Building a Microdata Infrastructure to Study Environmental Risk and Inequality: The Census Environmental Impacts Frame

Jonathan Colmer, Kendall Houghton, Cameron Scalera, John Voorheis[1] and Jenny Withrow

It is well-established that low-income and disadvantaged communities are disproportionately exposed to environmental risk. However, much of the work in this space exploits differences in exposure between locations, using the demographic characteristics of neighborhoods, to make claims about environmental inequality. Until now, detailed microdata combining information on the demographic, economic, and residential circumstances of individuals has not been available to researchers. This project presents a new microdata infrastructure using confidential Census Bureau microdata drawn from surveys, administrative records, and Decennial censuses – the Census Environmental Impacts Frame (EIF). This work allows us to move from a *place*-based to a *person*-based understanding of environmental inequality. This project presents the structure and composition of the EIF and applies it to the study of wildfire exposure as an illustration of what the EIF can provide.

Historically, research on the distributional effects of environmental risks has been limited due to insufficient sample sizes in survey data, which have impeded serious study of heterogeneity along margins of race and income, or due to the absence of socio-demographic information in administrative tax data. The EIF removes this constraint. The EIF will provide researchers with the ability to follow individuals over time and space, providing detailed information on exposure to environmental risks, alongside their economic well-being, labor market activities, demographic characteristics, and family structure. This framework presents new opportunities for researchers to engage in a more systematic and comprehensive analysis of how exposure to environmental risks are distributed, their causes and consequences, and how environmental inequality shapes more systemic patterns of economic opportunity and inequality through the economic and policy environment. The EIF can also be extended and updated as new data becomes available, allowing us to extend the economic panel and incorporate advances in the measurement of environmental risk as computational advances and new frontier measures become available.

The core of the EIF is a population spine, consisting of all individuals who have ever received a Social Security Number, with information on individual's place and date of birth, basic demographics, and date of death if deceased. We construct residential histories for individuals resolved to exact latitude and longitude using administrative records from the Internal Revenue Service, Department of Housing and Urban Development, Center for Medicare and Medicaid Statistics and the US Postal Service. We construct a panel of economic characteristics, containing information on family income and structure, individual earnings, and the characteristics of the firms for which individuals work. Finally, we overlay a core set of environmental hazard geospatial data (air pollution, weather extremes, flood risk, wildfires, toxic releases, and hurricane exposure).

The EIF's structure is ideally suited for exploring issues around the distribution of environmental hazards across income and demographic groups, particularly as it intersects with the development of green

[1] Corresponding Author. Any opinions and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Census Bureau.

national accounts. Following recent developments in the national accounts literature which have paid particular care to distribution (Piketty, Saez and Zucman, 2018; Auten and Splinter, 2019), efforts to build a green national accounting infrastructure in the United States should take into account the distribution of environmental costs and benefits. This exercise requires a microdata infrastructure such as the EIF, as it is necessary to measure how environmental damages by the distribution of (national) income.

As an illustration of the EIF's potential, we use a prototype to study the demographic composition of exposure to wildfire risk and wildfire smoke air pollution across individuals over time. We contrast this exposure with overall particulate matter air pollution. We intersect the EIF with daily fire burn extent data from U.S. Geologic Survey, wildfire smoke plume data from NOAA's Hazard Mapping System, and newly available satellite data on fine particulate matter. We are thus able to explore both the level of inequality in exposure to wildfire smoke and other forms of particulate matter pollution, as well how disparities in exposure across individuals has evolved over time.

We document that, on average, Non-Hispanic Whites endure disproportionately more exposure to above-average levels of annual wildfire smoke relative to other race and ethnic groups with very little variation in wildfire smoke exposure across the income distribution within race and ethnic groups. We also observe large urban-rural gaps in smoke exposure, with much more intense exposure for residents of rural areas. However, over time these demographic patterns are changing. Between 2006 and 2019, the number of wildfire smoke days experienced on average has tripled. The average American now experiences between 20 and 30 days of smoke a year. Among different race and ethnicity groups, Black Americans experienced the highest average year-to-year growth rate in smoke. While exposure to wildfire-related smoke is worsening for Black Americans, overall levels of ambient particulate matter exposure and absolute Black-White particulate matter disparities have improved over time (Colmer et al., 2020; Currie, Walker, and Voorheis, 2021). We examine the degree to which these changing patterns of wildfire smoke exposure may undermine these improvements.