The Effect of Birth and Contemporaneous Pollution Exposure on Health Outcomes

D. Phaneuf (PI), C.Ilin (Co-PI)

Data requested (ongoing IRB process):

Source 1: California Office of Statewide Health Planning and Development

- Patient Discharge Data (1991-2017)
- Emergency Department Data (2005-2017)

Source 2: California Department of Public Health

- Vital Statistics Data (1991-2012)

Project Description

There is a large literature in environmental economics that examines the contemporaneous relationship between pollution and health outcomes. For example, Schlenker and Walker (2016) measure how hospital admissions for respiratory and cardiac conditions on a given day are affected by local air pollution concentrations. A newer literature examines how exposure to pollution conditions in utero or during the first year of life can affect economic outcomes later in life. For example, Isen et al. (2017) examine how exposure to total suspended solids during the year of birth can influence wages and labor force participation in early adulthood. These two strands of research have shown that both short run and long run exposure to pollution can have sizeable impacts on health and economic outcomes. However, little is known about how the two pathways interact. We propose a research design that will allow us to jointly assess the role of birth and contemporaneous pollution exposure on a suite of health outcomes.

To illustrate, suppose we observe a sample of B births each indexed by (i,bz,b), where i is an individual, bz is the location (county or zip code) where the person was born, and b is the year of birth. We also observe later in life hospital admissions for a subset of the births. Denote the set of births for whom any type of later hospital care is observed by $N<B$. In what follows we refer to the sample of N individuals as our analysis sample. Let H_{izt} be an indicator variable equal to one if person i visits a hospital at time $t > b$ for respiratory or cardiac conditions, while living in location z. For concreteness, we assume that t denotes a unique quarter/year combination, so that the temporal unit of analysis is a quarter.

We are interested in estimating how birth and contemporaneous pollution exposure affects H_{izt}. Let $P_{i,bz,b}$ denote the pollution level at place bz during year b. We assign this level of birth exposure to the cohort of individuals in the analysis sample who were born in the (bz,b) year and place. Likewise, let P_{izt} denote the pollution level a person is exposed to during time t while living at place z. We assign this pollution level to all people in the analysis sample living in location z at time t. With these variables in hand we are interested in specifications of the following type:

$$H_{izt} = \beta_1 P_{i,bz,b} + \beta_2 P_{izt} + \beta_3 P_{izt} \cdot P_{i,bz,b} + \gamma X_{izt} + \theta_i + \tau_t + \varepsilon_{izt}, \quad i = 1,\ldots,N, \quad t = 1,\ldots,T, \qquad (1)$$

where T is the number of quarters we will include in the analysis, X_{izt} is a vector of controls, θ_i is a person fixed effect, and τ_t is a fixed effect for quarter t. We are first interested in estimating (β_1,β_2,β_3), which together measure the average joint impact of birth and contemporaneous exposure on the probability of a respiratory or cardiac related hospital admission.

The vector of controls X_{izt} can also include interaction terms between individual characteristics and the pollution variables, which will allow us to estimate heterogeneous responses to pollution exposure. For example, observing race or other socio-economic indicators will allow us to measure if disadvantaged populations are differentially affected by pollution. Similarly, observing conditions of a person's birth, such as birth weight or early life health conditions, will allow us to test if disadvantages stemming from in utero or neonatal conditions affect susceptibility to pollution later in life.

Note the importance of limiting the analysis sample to the N individuals observed at some point in the hospital admissions data. This is because we want to be able to observe current location for all analysis sample members, in order to assign current pollution exposure. By seeing people in the hospital use files we can make a good guess at their location. In contrast, if we do not see an individual after observing his birth, we do not know where he lives later in life. This means that the analysis sample is the set of people seen in the hospital data, and we assume that the people in the hospital for non-cardiac or non-respiratory conditions – especially for accidents or trauma not related to sickness – are good counterfactuals for people who visit for cardiac or respiratory reasons. An alternative to this is to assume that people stay in their birth county if we do not observe them later in life.

Note as well the importance of observing characteristics of individual birth circumstances. While equation (1) includes a person fixed effect, it will likely be the case that other specifications will use a coarser set of controls. For example, an alternative to (1) is

$$H_{izt} = \beta_1 P_{i,bz,b} + \beta_2 P_{izt} + \beta_3 P_{izt} \cdot P_{i,bz,b} + \gamma X_{izt} + \theta_z + \tau_t + \varepsilon_{izt}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2}$$

where the only difference is the use of a zip code fixed effect θ_z rather than person fixed effect. This implies that our vector of controls X_{ict} will need to include a larger set of variables in order to adequately account for the many different drivers of individuals' overall health.

Literature:

[1] Schlenker, Wolfram, and W. Reed Walker. "Airports, air pollution, and contemporaneous health." *The Review of Economic Studies* 83.2 (2015): 768-809.

[2] Isen, Adam, Maya Rossin-Slater, and W. Reed Walker. "Every breath you take—every dollar you'll make: The long-term consequences of the clean air act of 1970." *Journal of Political Economy* 125.3 (2017): 848-902.