Health and the City: Urban Congestion and Air Pollution in Brazil *

Marcos A. Rangel and Romina Tome

Sanford School of Public Policy, Duke University

August, 2018

Abstract

Congestion of urbanized centers is a trademark accompanies of economic development across the globe. These economic forces pose a challenge with respect to environmental conditions and, therefore, welfare for populations in those areas. Nevertheless, the evidence to inform policy-makers in the developing world is still scarce. In this paper, we take advantage of the meteorological phenomenon of thermal inversion which in urban areas arguably exogenously lock pollutants closer to the ground to estimate the causal effects of pollution on infants health at birth. We employ detailed data from birth records around the metropolitan area of São Paulo, Brazil, between 2002 and 2009. Our preliminary results confirm a positive relationship between thermal inversions and several air pollutants and suggest that exposure during the last three months of pregnancy to thermal inversions has a negative effect on health at birth.

*Work in progress. Please, do not quote or cite without authors' permission.

Intoduction

Congestion of urbanized centers is a trademark of economic development across the globe. These economic forces pose a challenge with respect to environmental conditions and, therefore, welfare for populations in those areas. While empirical evidence has helped shape environmental regulation in developed nations, there is a scarcity of evidence to inform policy-makers in the developing world. The level of air pollution is higher and population's health is worse in the latter than in the former preventing to extrapolate the results found in richer countries, and poorer nations could benefit from further knowledge of potential nonlinear effects of pollution.

In this paper, we focus on air pollution and undertake a causal inference study of its impacts on the health of infants *in utero* in a metropolitan area of Brazil. Specifically, we study the impact of prenatal exposure to different pollutants on health at birth for children born in the São Paulo Metropolitan Area – a place where congestion and traffic represents one of the main sources of pollution. The identification of the causal effect of air pollution on health is challenging for many well known reasons, such as families self-selection into residential location and avoidance behaviors, and the link between economic activity and pollution. To address these concerns, we take advantage of the meteorological phenomenon of thermal inversion which in urban areas arguably exogenously locks pollutants closer to the ground. Relevant for our identification strategy, the formation of thermal inversions do not have any direct effect on health conditional on weather conditions.

Our empirical strategy consists in two parts. We focus first on understanding the effect of thermal inversion on air pollution. Using data on concentrations of five pollutants and weather conditions between the years 2001 and 2009, we compare the effect of all inversions and those that are more intense based on the height of their base and the change in temperature. We find that the larger effects on the accumulation of pollution is coming from those inversions that occur closer to the ground. For PM_{10}, an additional inversion within the last week increases the accumulation of pollution by 0.4 micrograms per cubic meter in that week. Our second estimation looks at the link between thermal inversions and health at birth. By employing detailed data from vital records around the São Paulo Metropolitan Area (SPMA), Brazil, between 2002 and 2009 we detail cumulative effects of air pollution on infants' health at birth.

Our preliminary results show that exposure to thermal inversions that occur below 1.3 km (relative to sea level) harms health at birth. Specifically, an extra inversion per week

occurring in the last 13 weeks of gestation leads to a decrease in birth weight by 4.4 grams, an increase of 2.4% in the incidence of low birth weight, and a raise of 4.2% in the incidence of very low birth weight. These effects on birth weight are partly explained by a reduction in gestational length: there is an increase of 1.7% and 10% in rates of preterm and very preterm births, respectively, per an additional per week inversion that occurs in the last gestational period. Our results are precisely estimated. We do not find effect of additional inversions in earlier periods of gestation.

Our work builds on Arceo et al., (2016), who study the contemporaneous effect of pollution on infant mortality in Mexico City using thermal inversions as an instrument for pollution. We complement this study by examine the effects of prenatal exposure to inversion episodes on birth outcomes.

We go on to extend our analysis to investigate urban policy and technological innovations during the period we study. In particular we examine the regulation of minimum ethanol content in gasoline and the development and adoption of flex-fuel cars (running 100% on ethanol).

Background

The SPMA consist in 39 municipalities that extends over 8.000 km^2 with altitudes that go from 650 to 1200 meters. More than 19.7 million inhabitants reside there in 2009 – making it one of the ten largest urban conglomerates in the world. The municipality of Sao Paulo, with 11 million people, is in the center of this dynamic urban area.

The number of vehicles increased significantly between 2001 and 2009: from 6 millions to 9.7 millions (CETESB, 2002, 2010). That represents an increase in the number of vehicles per 1,000 inhabitants from 331 to 492. While this is lower than some cities in the United States, it is still higher than most other Latin American cities and contributes to a mobility crisis; São Pauloâs infrastructure was not built to accommodate such a large number of private vehicles. For instance, during the first decade of this century, the daily average traffic jams during rush hours in the city of São Paulo has been 118 kilometer (112 miles), and cars circulate during rush hour with an average speed of 19.30 km/h (Rolnik and Klintowitz, 2011). As a result, people waste almost one month each year trapped in traffic.

Considering the high level of congestion in the SPMA, it is not surprising that traffic is one of the main sources of pollution. Indeed, in 2009, motor vehicle emissions were responsible of 97% of the emissions of carbon monoxide (CO), 96% of nitrogen oxides (NO_x), 40% of

particles matter under 10 micrometers (PM_{10}), 97% of hydrocarbons (HC), and 32% of sulfur dioxide (SO_2) (CETESB 2010).[1] Hydrocarbons, nitrogen oxide, and other volatile organic compounds contribute to the formation of ozone by reacting with sunlight (Kinney et al., 2011). Moreover, a comparison of the emissions of CO and NO_x during weekdays and weekends shows peack of them at rush hour in the morning and evenings of the former; ozone concentrations behave differently with higher concentrations during the afternoon between 2 pm and 3 pm given the solar radiation at that time (Carvalho et al., 2015).

The increase in the fleet of cars in the last years coincide with the introduction of policies at the state and local level oriented to reduce congestion, and they might have prevented an even higher escalation of the total emissions (Carvalho et al., 2015). Furthermore, other factors that might have been related to the dynamics of emissions from traffic over time are changes in the price and production of gasoline and ethanol.

The other sources of pollution in this area are industrial processes, waste burning, and handling and storage of fuels (CETSB, 2002). Indeed, industrial processes – with almost 2,000 high polluting firms – account for the rest of the emissions of CO, HC, NO_x, and SO_2. In the case of PM_{10}, secondary aerosol formation and resuspension of soil particles, are a bigger contribution to emissions than industrial processes (CETERB, 2010).

In the SPMA, meteorological conditions are unfavorable to the dispersion of pollutant: during winters, there are frequent low altitude thermal inversions (as we show below), and during other seasons, strong solar radiation occur. Moreover, the topography of the SPMA contributes to the deterioration of air quality considering that the area is formed by floodplain surrounded by mountains in the north and northwest, receiving predominant winds from the ocean at southeast – approximately 60 kilometers away (Ribeiro and Assuncao, 2001).

Conceptualization of potential health impacts

Urban air pollution has become one of the most critical issues in metropolitan cities everywhere. The rapid growth of urban transportation and industries that characterized the urbanization in many developing countries makes the environmental problems much greater. While these processes are positively related to economic activity, they also introduce a health burden. We focus in this paper on the consequences of exposure to air pollution on health at birth.

[1] Overtime, the contribution of vehicle emissions do not change significantly for CO, NO_x, HC, and PM_{10}; the contribution was higher for SO_2 in 2001 reaching 56% (CETESB, 2002).

Pollution may affect infants health while in the womb and after birth. Before birth, exposure to air pollution may affect fetal development through the effect on the mother's health that reduces nutrients required to development. It may also have a direct effect on infants if it crosses the protective barrier of the placenta. Consequences of such exposure include interrupted placental development, fetal growth restriction, higher likelihood of a preterm delivery, heart anomalies, and lower weight at birth. After birth, air pollution may cause several health issues such as respiratory and cardiac troubles. These later effects might be influenced by the effects before birth that make children more vulnerable, yet there is not a clear understanding of the functional form or magnitude of these interactions.

A growing body of the economic literature has shown that pre- and post-natal exposure to urban air pollution has an adverse effect on infant health. Findings in these studies point to the same conclusion: air pollution has a detrimental effect on health at birth and increases infant mortality (e.g., Chay and Greenstone, 2003; Currie, Neidell, and Schmieder, 2009; Currie and Walker, 2011, Knittel et al., 2016). Prenatal exposure to air pollution also has consequences after birth, for instance, on performance at school (Sanders, 2012). While these studies provide evidence of the effects on health in developed countries, their implications for urban air pollution in developing countries remain unclear. In particular, the differences in air pollution levels and population's health between developed and developing nations restrict the extrapolation of results from the former.

In developing countries, epidemiological studies have shown a correlation between air pollution and health (see, for example, Gouveia, et al., 2003; Pereira et al., 2004; Borja-Aburro et al., 1997). These cross-sectional articles fail to account for several sources of confounding effects. First, families may self-select into residential location. If those who prefer to live in cleaner places are also healthier, wealthier, or have greater access to health care, and/or make higher quality investment in their children, the results will be biased upward. Second, the estimates would be biased downward if families adopt avoidance behaviors when they face environmental risks. Third, air pollution may be correlated with economic activity, and the latter also may affect infant health directly.

One of the few papers using applied microeconometric techniques in a developing country, instruments air pollution with thermal inversion episodes in Mexico City (Arceo-Gomez et al., 2016). The authors link the number of thermal inversions per week, weekly pollution levels, and contemporary infant mortality. They show that increasing PM_{10} or CO by 1% raises infant mortality by 0.42% or 0.23%, respectively. Our work builds on this study. As

an extension, we explore the effects of pollution on fetal development. This analysis will allow us to understand how the effects after birth differ depending on the level of exposure before birth. Moreover, even though we also use thermal inversions to address the well known concerns related to estimating the effects of air pollution on health, we take a more conservative approach to minimize the difficulties that the lack of data on all pollutants brings (Jans et al., 2014).

Meteorological shocks

Thermal inversion layers occur when temperature increases with altitude.[2] Normally, temperature decreases with height allowing air pollutants to disperse. When there is a mass of hot air on top of the mass of cold air, pollutants are trapped close to the ground. Indeed, heavy air pollution episodes have been associated with thermal inversions (e.g., in London, Logan, Santiago, among others).

Several factors contribute to the formation of a thermal inversion. For instance, a necessary (but not sufficient) condition to have an inversion episode is stable air that inhibit vertical and horizontal mixing near the ground favoring the development of inversions (Jacobson 2005). This condition is common during nights and winters when cold ground temperatures cool the air that is close to the ground creating warm air over cold air. When sun warms the ground during the daylight hours, inversions weaken or disappear. Wind speed and rain also contribute to weaken inversions.

The shape of the landscape has a role in the formation, lifetime, and intensity of inversion (Redon et al., 2014). In valleys, denser and heavier cool air flow down the slopes and settler under the warm air leading to stronger effects. Inversions can be produced when a layer of cool air descends through a layer of hot air from vertical air movements or from horizontal movements of air at different temperatures (Jacobson, 2005).

While related to a combination of meteorological and topological conditions, the formation of thermal inversion is independent of economic activity, and they they do not represent a health risk by themselves. Yet, the accumulation of pollution that occurs when there is an inversion episode is likely to be risky. Moreover, the meteorological factors associated with them (e.g., temperature) may directly impact economic activity and health. The effect on health may be larger when inversions lock pollutants closer to the ground and when there is large variation in temperature leading to stronger locking of pollutants.

[2]Definition provided by the Meteorological Glossary of the American Meteorological Society.

Data

To study the effects of exposure to air pollution on health at birth, we combine several datasets. The following sections describes each data source. We look at the universe of births in the SPMA between 2002 and 2009. To being able to analyze the time *in utero* for all of them, we examine air pollution and thermal inversions in the period from April, 2001 to December, 2009.

Pollution

Air pollution data come from air monitoring stations operated by the State of São Paulo's environmental agency CETESB – *Companhia Ambiental do Estado de São Paulo*. We study five pollutants: particulate matter under 10 micrometers (PM_{10}), carbon monoxide (CO), ozone (O_3), nitrogen oxides (NO_x), and sulphur dioxide (SO_2). For the period 2001-2009, PM10 data is available for 25 stations; CO is collected in 16 stations; O_3 is available in 15 stations; NO_x is drawn from 12 stations; and SO_2 is available for 8 stations.

Data are organized in hourly observations of pollution concentrations at the station level. For all pollutants except CO, we use these observations to calculate daily average concentration of pollutants for every day with at least 8 hours of raw data. For carbon monoxide, we use the hourly observations to calculate the maximum daily 8-hour average. Then, for all pollutants, we compute rolling week averages. That is, for each date t, we take the average of the daily measure for the week from $t-6$ to t.

We present in the top panel of Table 1 descriptive statistics of the pollutants we use in the regressions. The mean level of pollution in the SPMA is higher than in California but lower than in Mexico City based on levels of PM_{10}, CO, and SO_2. For instance, the average level of PM_{10} in our data reaches 42 micrograms per cubic meter ($\mu g/m^3$) whereas that mean in California is below 30 $\mu g/m^3$ (Knittel at al., 2016) and it is around 67 $\mu g/m^3$ in Mexico City (Arceo et al., 2016). However, the mean levels of ozone in the SPMA and Mexico City are similar (around 33 $\mu g/m^3$).

Weather

Weather covariates are key elements in our analysis given that they are directly related to thermal inversions, economic activity, and health. Temperature, relative humidity, and wind data are collected by CETESB. Rainfall data come from the São Paulo department

of water and energy DAEE -*Departamento de Águas e Energia Elétrica*. We convert daily hourly observations of all weather variables, except wind, into daily averages if there are at least 8 hours of raw data for that date. We then compute rolling week averages from daily interpolated data.

For wind, we consider the daily prevailing wind as our measure of daily summary. From raw data that are coded as angles in degrees (0 indicates wind from due North, and 180 corresponds to wind from due South), we define the daily prevailing wind as the sector of the wind rose that contains the most hourly measurements we use fixed octants in our analysis.

Not all the CETESB stations that collect air pollution data, recover weather data too, and some stations that provide weather data do not measure the concentration of pollutants. For those stations collecting only pollution data, we construct weighted daily averages of weather covariates employing data from stations within a radius of 20 miles. We use the inverse of the distance between stations as weights.

In the middle panel of Table 1, we present mean and standard deviation of weather covariates. These statistics are based on observations measured at the 25 stations that collect PM_{10} data. The statistics do not change significantly when we look at fewer stations such as the ones measuring other pollutants.

Thermal Inversions

To identify thermal inversion episodes we use vertical temperature profile of data collected by CETESB and the University of Wyoming. Balloons are launched twice each day, at 12 UTC and 00 UTC (10 am and 10 pm of the night before in Brazil, respectively); they gather temperature, humidity, and wind data as they ascend through the troposphere.

We use all diurnal data to find boundary layer inversions every time a temperature at a given altitude was warmer than the temperature at an altitude below it. The opposite non-monotonic temperature gradient allows to identify the top of an inversion. More than one inversion can be found each day at different altitudes. We consider in the analysis those thermal inversions that occurs in layers above 731 meters of the ground, and we focus for each date on the inversion that occurs closer to the ground.[3]

We use these daily measures to construct for each date rolling week counts of inversion

[3]The station collecting thermal inversion data is located at 731 meters above sea level, so we ignore inversion episodes that occurred below that level.

episodes between that date and the previous six dates. In less than 7.2% of the dates of our period of analysis, thermal inversion data were not collected. We take a conservative approach to deal with these missing values: we assign a zero to each rolling week with at least one date with missing information, and we control semi-parametrically in our specification for the number of missing values in each week.

Considering that inversions differ in certain characteristics, we explore in our analysis whether certain type of episodes have a larger effect of locking pollutants closer to the ground. Specifically, to characterize the strength of each inversion, we distinguish inversions by the height of their base and compute the difference between the temperature at the the top and base levels of the inversion layer. We classify inversions as being close to the ground when they occur below 1.331 meters from the sea level. The 1.3 km limit is based on the fact that the station collecting inversion data is located 731 meters above sea level and that inversions that occur below 600 meters above the ground are expected to have stronger effects (Jans et al., 2014). Moreover, we divide the inversion episodes in two groups: those that have a "high" and "low" change in temperature considering if that change in above or below the percentile 75 of the distribution of the differences in temperature for all inversions.

In the bottom panel of Table 1, we present descriptive statistics of thermal inversion episodes.[4] In 71% of the weeks there are at least one inversion episode, and on average, there are 4.6 inversions per week. Thermal inversions occur close to the ground (below 1.3 km) in 66% of the weeks. The mean number of these episodes is 2.2 per week. Table 1 also shows that in 29% of the weeks there is at least one date without inversion data, and they represent on average less than a date per week.

Figures 1 and 2 describes characteristics of the thermal inversion patterns in Brazil. Figure 1 provides evidence of the accumulation of pollutants when there are thermal inversion episodes. For each pollutant, we present three distributions based on having a thermal inversion below 600 meters, having a thermal inversion above 600 meters, or not having a thermal inversion (or missing information). As expected, the concentration of pollutants has a distribution centered at a higher level of pollution when an inversion occurs closer to the ground. This pattern is not seen for O_3 a pollutant that is affected in a different way than other pollutants by sunlight and weather conditions.

The seasonality patterns associated with thermal inversions are clearly illustrated in

[4]The statistics in Table 1 are based on the final data; that is, dates with missing inversion data was replaced by 0.

Figure 2.[5] Based on the results presented in the previous paragraph, Figure 2 focuses on inversions that occur up to 600 meters from the ground. For each calendar month, it shows the average number of inversions in a week (bars), and it compares the count with the weekly average of weather variables: ground temperature, relative humidity, and rainfall. Weather conditions and thermal inversions are negatively correlated. During the spring (October to December) and the summer (January to March), the average weekly temperature, humidity, and rainfall are higher than during the fall (April to June) and the winter (July to September), while the average number of weekly inversions is lower during warmer seasons than the colder ones.

Birth Data

We study singletons born in hospitals in the SPMA between 2002 and 2009. Data come from individual-records in DATASUS, the Brazilian Ministry of Health's Usage Information System. Brazil vital registration system provide a complete coverage of all birth. Indeed, in 2010, 99% of births and infants death were registered (IBGE, 2011). Vital records for all births in the SPMA are included in the analysis. They provide information about newborns' health and their mothers' characteristics. Relevant for our study, these confidential data allow us to identify where children reside .

To link daily rolling weeks of weather and pollution data to vital records we use the information about location of residence available in vital records. We approximate local conditions by averaging location-level weather and pollution in each date. We proceed in three steps based on the information available. First, for those newborns whose mothers reside in the São Paulo City and we recover their district of residence, we use contour data from *Centro de Estudos da Metropole* (CEM/USP) to identify the district polygon centroid. Second, for infants in the capital with unidentified district, we use the weighted polygon centroid employing 2000-Census population as weights for each district listed on the CEM/USP data. Third, for newborns who reside at birth in municipalities outside the São Paulo City, we use the weighted centroid using all districts within that municipality and 2000-Census population as weights. So, we identify 135 locations of residence (97 districts in the capital and 38 municipalities outside of it) and employ those that have a weather station within 20 miles in the regressions below.

[5]For illustration purposes, weather conditions are estimated using reading from the station that collects thermal inversion data – Santana station. This station also collects some weather variables; for the rest, we use a distance-weighted average of observations in stations located close to Santana station.

Then, we calculate the distance between a centroid and each monitor station using latitude and longitude location data. We compute a weighted average of the weather variables and pollutants concentration weighting the data measured in each station by the inverse of the distance from the station to the centroid.

Our outcomes are birth weight and prematurity. Weight at birth is measured continuously in grams and as an indicator for low or very low birth weight (<2,500 grams or <1,500 grams, respectively). Gestational age at birth is coded in categories in the Brazilian birth certificates. We define indicators for preterm (<37 gestational weeks) and very preterm (<32 gestational weeks).

To reduce computational demands, we collapse observations at individual level by date of birth and location of residence. Our final dataset includes 316,970 location-date cells. We perform the analysis of the effect of air pollution on infant health at the location-date level, and we weight the analyses by cell size.

In Table 2, we show descriptive statistics of the health outcomes. Newborns weight on average 3.2 kilograms. 82 per 1,000 infants are low birth weight (< 2,500 grams), and 12 in 1,000 weight at birth less than 1,500 grams (very low birth weight). Prematury (< 37 weeks) is found in almost 72 out of 1,000 newborns; 11 in 1,000 infants are very premature (< 32 weeks). We also include in Panel B of Table 2 the mean and standard deviation of the covariates we use in our empirical strategy – that coincide with those in Table 1.

Estimation strategy

We aim to estimate the effect of air pollution on infant health at birth. Selection into residential location, avoidance behaviors, and the correlation between air pollution and economic activity prevent the estimation of unbiased effects from the simple study of pollution and health cross-sectional data. Furthermore, the link between thermal inversions and pollutants that are not in our data limit the interpretation of the effects of each particular pollutant using inversions as an instrument of it. In other words, many pollutants are locked in an inversion episode. These pollutants might be correlated with each other, and they also have a direct effect on infants' health or economic activity. So, the unobservables in a second-stage regression for an IV estimation would be correlated to the outcome and to the instrument.

In order to avert these concerns, we first formally test the pollution effects of thermal

inversion episodes, which is akin to a first stage relationship. In these regressions, we explore the effects of inversions with different characteristics. We use the results from out first specification to inform the models in our second estimation that link thermal inversions to health outcomes. In the last one, we estimate an intention-to-treat effect.

Effects of Thermal Inversions on Air Pollution

To explore the effect of thermal inversion episodes on pollution concentrations, we focus on PM_{10}, but we also examine other four pollutants (CO, O_3, NO_x, and SO_2). Using observations at the station (s)-date (t) level, we estimate models of the form:

$$Poll_{st} = \beta_0 + \beta_1 \; Inv_t + X'_{st} \; \gamma + \mu_{ws} + \mu_y + \epsilon_{st} \tag{1}$$

where $Poll_{st}$ is an average daily pollution concentration from date $t-6$ to t measured by station s. Inv_t indicates the number of thermal inversion episodes that occurred from date $t-6$ to t. The vector of covariates, X_{st}, includes an indicator for having at least one date between $t-6$ and t with missing inversion data, a count of the number of dates in that week without inversion data, and weather variables: cubic polynomials in weekly average temperature, weekly maximum temperature, weekly minimum temperature, weekly average humidity, direction of wind (fixed octants), and weekly average rainfall. We include temperature, humidity, wind, and rainfall because thermal inversions formation and population's health are related to the,. Equation (1) also includes station-by-weak of the year fixed effects, μ_{ws}, to control for seasonality effects within each location, as well as year fixed effects, μ_y, to control for common changes over time. We use two-way clustering, and standard errors are clustered at the station and date level to account for potential spatial correlation and common weather trends.

In Model (1), we estimate the effect of any extra inversion episode between date $t-6$ and t on the concentration of pollutants. Considering that some characteristics of inversion episodes might affect their scope to lock pollutants, we run two additional models differentiating inversions by their distance from the ground and their intensity. Specifically, these models follow the first one and focus on alternative independent variables.

In Model (2), the central explanatory variable, $Inv1km$, indicates the number of inversion episodes that occurred from date $t-6$ to t below 1.331 meters from the sea level. This height limit considered the altitude of the monitor station that measures inversions and the altitude

from the ground that is considered in the literature for inversions with stronger effects. The specification we use for Model (3) includes in our first model a variable that accounts for the number of inversion between date $t-6$ and t that have a temperature difference between the top and the base that is above the 75th percentile of the difference in temperature for all the inversions ($InvHigh\Delta Temp$).

Table 3 presents the results of these models. Column (1) shows that an extra thermal inversion in the preceding week increases the average concentration of PM_{10} by 0.2 units in that week. The accumulation of CO, NO_x, and SO_2 also raises with one additional inversion in the past week. The negative effect on O_3 is not surprising considering that this pollutant's dynamic depends on different factors than the rest. Column (2) and (3) reveal that inversions closer to the ground and those with a large temperature differential lead to stronger effects on PM_{10}. In particular, an extra intense inversion in the past week raises that pollutant by 0.4 units in the same time frame. Yet, when we look at other pollutants, it seems that the strength is coming from the closeness from the ground.

Effects of Thermal Inversions on Infant Health

To study the effects of exposure *in utero* to air pollution on health at birth, we consider a specification that follows our Model (2). Following Rangel and Vogl (2016), we use a distributed lag model on data aggregated at the location and date of birth level. Specifically, given that we cannot backdate conception from Brazilian birth records, we use lags for 38 weeks from the birthdate and assume 9-month pregnancies. For births in location l on date t, we estimate,

$$H_{lt} = \sum_{s=0}^{38} \beta_s Inv1km_{t-7s} + \sum_{s=0}^{38} X'_{l,t-7s}\ \gamma_s + \mu_{wl} + \mu_y + \epsilon_{lt} \tag{2}$$

where H_{lt} is an average birth outcome. The central explanatory variables are $Inv1km_{t-7s}$, the count of thermal inversion episodes close to the ground for the week leading up to date $t-7s$, where s is measured in weeks and t is measured in days. The vector of covariates is defined as in equation (1) for each one of the lagged weeks.

From equation (2) we recover many week-specific estimates, yet they might be individually imprecise and unwieldy to report (Almon, 1965; Sargen, 1980). Thus, we report 13-week sums of the coefficients that correspond to the last, second-to-last, and third-to-last periods

before birth – alike trimesters of gestation,

$$\beta_{\{\underline{T},\bar{T}\}} \equiv \sum_{s=\underline{T}}^{\bar{T}} \beta_s \tag{3}$$

where the periods $\{\underline{T},\bar{T}\}$ are $\{0,12\}$, $\{13,25\}$, and $\{26,38\}$. Each of these coefficients indicates the effect of having an extra thermal inversion per week during periods of approximately three months. Defining exposure backwards from the date of birth is not ideal: the timing and length of pregnancy are endogenous. For some premature infants, the 38-weeks period might include weeks before their conception. Moreover, those who are exposed to inversions since early in their time *in utero* might be positively selected because they were not born prematurely. Thus, estimates from the first three months of pregnancy should be interpreted carefully.

Results

Table 4 shows estimations of equation (2). We report full-sample estimates of the 13-week sums of coefficients for each of our birth outcomes. That is, they show the effect of an additional thermal inversion episode per week in each period of 13 weeks.

Columns (1), (2), and (3) present the estimates related to birth weight. In columns (1), the results indicate that an extra inversion per week occurring in the last 13 weeks of gestation leads to a reduction birth weight by 4.4 grams; an effect that is statistically significant at the 1% level. The impact on birth weight is actually felt at the lower tail of the birth weight distribution: an additional inversion per week in the last period of gestation increases the incidence of low birth weight (< 2,500 grams) by 2.4% – an effect of 2 per 1000 on a base risk of 82 per 1,000 –and raises the incidence of very low birth weight (< 1,500 grams) by 4.2% – an effect of 0.5 per 1000 on a base level of 12 per 1,000. For the two earlier gestational periods, we do not find statistically significant effects of inversions. Moreover, their differential effects might be related to concerns about selection form prematuriy.

Columns (4) and (5) show that the effect of inversions on birth weight might be explained in part by their influence on the length of gestation. Rates of preterm birth (<37 weeks) and very preterm births (< 32 weeks) increases by 1.7% and 10%, respectively – both are effects of 1.2 per 1,000 on a risk base of 71 and 12 per 1,000, respectively.

Overall, this table reports that exposure to thermal inversions during the last 13 weeks

of gestation have a negative effect on infants health at birth. Considering the evidence that health early in life is a predictor of future outcomes such as earnings and education (see Almond and Currie, 2011, for a review of the literature), our results suggest lasting negative consequences of air pollution on new generations.

Discussion and Next Steps

Urban air pollution is one of the most critical issues worldwide. Growth in urban transportation and congestion are key elements behind that. A growing number of studies focused on developed countries have shown that prenatal exposure to air pollution harm health at birth and leads to an increase of infant mortality. Central differences between developed and developing countries limit the understanding of the consequences of air pollution in the latter based on results from richer countries. Thus, the magnitude and scope of the impact of exposure to air pollution on infants health in less-developed countries still remain unclear.

In this paper, we use data from one of the largest urban conglomerate, the Metropolitan Area of São Paulo. Taking advantage of the meteorological phenomenon of thermal inversion which arguably exogenously locks pollutants closer to the ground, we find that exposure to inversion episodes during the last three months of gestation decreases birth weight and increases the changes of prematurity. These results are strongly significant.

A future version of this paper will expand in at least two ways. First, we will more directly link traffic to health outcomes by interacting thermal inversions with congestion data. Second, we will investigate urban policy and technological innovations during the period we study. In particular we examine the regulation of minimum ethanol content in gasoline and the development and adoption of flex-fuel cars (running 100% on ethanol).

References

Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*, 33(1): 178-196.

Almond, D. and Currie, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, 25(3): 153-72.

Arceo-Gomez, E., Hanna, R., and Oliva, P. (2016). Does the Effect of Pollution on Infant Mortality Differ between Developing and Developed Countries? Evidence from Mexico City. *Economic Journal*, 126: 257-280.

Carvalho, V., Freitas, E., Martins, L., Martins, J., Mazzoli, C. and Andrade, M. (2015). Air quality status and trends over the Metropolitan Area of Sa?o Paulo, Brazil as a result of emission control policies. *Environmental Science & Policy*, 47: 68-79.

Companhia de Tecnologia de Saneamento Ambiental (CETESB) (2002). Relatòrio de qualidade do ar no Estado de São Paulo 2001.

Companhia de Tecnologia de Saneamento Ambiental (CETESB) (2010). Qualidade do ar no Estado de São Paulo 2009. Sèrie Relatòrios.

Chay, K., and Greenstone, M. (2003). The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *Quarterly Journal of Economics*, 118(3): 1121-1167.

Currie, J., Neidell, M., and Schmieder, J. (2009). Air Pollution and Infant Health: Lessons from New Jersey. *Journal of health Economics*, 28(3): 688-703.

Currie, J. and Walker, R. (2011). Traffic Congestion and Infant Health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1): 65-90.

Jacobson, M. (2005). Fundamental of Atmospheric Modeling. Cambridge: Cambridge University Press.

Jans, J., Johansson, P., and Nilsson P. (2014). Economic status, air quality, and child health: evidence from inversion episodes. IZA Working Paper 7929.

Kinney, P., Gatari, M., Volavka-Close, N., et al. (2011). Traffic impacts on PM2.5 air quality in Nairobi, Kenya. *Environmental Science & Policy*, 14: 369-378.

Knittel, C., Miller, D., and Snaders, N. (2016). Caution, drivers! Children present: traffic, pollution, and infant health. *The Review of Economics and Statistics*, 98(2): 350-366.

Instituto Brasileiro de Geografia e Estatstica (IBGE). (2011). Estatsticas do Registro Civil (Vol. 37). Rio de Janeiro, Brazil: Instituto Brasileiro de Geografia e Estatstica.

Rangel, M. and Vogl, T. (2016). Agricultural fires and infant health. NBER, Working

Paper 22955.

Redon, A., Salazar, J., and Palacio, C. (2014). Effects of Urbanization on the Temperature Inversion Breakup in a Mountain Valley with Implications for Air Quality. *Journal of Applied Meteorology and Climatology*, 53: 840-858.

Ribeiro, H., and Assuncao, J. (2001). Historical overview of air pollution in Sao Paulo metropolitan area, Brazil: influence of mobile sources and related health effects. *Transactions on the Built Environment*, 52: 351-360.

Rolnik, R., and Klintowitz, D. (2011). (Im) Mobility in the city of São Paulo. *Estudos Avançados*, 25(71): 89-108.

Sanders, N. (2012). What Doesn't Kill You Makes You Weaker: Prenatal Pollution Exposure and Educational Outcomes, *Journal of Human Resources*, 47: 826-850.

Sargan, J. (1980). The Consumer Price Equation in the Post War British Economy: An Exercise in Equation Specification Testing. *Review of Economic Studies*, 47(1): 113-135.

Figure 1: Relationships between pollutants and thermal inversions

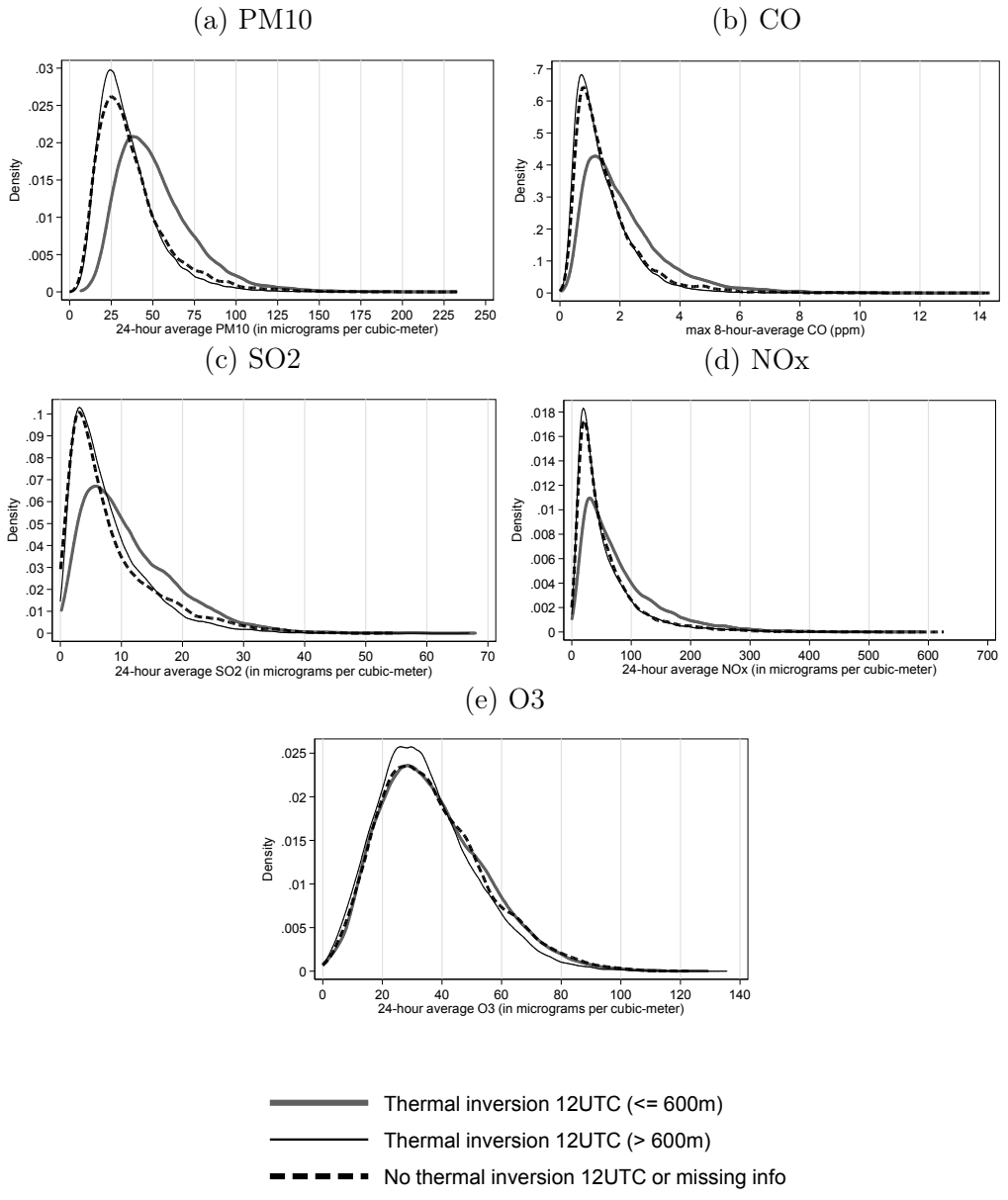
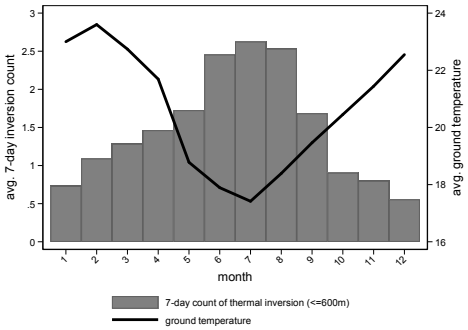
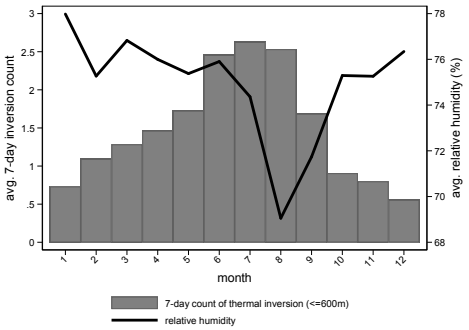


Figure 2: Average of the number of thermal inversions in a week by calendar month

(a) Ground temperature



(b) Relative humidity



(c) Rainfall

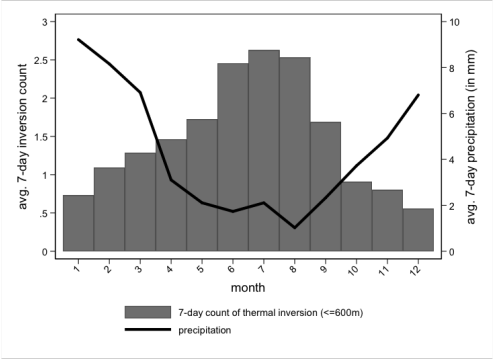


Table 1: Descriptive Statistics: pollution, weather conditions, and thermal inversions

	Mean	SD	Observations	Stations	Dates
A. Pollutants					
Particulate Matter (PM_{10}) ($\mu g/m^3$)	42.2	17.3	57,162	25	2286
Carbon Monoxide (CO) (*ppm*)	1.3	0.6	34,416	16	2,151
Ozone (O_3) ($\mu g/m^3$)	33.5	12.6	34,359	15	2,291
Nitrogen Oxides (NO_x) (*ppb*)	79.6	57.6	21,394	12	1,783
Sulphur Dioxide (SO_2) ($\mu g/m^3$)	11.4	5.9	14,421	8	1,803
B. Weather variables					
Mean Temperature (*Celsius*)	20.3	2.7	57,162	25	2,286
Max Temperature (*Celsius*)	25.8	3.2	57,162	25	2,286
Min Temperature (*Celsius*)	16.5	2.7	57,162	25	2,286
Mean Relative Humidity (*percentage*)	76.3	8.3	57,162	25	2,286
Max Relative Humidity (*percentage*)	92.4	6.2	57,162	25	2,286
Min Relative Humidity (*percentage*)	52.5	11.3	57,162	25	2,286
Rainfall (*millimetre*)	4.1	4.7	57162	25	2286
Percent of winds originating from octant...					
Wind NNE	9.8	7.2	57,162	25	2,286
Wind ENE	13.4	7.6	57,162	25	2,286
Wind ESE	15.6	10.7	57,162	25	2,286
Wind SSE	28.6	14.3	57,162	25	2,286
Wind SSW	7.2	6.0	57,162	25	2,286
Wind WSW	2.4	2.2	57,162	25	2,286
Wind WNW	7.5	7.1	57,162	25	2,286
Wind NNW	10.2	8.2	57,162	25	2,286
Missing or Without prevailing wind	5.2	10.0	57,162	25	2,286
C. Thermal inversions					
At least 1 inversion (*percentage*)	71	45	3,197	1	3,197
Number of inversions	4.6	3.1	3197	1	3,197
At least 1 inversion below 1.3 km (*percentage*)	66	47	3,197	1	3197
Number of inversions below 1.3 km	2.2	2.1	3,197	1	3,197
Number of inversions High Δ Temp	1.2	1.6	3,197	1	3,197
At least 1 date with missing data (*percentage*)	29	45	3,197	1	3,197
Number of missing data dates	0.5	1.1	3,197	1	3,197

Notes: Observations are at station-date level; they are measures over the preceding week.

Table 2: Descriptive Statistics: health outcomes

	Mean	SD	Min	Max	Observations	Weights
Panel A. Health outcomes						
Birth weight	3152.6	75.9	1085	4560	326,661	16,477,732
Low birth weight (*per 1,000*)	82.1	38.8	0.0	1000	326,661	16,477,732
Very low birth weight (*per 1,000*)	12.1	15.4	0.0	1000	326,661	16,477,732
Preterm (*per 1,000*)	71.8	38.7	0.0	1000	326,661	16,477,732
Very preterm (*per 1,000*)	11.6	15.1	0.0	1000	326,661	16,477,732
Panel B. Covariates						
Mean Temperature (*Celsius*)	20.4	2.6	12.2	27.7	326,661	16,477,732
Max Temperature (*Celsius*)	25.8	3.1	15.1	34.5	326,661	16,477,732
Min Temperature (*Celsius*)	16.6	2.7	7.7	23.1	326,661	16,477,732
Mean Relative Humidity (*percentage*)	76.0	7.1	50.8	99.2	326,661	16,477,732
Rainfall (*millimetre*)	4.1	4.4	0.0	36.1	326,661	16,477,732
Percent of winds originating from octant...						
Wind missing	5.6	7.3	0.0	68.3	326,661	16,477,732
Wind NNE	9.8	5.6	0.0	46.4	326,661	16,477,732
Wind ENE	12.8	5.8	0.0	51.0	326,661	16,477,732
Wind ESE	15.7	7.2	0.1	66.6	326,661	16,477,732
Wind SSW	6.6	4.0	0.0	35.7	326,661	16,477,732
Wind WSW	2.4	1.7	0.0	21.8	326,661	16,477,732
Wind WNW	8.3	6.1	0.0	64.9	326,661	16,477,732
Wind NNW	10.2	7.6	0.0	48.3	326,661	16,477,732
Number of inversions below 1.3 km	2.3	2.1	0.0	7.0	326,661	16,477,732
At least 1 date with missing data	27.7	44.8	0.0	100.0	326,661	16,477,732
Number of missing data dates	0.5	1.2	0.0	7.0	326,661	16,477,732

Notes: Observations are at location-date level; they are measures over the preceding week.

Table 3: Effect of Thermal Inversions on Pollution

	(1)	(2)	(3)
Panel A: PM_{10}			
Inversions (all)	0.231		
	(0.152)		
Inversions 1km		0.446***	
		(0.108)	
Inversions High ΔTemp			0.404***
			(0.117)
Panel B: CO			
Inversions (all)	5.359		
	(5.031)		
Inversions 1km		11.037***	
		(3.622)	
Inversions High ΔTemp			3.917
Panel C: O_3			
Inversions (all)	-0.480*		
	(0.233)		
Inversions 1km		0.011	
		(0.083)	
Inversions High ΔTemp			0.084
			(0.072)
Panel D: NO_x			
Inversions (all)	1.317*		
	(0.605)		
Inversions 1km		0.506	
		(0.343)	
Inversions High ΔTemp			0.512
			(0.436)
Panel F: SO_2			
Inversions (all)	0.269***		
	(0.070)		
Inversions 1km		0.123**	
		(0.048)	
Inversions High ΔTemp			-0.006
			(0.042)

Notes: Observations are at station-date level; they are measures over the preceding week. The sample includes 3,197 dates from April 2001 to December 2009. Each of the columns (1) to (3) presents the results of a model. *Inversions 1km* indicates the number of inversions that occurred up to 1,331 meters from the sea level. *inversions High ΔTemp* stands for the number of episodes which difference in temperature at the base and top of the inversion that is above the percentile 75 of the distribution of all inversions. All regressions control for an indicator of having missing information data over the preceding week, the number of dates with missing data in that week, and weather controls – cubic polynomials in mean temperature, maximum temperature, minimum temperature, humidity, direction of wind (fixed octants), and rainfall. In all specifications we include year fixed effects, week of the year x station fixed effects. Two-way clustering is employed; standard errors are clustered at station and date level. *** significant at 1% level, ** significant at 5% level, * significant at 10% level.

Panel A: 57,162 observations measured in 25 stations. Mean dependent variable: 42.2. SD dependent variable: 17.3. *Panel B:* 34,416 observations measured in 16 stations. Mean dependent variable: 1260.4; SD dependent variable: 589.3. *Panel C:* 34,359 observations measured in 15 stations. Mean dependent variable: 33.5; SD dependent variable: 12.6. *Panel D:* 21,394 observations measured in 12 stations. Mean dependent variable: 79.6; SD dependent variable: 57.6. *Panel E:* 14,421 observations measured in 8 stations. Mean dependent variable: 11.4; SD dependent variable: 5.9.

Table 4: Effect of Thermal Inversions on Infant Health

	(1)	(2)	(3)	(4)	(5)
	Birth weight	Low birth weight (per 1,000)	Very low birth weight (per 1,000)	Preterm (per 1,000)	Very premature (per 1,000)
Inversions-1km weeks t and t-12	-4.3993***	1.8792***	0.5496**	1.2386*	1.1609***
	(1.2817)	(0.6429)	(0.2687)	(0.7103)	(0.3759)
Inversions-1km weeks t-13 and t-25	0.9046	0.0214	-0.3538	0.9310	-0.1239
	(1.2095)	(0.7433)	(0.2563)	(0.8697)	(0.2633)
Inversions-1km weeks t-26 and t-38	-0.6389	0.0864	0.0092	0.8881	-0.2392
	(1.2897)	(0.6992)	(0.3197)	(0.9697)	(0.3419)
Observations	316,970	316,970	316,970	316,970	316,970
Ymean	3,152.6	82.1	12.1	71.8	11.6

Notes: Observations are at location-date level and are weighted by the size of the local birth-cohort in that location-day. Estimates are 13-week sums of coefficients on weekly counts of thermal inversions that occurred up to 1,331 meters from the sea level. All regressions control for an indicator of having missing information data over the preceding week, the number of dates with missing data in that week, and weather controls – cubic polynomials in mean temperature, maximum temperature, minimum temperature, humidity, direction of wind (fixed octants), and rainfall. In all specifications we include year fixed effects, week of the year x location fixed effects. Two-way clustering is employed; standard errors are clustered at location and date level. *** significant at 1% level, ** significant at 5% level, * significant at 10% level.