30 as omitted category), education dummies (less than high school, high school diploma only, some college, with college degree and above as omitted category), headship (head is female, head is male with no wife/cohabitant, with omitted category as head is male with wife/cohabitant), race and ethnicity of head (black non-Hispanic, Hispanic, other race non-Hispanic, with white non-Hispanic as omitted category), and family poverty status (whether below the 200 percent poverty line based on

family income in the previous year). All variables are the status of the individual or family as of the 2007 interview.

Neighborhood and metropolitan characteristics include: neighborhood poverty rate dummies at 2007 residence census tract, MSA home price change between 2007 and 2009 (decline less than 10 percent, decline more than 10 percent, with no decline/price increase as omitted category), and census tract percent of foreclosures started over the past 18 months through June 2008. All geographic variables are tied to the individual's 2007 residence or to changes between 2007 and 2009.

As stated, neighborhood poverty in the multivariate models was based upon where individuals resided in 2007. However, individuals and families move residences and are not always rooted to the same neighborhood over time. Analysis of movement between neighborhoods demonstrates that just 12 percent of the total study sample moved to a neighborhood with a different poverty level between 2007 and 2009. With so little movement happening between neighborhoods of different poverty levels during the Great Recession, results in the multivariate models cannot be attributed to geographic mobility among the study sample.

All dependent variables in the regression models are discrete variables and a multinomial logit model is used, with the marginal effect reported rather than the coefficients. That is, the change in the likelihood for an infinitesimal change in each continuous variable and the discrete change in the likelihood for dummy variables are reported. Standard errors are not clustered by individual/family as outcome variables, nor are they clustered by census tract, as about 60 percent of census tracts in our sample contain only one family, and there are multilevel geographic characteristics such as census tract, MSA, and county.

Discussion on Neighborhood Effect

Empirical studies on neighborhood effects are subject to multiple estimation problems such as omitted variable bias, endogenous neighborhood selection, and the reflection problem.[4] Some studies use fixed effects or first difference estimators when panel data are available. Other studies use experimental data to control for selection bias—that individuals with certain characteristics choose to live in a certain neighborhood. Quasi-experimental approaches such as using regional variation as an instrumental variable are also used in the literature.

This report focuses on the disparate impact of the Great Recession on residents of neighborhoods with different poverty levels. This study does not identify a causal neighborhood effect. Rather,

the report describes how the Great Recession is associated with various economic outcomes in poor and nonpoor neighborhoods. The study goes beyond existing research on the relationship between the Great Recession and economic outcomes for individuals and families with certain characteristics (e.g., gender, age, education, and race and ethnicity). The possibility of identifying a pure neighborhood effect is limited by the nature of the data—about 60 percent of the neighborhoods in the sample do not have multiple families to control for fixed or random neighborhood effects.

Nonetheless, additional analysis examines whether the association of the neighborhood poverty and economic outcomes in the Great Recession is related to the role of individual characteristics. First, the raw correlation between neighborhood poverty rate and family poverty status (0/1) is only 0.23. Second, the links between neighborhood poverty and family poverty status are examined using a sequence of regression models: (1) model with three neighborhood poverty dummies but not family poverty status; (2) model with family poverty status but not neighborhood poverty measure; (3) model with both family and neighborhood poverty[5]; and (4) model with both family and neighborhood poverty, with a continuous measure of poverty rate rather than three dummies.[6]

The findings indicate that including or excluding one poverty measure (family/neighborhood poverty) does not affect the precision of the other poverty measure estimate (i.e., the standard error). After including family poverty status, the estimated effect of neighborhood poverty dummies does shrink a bit and suggests that family level poverty explains some variation in economic outcomes. Thus, this variable is retained in the regression model. Using a discrete or continuous measure of neighborhood poverty has little effect on estimates of family poverty status or of other explanatory variables. Results presented in this report use discrete measures of neighborhood poverty, as its relation to economic outcomes is considered to be nonlinear.

Endnotes

1 For example, Sharkey, Patrick, 2009. "Neighborhoods and the Black-White Mobility Gap." Washington, DC: The Pew Charitable Trusts, Economic Mobility Project. http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/Economic_Mobility/PEW_SHARKEY_v12.pdf?n=1399.

2 Estimated foreclosure rate is calculated from Federal Reserve Home Mortgage Disclosure Act data on high-cost loans, Office of Federal Housing Enterprise Oversight data on falling home prices, and Bureau of Labor Statistics data on county unemployment rates. More technical details can be found at http://www.huduser.org/DATASETS/Desc_%20NSP_data.doc.

3 Both neighborhood and individual characteristics are in the model. Individual characteristics are not indexed with t (time period), as the outcomes are one-time changes between 2007 and 2009. Ideally, there would be a control for neighborhood fixed effects. However, 59 percent of census tracts in the PSID data contain only one family per census tract. Controlling for neighborhood fixed effects would leave out 59 percent of the observations.

4 For example, Manski, Charles F., 1989. "Anatomy of the Selection Problem." *Journal of Human Resources* 24(3), 343-60. And Manski, Charles F., 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives* 14(3), 115-36. For a literature review on theoretical and empirical studies of neighborhood effects, see Haurin, Donald R., Robert D. Dietz, and Bruce A. Weinberg, 2003. "The Impact of Neighborhood Homeownership Rates: A Review of the Theoretical and Empirical Literature." *Journal of Housing Research* 13(2): 119-51.

5 As the correlation between family and neighborhood poverty is only about 0.2, including both variables in the regression does not have a multicollinearity issue.

6 In all these regression models, explanatory variables do not include any neighborhood or metropolitan characteristics (other than neighborhood poverty).