Clearing the Air on the Benefits and Costs of Road Infrastructure*

Clare Balboni
LSE
c.a.balboni@lse.ac.uk

Aaron Berman
MIT
bermana@mit.edu

Johannes Boehm
Geneva Graduate Institute
johannes.boehm@graduateinstitute.ch

Lorenzo Marzano
UPF
lorenzo.marzano@bse.eu

Mazhar Waseem
University of Manchester
mazhar.waseem@manchester.ac.uk

June 13, 2025

First draft, preliminary

Abstract

Investments in transportation infrastructure can yield sizable trade and commuting cost gains, but may also contribute to damaging local air pollution from increased vehicle traffic. Accounting for such costs is especially important in developing country cities, which are making significant investments in expanding road networks but where high ambient pollution concentrations contribute to severe health impacts. We develop a quantitative urban equilibrium model, integrating a novel granular atmospheric model of pollution dispersion, in order to estimate the impacts of investing in urban roads accounting for both gains from economic integration and costs from local pollution exposure. We estimate the model using high-resolution data on emissions, pollution transport, commuting and goods trade in Lahore, Pakistan – one of the world's most polluted cities. Counterfactual simulations consider the aggregate and distributional implications of major road infrastructure projects in the city.

KEYWORDS: transportation, road infrastructure, pollution, commuting, trade
JEL: F18, Q53, Q56, R13, R40

*We thank Omar Masoud and the Punjab Urban Unit for sharing data. We are grateful to Somei Miyashita for outstanding research assistance. This project is an output from the research initiative "Private Enterprise Development in Low-Income Countries" (PEDL), a programme funded jointly by the Centre for Economic Policy Research (CEPR) and the Foreign, Commonwealth and Development Office (FCDO), contract reference PEDL_LOA_10011_Boehm. The views expressed are not necessarily those of CEPR or FCDO.

1 Introduction

Expanding transportation infrastructure is widely regarded as a catalyst for economic development, facilitating market integration and labor mobility across and within cities (Redding and Turner, 2015; Heblich, Redding and Sturm, 2020). Such investments, however, may also have important environmental consequences. Air pollution from road traffic is a textbook example of a local externality, which can impose significant costs on local residents by increasing their exposure to harmful pollutants (Currie and Walker, 2011; Bassi et al., 2022; Knittel, Miller and Sanders, 2016). These costs may be especially pronounced in rapidly urbanizing regions of the developing world, which are often characterized by high traffic congestion and weak environmental regulation, and where infrastructure development is often pursued with limited consideration of environmental consequences (Kahn, 2007; Bryan, Glaeser and Tsivanidis, 2020; Oguge et al., 2024).

In this paper, we evaluate the aggregate and distributional consequences of urban road infrastructure investments while explicitly accounting for the local pollution externalities that they impose. We develop a quantitative spatial model that captures the trade and commuting cost gains from improving urban transportation infrastructure, and combine this economic model with a detailed atmospheric model of pollution dispersion to assess the net welfare effects of road infrastructure in Lahore, Pakistan. Lahore is one of the most polluted cities in the world: the average resident loses 5.3 years of life expectancy from particulate matter pollution (Greenstone and Fan, 2019), to which vehicular emissions are a principal contributor (Shabbir, 2024). In this context, we assemble a combination of high-resolution datasets capturing commuting, goods trade, production output, and pollution in order to estimate how local emissions may alter assessments of the welfare gains from road improvements.

We present a series of motivating empirical results leveraging high-frequency pollution data from a network of air quality monitors across Lahore. Pollution concentrations are very high on average, and demonstrate pronounced spikes during the morning and evening rush hours, consistent with a significant role for traffic in generating urban air pollution. These impacts are extremely localized, highlighting the importance of accounting for pollution at a fine spatial resolution in order to capture spatially heterogeneous impacts of traffic emissions. In line with findings from the wider literature (Balboni and Shapiro, 2025), we find that the extent and spatial distribution of pollution depend on localized meteorological factors such as wind speed and direction, cloud cover, and temperature. These complex environmental determinants of pollutant dispersion underscore the need to augment economic models of changes in transportation patterns with realistic physical pollution transport mechanisms to capture the magnitude and incidence of the external costs of road infrastructure.

To this end, we develop a quantitative spatial equilibrium model of Lahore that integrates commuting and trade frameworks with a detailed pollution dispersion model. The economic core of the model, building on Allen and Arkolakis (2022), features endogenous workplace choice, trade of goods and commuters across locations, and transport route selection, allowing us to capture the

benefits of improved connectivity conferred by transportation infrastructure investments. Standard agglomeration externalities increase productivity, while congestion externalities raise travel costs as traffic intensifies (Kreindler, 2024; Almagro et al., 2024). We assume that local pollution concentrations at individuals' location of residence and workplace reduce their utility and productivity, respectively, in line with a growing empirical literature (Adhvaryu, Kala and Nyshadham, 2022; Hoffmann and Rud, 2024; Chay and Greenstone, 2005; Freeman et al., 2019).

The central innovation of our framework is the incorporation of a pollution emission and dispersion model based on the ADMS-Urban system (Carruthers et al., 1998; Hood et al., 2018), which explicitly accounts for the spatial distribution of air pollution as a function of emissions from production and transportation, and environmental conditions. Our analysis focuses on key local pollutants with established links to human health and economic outcomes, in particular concentrations of particulate matter with particle diameters of less than 2.5 or 10 micrometers ($PM_{2.5}$ and PM_{10}), as well as nitrogen oxides (NO_x).[1] We calibrate the pollution dispersion model to the setting of urban Lahore at a fine spatial resolution using detailed input data on local meteorological conditions, background chemical concentrations, traffic counts and vehicle fleet composition, emissions factors and sources, and validate its predictions using data from 114 ground-level pollution sensors.

The quantitative implementation of the economic model leverages data on firm sales transactions from VAT records, traffic surveys, and fine-grained signals from cellphones (to capture commuting patterns) and from commercial trucks (to capture goods trade). The model predicts that welfare changes can be estimated by observing how trade and the location of economic activity and traffic flows change in response to exogenous changes in transportation costs that result from road investments (Redding and Rossi-Hansberg, 2017). Our integrated environmental-spatial framework preserves the ability to solve the model in changes (the "exact-hat algebra" introduced by Dekle, Eaton and Kortum, 2008), which we use to perform counterfactual analyses of road infrastructure policies.

Using this quantitative framework, we conduct counterfactual exercises to estimate the benefits and costs of major road infrastructure projects in Lahore, accounting for both gains from market integration and damages from local air pollution. We evaluate the spatial distribution of economic and pollution-related impacts resulting from traffic restrictions along the Lahore ring road, a major orbital highway that was opened in three phases around the central city from 2012 to 2024, and carries over 400,000 vehicles daily.

Our analysis quantifies the heterogeneous welfare effects of changes in trade, commuting, congestion and air quality across different neighborhoods of the city that result from removal of the ring road. The aggregate estimated welfare loss from removing the ring road is 3.80%, with especially pronounced losses in areas of the city directly surrounding the ring road route where market access declines markedly. With the closure of the ring road, commuter traffic is estimated to substitute

[1] This contrasts with studies of how trade may influence emissions of global pollutants such as carbon dioxide, which have focused on international transportation of goods (Cristea et al., 2013; Shapiro, 2016).

away from the direct vicinity of the ring road towards neighboring areas in the east and north of the city. Consistent with changes in the location of commuter traffic, the closure of the ring road also results in a shift in the spatial distribution of elevated air pollution concentrations towards the east and north of Lahore.

A second set of simulations estimates how far accounting for this shift in the location of air pollution affects the estimated aggregate and distributional impacts of the ring road. Shutting down the effects of air pollution in the model suggests that welfare losses from removal of the ring road would instead be 4.30%, making clear that ignoring the impacts of local air pollution would result in a quantitatively sizable over-estimate of the welfare gains from the opening of the ring road. Accounting for pollution also leads to significant changes in the distributional impacts of the ring road: for 15% of spatial units, accounting for local air pollution leads to a different conclusion about whether construction of the ring road leads to an increase or decrease in overall welfare.

Finally, we simulate the net welfare impact of restricting traffic on individual road segments throughout the city. This exercise identifies 110 out of 623 road edges in Lahore, concentrated in the city outskirts, for which closure would yield positive net gains once the local pollution externalities associated with vehicle traffic are accounted for.

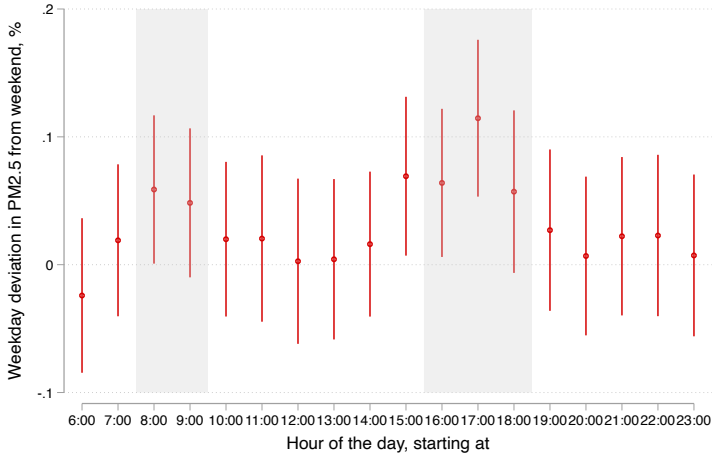
The rest of the paper is structured as follows. Section 2 presents motivating facts relating to vehicle emissions and local pollutant concentrations in Lahore, Pakistan. Section 3 presents the theoretical framework that we use to study the aggregate and distributional implications of transportation infrastructure investments. Section 4 describes the data from Lahore used in our analysis. Section 5 discusses how we use these datasets to estimate our theoretical model. Section 6 presents the results of the counterfactual simulations that we use to assess the benefits and costs of road investments in Lahore. Section 7 concludes.

2 Vehicle emissions and local pollutant concentrations in Lahore

The focus of our analysis is on understanding how far accounting for local air pollution externalities may alter estimates of the aggregate and distributional impacts of road infrastructure projects. A key motivation for this focus is that local air pollutants are a first-order health concern in our empirical setting, and that vehicle emissions are a significant— though spatially complex—driver of these pollutants. In 2016, 98% of Pakistan's population lived in areas where the annual average particulate pollution level exceeded World Health Organization guidelines of 10µg/m^3, and 97% in areas where it exceeded Pakistan's own $PM_{2.5}$ standard of 15µg/m^3, resulting in the loss of an estimated 5.3 years of life expectancy for an average Lahore resident (Greenstone and Fan, 2019).

Figure 1 suggests that road traffic is a notable contributor to particulate matter concentrations in Lahore. The estimates in the figure show residual $PM_{2.5}$ concentrations measured at 13 ground-level monitoring stations across the city between 2018 and 2023 by hour of the working day, controlling for natural diurnal variation using hourly concentrations during weekends (Saturday and Sunday), as well as sensor and day fixed effects, and weather conditions. Clear peaks with

Figure 1 $PM_{2.5}$ concentrations and rush hour periods in Lahore



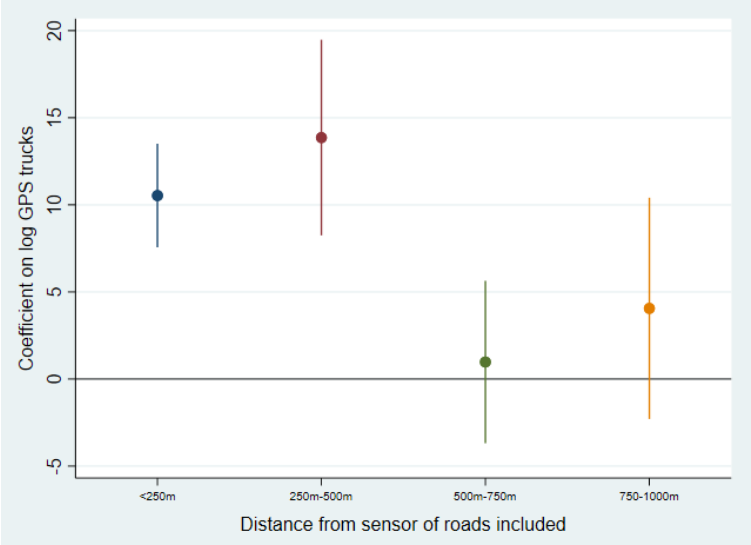
The graph shows the average weekday (Monday to Thursday) deviations in $PM_{2.5}$ concentrations, in percent, from the unconditional averages, by hour of the day. Shaded regions correspond to the morning and evening rush hours. The underlying regression excludes Fridays (because or less regular work hours due to Friday prayers), and includes sensor, day, and hour-of-the-day fixed effects as well as controls for third-order polynomials in temperature, humidity, wind speed, precipitation, and cloud cover. The regression weighs observations by the inverse of the number of observations available for each sensor, to correct skewness in data availability.

especially elevated pollutant concentrations are visible during the morning and evening rush hour periods (shaded in grey). Concentrations are about 5% higher during the morning rush hour, and 10% higher during the evening rush hour, compared to their unconditional averages during those hours of the day, which are already extremely high at approximately 100µg/m^3 on average during the morning rush hour and 65µg/m^3 during the evening rush hour. These are periods when vehicular traffic is at its highest, but do not coincide with peak emissions from other major sources of local pollutants such as industry and crop burning.

Average pollutant concentrations across Lahore will be a poor guide to local conditions and damages, which are highly localized. While traffic patterns have a clear impact on local pollutant concentrations on average across monitor-day observations, areas that are not adjacent to major traffic thoroughfares are likely to be largely shielded from these impacts. Figure 2 makes clear how localized the impacts of traffic on local pollution are. The figure plots the coefficient on traffic flows in a regression of $PM_{2.5}$ concentrations on local traffic and weather conditions, as the radius within which traffic flows are included increases. The results highlight that only roads within 500 meters of pollution sensors have a significant impact on local pollutant readings. This difference underscores the importance of accounting for local air pollution damages at a fine spatial resolution in estimating heterogeneous welfare impacts. Furthermore, this suggests that analysis relying solely on measured pollutant concentrations will be very sensitive to the precise location of air quality sensors, especially given the sparsity of air pollution monitors across Lahore (see Appendix Figure B.1).

Our counterfactual simulations depend crucially on how changes in traffic patterns influence local pollutant concentrations across the city. Atmospheric processes complicate the relationship

Figure 2 Local traffic contributions to recorded $PM_{2.5}$ concentrations



Each point represents the estimated coefficient on Log GPS trucks from regressions using specification (5) in Table 1, where the distance within which roads are included in the traffic variable changes across regressions (eg. 250-500m only includes GPS signals from roads at a distance between 250m and 500m from each sensor).

between the locations where such pollutants are emitted and areas where concentrations are highest. Table 1 presents correlations between measured $PM_{2.5}$ concentrations and local conditions including traffic, temperature and humidity at the monitor level, and additional weather and wind conditions at a resolution of 0.25° x 0.25°. While traffic can again be seen to be an important driver of local particulate matter concentrations, the latter are also positively correlated with cloud cover and negatively with wind speed, temperature, precipitation and the ventilation coefficient.[2] Such relationships, and their interactions, will be important in determining how changes in traffic, induced by counterfactual road investments, alter pollutant concentrations across different regions of the city. We use the ADMS-Urban model, calibrated to local conditions in Lahore as described in Section 4, to capture the detailed atmospheric transport and chemistry processes via which local pollutant emissions at a given point drive ambient concentrations across all other locations in the city.

[2]The ventilation coefficient is a key meteorological parameter that captures the ability of the atmosphere to disperse pollutants and is defined as the boundary layer (mixing) height multiplied by the wind speed.

Table 1 Correlates of $PM_{2.5}$ concentrations

	Dependent variable: PM2.5 concentration (ug/m3)				
	(1)	(2)	(3)	(4)	(5)
Weekend dummy	-17.93***	-22.15***	-19.90***	-24.59***	-22.65***
	(2.21)	(2.07)	(2.98)	(2.88)	(2.74)
Ramadan dummy	-21.38***	-20.07***	-49.55***	-13.46	-47.25***
	(5.57)	(5.70)	(9.52)	(11.2)	(9.71)
Log GPS trucks			7.479***	7.916***	9.438***
			(2.01)	(1.94)	(1.90)
Wind speed (std)		-12.80		-13.68***	-37.15**
		(10.9)		(1.57)	(11.9)
Temperature (std)		-45.99***		-57.92***	-50.57***
		(5.51)		(3.90)	(5.99)
Humidity (std)		-38.70***		16.90***	-31.66***
		(4.31)		(2.00)	(6.77)
Cloud cover (std)		34.91*		7.291***	-2.016
		(17.0)		(1.52)	(20.7)
Precipitation (std)		-0.311		-2.194*	-1.291
		(3.39)		(1.11)	(3.98)
Inv Ventilation coeff (std)		16.50***		0.314	24.27***
		(4.78)		(1.78)	(6.11)
Sensor FE	X	X	X	X	X
Hour of the day FE	X	X	X	X	X
Week FE	X	X	X	X	X
Additional controls		X			X
R^2	0.595	0.647	0.599	0.627	0.658
Observations	112329	103693	34380	31727	31727

Standard errors in parentheses. GPS trucks denote the number of tracked vehicles passing on roads within 500 metres on a given day. Temperature and humidity are taken from the sensor data (hourly average when missing), whereas other meteorological variable are linearly interpolated from ERA5 reanalysis data. Inv ventilation coefficient denotes the inverse of boundary layer height times wind speed. All control variables are standardised. Additional control variables include dummy variable for the wind direction quadrant and the squares and cube of control variables. Observations from each sensor is weighted by the inverse of the number of total observations in the dataset.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

3 Model

We construct a model of Lahore which captures workers that commute between locations of residence and work, and firms which produce and trade goods varieties. Workers' choice of workplace location and commuting routes are influenced by emissions, which are endogenously generated by production, trade and commuting, and transported by exogenous atmospheric processes. The model builds on canonical models of commuting (Allen and Arkolakis (2022); Monte, Redding and Rossi-Hansberg (2018)), but integrates spatial pollution externalities that are consistent with pollution dispersion models (Jensen et al. (2017); Hood et al. (2018); Hall et al. (2024)).

3.1 Workers and firms

The geography consists of a graph G with vertices ("locations") I that are connected by edges ("roads") E. We take the distribution of residential locations as exogenous, and assume that in each location i there is a mass of L_i^0 of ex-ante identical workers.[3] Workers choose a location in which to work and a route from their location of residence to their work location. Denote by $\mathfrak{R}_{ji}$ the set of all routes (i.e. *walks* of finite length) on G from a location i to a location j.

Workers in location j supply one unit of labor in that location. Pollution in work locations reduces effective labor supply, so that each worker in location j supplies $m_j^{-\sigma_w}$ effective units of labor, where m_j is the pollutant concentration in j and σ_w is the elasticity of workers' productivity with respect to concentration. This is consistent with widespread evidence of the detrimental impacts of air pollution for worker productivity (Holub and Thies (2023); Hoffmann and Rud (2024); Sarmiento (2022)), including in settings where particulate pollution levels are comparable to those observed in our empirical setting (Adhvaryu, Kala and Nyshadham (2022)).[4]

Labor markets are perfectly competitive within each location, so that workers earn a labor income of $m_j^{-\sigma_w} w_j$. Workers are assumed to purchase consumption goods in their workplace location.[5] They derive linear utility from consuming a bundle c_j of varieties on a unit interval, $c_j(s)$, with associated prices $p_j(s)$ and ideal bundle price index p_j:

$$c_j = \left[\int_0^1 c_j(s)^{\frac{\varepsilon-1}{\varepsilon}} ds\right]^{\frac{\varepsilon}{\varepsilon-1}}, \qquad p_j = \left[\int_0^1 p_j(s)^{1-\varepsilon} ds\right]^{\frac{1}{1-\varepsilon}}.$$

In addition to consumption, utility is also shaped by commuting costs, a utility shock $\eta_{ji,r}$ that is specific to commuting route r, and pollution at the residential location. We assume that these affect utility multiplicatively: formally, when a worker ω in location i chooses to commute to work in j via route $r = (r_1, r_2, \ldots, r_K(r))$, her utility is

$$V_{ji,r}(\omega) = \underbrace{m_i^{-\sigma_h}}_{\text{Disutility from home pollution}} \times \underbrace{\frac{m_j^{-\sigma_w} \times w_j}{p_j}}_{\text{Real consumption}} \times \underbrace{\left[\prod_{l=1}^{K(l)} t_{r_{l-1},r_l}\right]^{-1}}_{\text{Commuting cost}} \times \underbrace{\eta_{ji,r}(\omega)}_{\text{Idiosyncratic shock}}. \quad (1)$$

The parameter σ_h denotes the elasticity of household utility with respect to pollution concentration

[3] Holding residential location fixed allows us to study the change in welfare for residents of each location. While there is evidence that local air pollution may influence residential sorting over decades or centuries (e.g. Heblich, Trew and Zylberberg, 2021), our estimation is over much shorter time horizons where such effects are unlikely to be first order.

[4] We abstract from the possibility that exposure to pollution at the home location may also influence labor supply or productivity at work. We are not aware of studies examining such lagged effects, which is often challenging given the difficulty of distinguishing home and work locations in the data. Studies in other contexts have found that a large share of employees live and work nearby, such that pollution exposure distributions at workplace and residence locations are nearly overlapping (Leroutier and Ollivier (2022)). This is also likely to be true in our setting, where 37% of observed individual residence-workplace pairs are located in the same mauza.

[5] This assumption implies that trade of each location is balanced, which simplifies the setup. The assumption can be relaxed, but much richer data on income flows would be required to take this generalized model to the data.

in the location of residence. Because the residential location is inelastic, this channel will not change workers' decisions, but we include it to be able to decompose the total impact of policy changes on welfare through different channels. Commuting costs t_e on each edge e depend on traffic. We assume that the idiosyncratic route-specific shock $\eta_{ji,r}$ is i.i.d Fréchet with

$$P\left(\eta_{ji,r} < x\right) = \exp\left(-x^{-\theta}\right).$$

Goods varieties are produced by perfectly competitive firms in each location i, using a technology that is linear in labor with unit productivity, $y_i(s) = l_i(s)$. Shipping the good to a destination j along a route $r \in \mathfrak{R}_{ji}$ incurs a deterministic iceberg cost $\prod_{k=1}^{K(r)} t_{r_{k-1},r_k}$ and a stochastic iceberg trade cost $\varepsilon_{ij}(r,s)^{-1}$ that satisfies

$$P(\varepsilon_{ij}(r,s) < x) = \exp\left(-A_i^{\theta} x^{-\theta}\right).$$

As a result, the cost (and therefore price) of offering one unit of variety s from location i in location j via route r is

$$p_{ji,r}(s) = w_i \frac{\prod_{k=1}^{K(r)} t_{r_{k-1},r_k}}{\varepsilon_{ji}(r,s)}.$$

We denote aggregate production in location i by $Y_i = \int_0^1 y_i(s)ds$ and aggregate labor use by $L_i = \int_0^1 l_i(s)ds$. Labor supply in each location is determined by commuting decisions and pollution concentration levels according to:

$$L_j = m_j^{-\sigma_w} \sum_i \omega_{ji} L_i^0 \tag{2}$$

where ω_{ji} is the share of households in location i that commute to location j. Under the assumption that the expected trade costs between locations remain positive despite the many route-specific cost draws[6], the commuting shares and sourcing shares follow the familiar gravity structure:

Lemma 1 *Assume that*

$$\tau_{ji} \equiv \left[\sum_{r \in \mathfrak{R}_{ji}} \left(\prod_{l=1}^{K(r)} t_{r_{l-1},r_l}^{-\theta}\right)\right]^{-\frac{1}{\theta}}$$

remains bounded away from zero. Then the probability that a household from i decides to work in location j is

$$\omega_{ji} = \frac{m_j^{-\theta\sigma_w} w_j^{\theta} p_j^{-\theta} \tau_{ji}^{-\theta}}{\sum_{j'} m_{j'}^{-\theta\sigma_w} w_{j'}^{\theta} p_{j'}^{-\theta} \tau_{j'i}^{-\theta}} \tag{3}$$

and the probability of sourcing a variety in j from i, as well as the share of varieties that j sources

[6]This is equivalent to assuming that the matrix with (j,i) element $t_{ji}^{-\theta}$ has a spectral radius of less than one. A necessary condition for this being the case is that the cost of remaining stationary is positive: $t_{ii} > 1$.

from i, and the expenditure share of j on goods from i, is

$$\pi_{ji} = \frac{(w_i/A_i)^{-\theta}\,\tau_{ji}^{-\theta}}{\sum_{i'}(w_{i'}/A_{i'})^{-\theta}\,\tau_{ji'}^{-\theta}}. \tag{4}$$

The model builds on the tractability of the route choice framework in Allen and Arkolakis (2022), which yields analytic expressions for traffic as a function of trade costs on each edge, and market access terms for nodes connected by each edge, as follows. First note that the consumer's price index can be written as

$$p_j = \left(\sum_k (w_k/A_k)^{-\theta}\,\tau_{jk}^{-\theta}\right)^{-\frac{1}{\theta}} \tag{5}$$

and denote the (inverse) producer market access by Π_i,

$$\Pi_i \equiv \left(\sum_j \tau_{ji}^{-\theta} p_j^{\theta} Y_j\right)^{-\frac{1}{\theta}} \tag{6}$$

Similar terms denote the (inverse) labor market access of households residing in location k, Ψ_k, and the (inverse) access of firms in location l to workers, denoted by Θ_l,

$$\Psi_k \equiv \left(\sum_i m_i^{-\theta\sigma_w}\tau_{ik}^{-\theta} w_i^{\theta} p_i^{-\theta}\right)^{-\frac{1}{\theta}} \qquad \Theta_l \equiv \left(\sum_j \tau_{lj}^{-\theta}\Psi_j^{\theta} L_j^0\right)^{-\frac{1}{\theta}} \tag{7}$$

The following lemma describes analytic expressions for goods and commuter traffic as a function of these market access terms and transportation costs:

Lemma 2 *The value of goods flows Ξ_e^y from location l to a location k connected via an edge e is given by*

$$\Xi_e^y = t_{kl}^{-\theta} \times p_l^{-\theta} \times \Pi_k^{-\theta} \tag{8}$$

and the mass of commuter flows Ξ_e^l on the same edge is

$$\Xi_e^l = t_{kl}^{-\theta} \times \Theta_l^{-\theta} \times \Psi_k^{-\theta} \tag{9}$$

3.2 Externalities

The model features three types of externalities. First, we model positive agglomeration externalities by assuming that the scale parameter of Fréchet productivity draws is increasing in local labor supplied with a constant elasticity $\alpha \geq 0$:

$$A_i = \bar{A}_i L_i^{\alpha}. \tag{10}$$

Second, congestion externalities arise because the deterministic cost of crossing an edge e in-

creases with the amount of traffic on the edge, Ξ_e, governed by an elasticity λ. We weight average goods traffic relative to commuter traffic using a fixed factor ν, following the approach taken in the transportation planning literature, which calculates the *passenger-car equivalents* of different road vehicles. The expression for transportation costs on an edge therefore becomes

$$t_e = \bar{t}_e \left(\nu \Xi_e^y + \Xi_e^l \right)^\lambda . \tag{11}$$

Finally, a key focus of our analysis is environmental externalities from air pollution. We model concentrations of air pollutants in each location j as an additive function of particles transported from other locations. A *pollution transportation matrix* $\Gamma = (\gamma_{ji})_{j,i}$ governs the steady state concentration increase γ_{ji} in location j when one unit of particles is emitted in location i. Emissions derive from production at a rate of μ_y units per unit of output, and from goods and commuter traffic at rates of μ_{Ξ_y} and μ_{Ξ_l} per unit of traffic, respectively. As a result, concentrations satisfy the relationship

$$m_j = \underbrace{\sum_{i \in I} \gamma_{ji} \mu_y Y_i}_{\text{Pollution from industry}} + \underbrace{\sum_{e \in E} \gamma_{je} \mu_{\Xi_y} \Xi_e^y}_{\text{Pollution from goods traffic}} + \underbrace{\sum_{e \in E} \gamma_{je} \mu_{\Xi_l} \Xi_e^l}_{\text{Pollution from commuter traffic}} . \tag{12}$$

This formulation assumes a linear relationship between emissions and concentrations of the pollutants considered in the analysis across different locations. Evidence from the literature on atmospheric pollution dispersion modeling suggests that this is a sensible assumption for the pollutants, distances and timescales considered in our analysis. NOx and PM from traffic decrease with distance from a source due to dispersion and deposition, as well as chemical conversion of NOx to PM. In the ADMS model, dispersion processes are linear in terms of the relationship between emissions changes and concentrations. While deposition and chemical processes may both be non-linear removal processes, existing estimates suggest that this is likely to lead to relatively small differences over short distances (Air Quality Expert Group, Department for Environment, Food and Rural Affairs (2004); Thunis et al. (2019)).

3.3 Equilibrium

The economy is in equilibrium when households and firms optimize their decisions, taking as given prices, wages, and concentration levels, and goods markets and labor markets clear. Formally, for a given graph G with productivity and infrastructure invariants $\{\bar{A}_i\}_{i \in I}$, $\{\bar{t}_e\}_{e \in E}$ and residential population levels $\{L_i^0\}_{i \in I}$, an equilibrium is a distribution of labor and production, $\{L_i, Y_i\}_{i \in I}$ such that

(i) labor supply results from work choices and concentration levels as in (2), and work and consumption choices (including commuting route choice) are the solution to the problem of maximizing household utility, equation (1), given wages, price indices, transportation costs, and concentration levels;

(ii) firms choose labor inputs and distribution route choices to maximize profits;

(iii) productivity levels satisfy (10), transportation costs satisfy (11), and concentration levels satisfy (12);

(iv) traffic levels are consistent with commuting and goods transportation route choices;

(v) goods and labor markets clear.

As a result of (v), the total income earned is spent on purchases from goods from across locations, and total sales equals total labor income:

$$Y_j = \sum_i X_{ji}, \qquad Y_j = \sum_i X_{ij}. \tag{13}$$

where X_{ji} denotes purchases of workers in j of goods produced in i.

Appendix C shows that these equations, along with the equations determining labor supply in each location (2), and the equations governing price indices, (5) to (7), can be rewritten to characterize the equilibrium as

$$Y_i^{1+\theta} \bar{A}_i^{-\theta} L_i^{-\theta(1+\alpha)} = \sum_j t_{ji}^{-\theta} Y_j^{1+\theta} \bar{A}_j^{-\theta} L_j^{-\theta(1+\alpha)} + Y_i p_i^{\theta} \tag{14}$$

$$p_i^{-\theta} = \sum_j t_{ij}^{-\theta} p_j^{-\theta} + Y_i^{-\theta} \bar{A}_i^{\theta} L_i^{\theta(1+\alpha)} \tag{15}$$

$$L_j^{1+\theta} Y_j^{-\theta} p_j^{\theta} m_j^{(1+\theta)\sigma_w} = \sum_i t_{ji}^{-\theta} L_i^{1+\theta} Y_i^{-\theta} p_i^{\theta} m_i^{(1+\theta)\sigma_w} + m_j^{-\theta\sigma_h} \bar{W}_j^{-\theta} L_j^0 \tag{16}$$

$$\bar{W}_i^{\theta} m_i^{\theta\sigma_h} = m_i^{-\theta\sigma_w} Y_i^{\theta} L_i^{-\theta} p_i^{-\theta} + \sum_j t_{ji}^{-\theta} \bar{W}_j^{\theta} m_j^{\theta\sigma_h} \tag{17}$$

where $\bar{W}_i = m_i^{-\sigma_h} \left(\sum_j \left(m_j^{-\sigma_w} \frac{1}{\tau_{ji}} \frac{Y_j}{p_j L_j} \right)^{\theta} \right)^{1/\theta}$ denotes a term that is proportional to expected utility of residents in i, concentration levels satisfy (12), and trade costs satisfy

$$t_{kl}^{-\theta} = \bar{t}_{kl}^{-\frac{\theta}{1+\lambda\theta}} \left(\nu p_l^{-\theta} Y_k^{1+\theta} \bar{A}_k^{-\theta} L_k^{-\theta(1+\alpha)} + \bar{W}_k^{\theta} m_k^{\theta\sigma_h} L_l m_l^{(1+\theta)\sigma_w} w_l^{-\theta} p_l^{\theta} \right)^{-\frac{\theta\lambda}{1+\lambda\theta}}. \tag{18}$$

Equations (14) to (17) constitute a system of $4 \times I$ equations in $4 \times I$ variables $\{Y_i, L_i, p_i, \bar{W}_i\}_{i \in I}$.

3.4 Counterfactuals

To evaluate the welfare impact of counterfactual changes in transportation infrastructure, we use the "exact-hat" approach pioneered by Dekle, Eaton and Kortum (2008). We write the system of

equations above in changes, denoting the change in variables by hats, $\hat{x} = x^{\text{post}}/x^{\text{pre}}$,

$$\hat{Y}_i^{1+\theta}\hat{L}_i^{-\theta(1+\alpha)} = \frac{Y_i}{Y_i + \sum_k \Xi_{ki}^y}\hat{Y}_i\hat{p}_i^{\theta} + \sum_j \frac{\Xi_{ji}^y}{Y_i + \sum_k \Xi_{ki}^y}\hat{t}_{ji}^{-\theta}\hat{Y}_j^{1+\theta}\hat{L}_j^{-\theta(1+\alpha)} \tag{19}$$

$$\hat{p}_i^{-\theta} = \frac{Y_i}{Y_i + \sum_k \Xi_{ik}^y}\hat{Y}_i^{-\theta}\hat{L}_i^{\theta(1+\alpha)} + \sum_j \frac{\Xi_{ij}^y}{Y_i + \sum_k \Xi_{ik}^y}\hat{t}_{ij}^{-\theta}\hat{p}_j^{-\theta} \tag{20}$$

$$\hat{L}_i^{1+\theta}\hat{Y}_i^{-\theta}\hat{p}_i^{\theta}\hat{m}_i^{(1+\theta)\sigma_w} = \frac{L_i^0}{\sum_k \Xi_{ik}^l + L_i^0}\hat{m}_i^{-\theta\sigma_h}\widehat{\bar{W}}_i^{-\theta}\hat{L}_i^0 + \sum_j \frac{\Xi_{ij}^l}{\sum_k \Xi_{ik}^l + L_i^0}\hat{t}_{ij}^{-\theta}\hat{L}_j^{1+\theta}\hat{Y}_j^{-\theta}\hat{p}_j^{\theta}\hat{m}_j^{(1+\theta)\sigma_w} \tag{21}$$

$$\widehat{\bar{W}}_i^{\theta}\hat{m}_i^{\theta\sigma_h} = \frac{m_i^{\sigma_w}L_i}{m_i^{\sigma_w}L_i + \sum_k \Xi_{ki}^l}\hat{m}_i^{-\theta\sigma_w}\hat{Y}_i^{\theta}\hat{L}_i^{-\theta}\hat{p}_i^{-\theta} + \sum_j \frac{\Xi_{ji}^l}{m_i^{\sigma_w}L_i + \sum_k \Xi_{ki}^l}\hat{t}_{ji}^{-\theta}\widehat{\bar{W}}_j^{\theta}\hat{m}_j^{\theta\sigma_h} \tag{22}$$

In addition to these key equilibrium equations, we have two additional equations expressing changes in transportation costs $\hat{t}_{kl}$ and concentrations $\hat{m}_i$ as functions of changes in the above objects,

$$\hat{t}_{kl}^{-\theta} = \widehat{\bar{t}}_{kl}^{-\frac{\theta}{1+\lambda\theta}}\left(\frac{\nu\Xi_{kl}^y}{\nu\Xi_{kl}^y + \Xi_{kl}^l}\hat{p}_l^{-\theta}\hat{Y}_k^{1+\theta}\hat{L}_k^{-\theta(1+\alpha)} + \left(\frac{\Xi_{kl}^l}{\nu\Xi_{kl}^y + \Xi_{kl}^l}\right)\widehat{\bar{W}}_k^{\theta}\hat{L}_l^{1+\theta}\hat{Y}_l^{-\theta}\hat{p}_l^{\theta}\hat{m}_k^{\theta\sigma_h}\hat{m}_l^{(1+\theta)\sigma_w}\right)^{-\frac{\theta\lambda}{1+\lambda\theta}} \tag{23}$$

$$\hat{m}_i = \sum_j \hat{Y}_j\mathbf{M}_{ij}^y + \sum_{j,k}\hat{\Xi}_{jk}^y\mathbf{M}_{i,jk}^{\Xi^y} + \sum_{j,k}\hat{\Xi}_{jk}^{\ell}\mathbf{M}_{i,jk}^{\Xi^{\ell}} \tag{24}$$

and, based on the analytic expressions for traffic flows,

$$\hat{\Xi}_{ij}^y = \hat{t}_{ij}^{-\theta}\hat{p}_j^{-\theta}\hat{\Pi}_i^{-\theta} = \hat{t}_{ij}^{-\theta}\hat{p}_j^{-\theta}\hat{Y}_i^{1+\theta}\hat{L}_i^{-\theta(1+\alpha)} \tag{25}$$

$$\hat{\Xi}_{ij}^{\ell} = \hat{t}_{ij}^{-\theta}\hat{\Theta}_j^{-\theta}\hat{\Psi}_i^{-\theta} = \hat{t}_{ij}^{-\theta}\hat{L}_j^{1+\theta}\hat{Y}_j^{-\theta}\hat{p}_j^{\theta}\widehat{\bar{W}}_i^{\theta}\hat{m}_i^{\theta\sigma_h}\hat{m}_j^{(1+\theta)\sigma_w} \tag{26}$$

Solving this system of equations requires as inputs: (i) the coefficient matrices in the baseline equilibrium, which capture incoming and outgoing goods and worker traffic for each location; (ii) the matrices $\mathbf{M}_{ij}^y$, $\mathbf{M}_{ij}^{\Xi^y}$, and $\mathbf{M}_{ij}^{\Xi^{\ell}}$, which capture the fraction of pollution concentrations in each location that derive from industrial and traffic emissions from other locations[7]; (iii) the passenger-car-equivalent shares of goods traffic on each edge, $\nu\Xi_{kl}^y/(\nu\Xi_{kl}^y + \Xi_{kl}^{\ell})$; and (iv) the counterfactual under consideration, which determines the change in the transportation cost component $\widehat{\bar{t}}$ on the road segments in question.

[7]The precise definition is given in Appendix C.

4 Data

To take the model described in Section 3 to the data, we compile a rich combination of detailed georeferenced datasets capturing pollution, commuting, transport and economic activity at an extremely fine spatial resolution across the city of Lahore, in Pakistan's Punjab province. This section describes the sources and key features of these datasets; Section 5 then describes how we map these to key variables in the model in order to estimate the welfare impacts of road investments and policy counterfactuals.

4.1 Administrative boundaries

The spatial units used for the analysis are 361 "mauzas" (or mouzas) across the district of Lahore, shown in Figure (4). Mauzas are administrative revenue divisions with delineated boundaries, which we obtain from Urban Unit (2021).

4.2 Road network

We obtain the network of major roads across Lahore from OpenStreetMap (2022). We use road segment shapefiles for all motorways, trunks, primary, secondary, and tertiary roads (and their associated link roads) that intersect Lahore district. Raw road segments are split at all intersection points to yield a network of road "edges". The raw directed dataset classifies edges as one-way or two-way based on travel direction attributes; we simplify this raw network by merging multi-lane roads represented as two or more proximate parallel one-way roads into a single road segment.

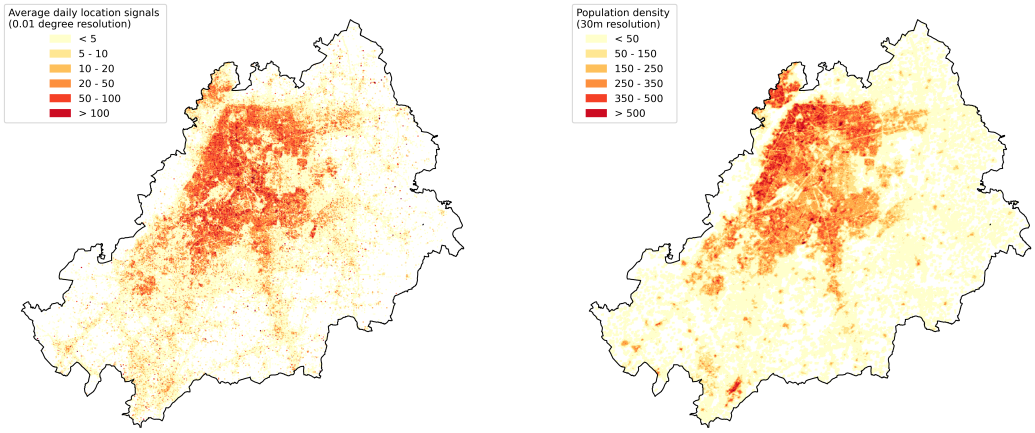
4.3 Mobile phone location data

Information on commuting flows between mauzas is based on anonymized mobile location data from Quadrant (2024). This dataset comprises approximately 1.2 billion GPS signals from over 16 million unique mobile devices across Pakistan during 2019. Each observation includes a device identifier, timestamp and geocoordinates.

We restrict attention to approximately 200 million signals from nearly 400,000 unique devices within Lahore district that satisfy the following criteria. First, we exclude signals with horizontal accuracy over 1km. Second, we consider only devices with more than 50 observations over the sample period, in order to be able to obtain meaningful information about the home and work locations of the device owner. Finally, we exclude a small number of erroneous observations, for instance where a device appears to travel an unrealistically large distance in a short period.[8] Panel (a) of Figure 3 shows the spatial distribution of GPS signals from mobile devices in this sample, which are intuitively most strongly concentrated in areas of the city with the highest population density, as shown in panel (b).

[8] To filter these cases, we follow Verma et al. (2024) and remove observations that have a segment speed over 50 m/s or acceleration outside the range of -10 to 10 m/s^2.

Figure 3 Mobile location data and population density



(a) GPS signals (b) Population density

This figure presents maps of GPS signals from mobile location data and population density in Lahore district. Panel (a) depicts the daily average number of recorded location signals in 2019 at a spatial resolution of 0.01°× 0.01°. Panel (b) illustrates population density estimates at a 30m resolution, based on data provided by Meta (2021).

4.4 Truck location data

We obtain data on the movement of trucks along Lahore's road network in 2018 from GPS trackers installed in more than 4,900 commercial trucks over this period. The data are obtained from a large original equipment manufacturer and described in more detail in Balboni, Boehm and Waseem (2023). GPS signals in this dataset include a timestamp, precise location and speed reading.

The GPS signals in this dataset are projected onto the closest edge in the road network described in Section 4.2 that lies within 10 meters of the observation coordinates, dropping observations that are at larger distances from the road network. This yields a total of 18 million (edge-day-vehicle) observations across Lahore during 2018.

4.5 Traffic surveys

Traffic survey data, which counts vehicles by vehicle type at road-side stations across the provinces of Punjab and Islamabad, yields information on location-specific traffic volumes and the vehicle fleet. We obtain such data from six surveys conducted by Government Ministries and international agencies between 2010 and 2020, described at Appendix B.2. Together these provide information across 671 traffic survey stations, 222 of which include precise coordinates, and the rest of which we manually geolocate based on descriptions of the survey station locations.

4.6 Firm location data

Data on the location of all VAT-registered firms over July 2011–June 2018 comes from Pakistan's Federal Board of Revenue.[9] Name and address information was used to geocode firm locations as described in Balboni, Boehm and Waseem (2023), which also describes the cleaning of this dataset to exclude firms with incomplete or potentially misreported transaction data. This yields a total of 15,294 geocoded firm locations across Lahore district.

4.7 Atmospheric dispersion model of pollution

In order to capture the generation and transportation of observed and counterfactual pollution emissions at a fine spatial and temporal resolution, we use a leading atmospheric dispersion model, newly calibrated to Punjab, Pakistan. The model used is the Multi-Model Air Quality System (MAQS), developed from the Atmospheric Dispersion Modeling System (ADMS-Urban) Regional Model Link system as described in Carruthers et al. (1998); Hood et al. (2018). The model captures both pollution from local sources (as in San José et al. (2018); Heblich, Trew and Zylberberg (2021)) and long-range chemical transport (as in Choi, Fuentes and Reich (2009); Fann et al. (2011); Peng et al. (2018)). It yields the concentration of $PM_{2.5}$, PM_{10}, NO_x and CO across space and time based on input data on meteorological conditions, background chemical concentrations, a gridded inventory of spatially aggregated emissions, and local emission sources. We compile detailed data on these inputs in Punjab; a more detailed description of these and the model is available at Appendix B.3.

4.8 Pollution monitor data

Air quality data was obtained from ground-level sensors covering 114 sensors across Punjab from 2016 to 2023. Of these, 14 report hourly $PM_{2.5}$ data, while the rest report daily information on PM_{10}, NO_2, NO, CO or SO_2 concentrations. These data were compiled from the Pakistan Air Quality Initiative, Pakistan Environmental Protection Agency, the Urban Unit, and pollution readings at US consulates, and are described in more detail at Appendix B.

4.9 Residential population data

The spatial distribution of residential population comes from Meta (2021), which provides a raster dataset of population density estimates at a 30-meter resolution worldwide.

[9] VAT registration is required for all importers, wholesalers and distributors, as well as manufacturers and retailers with revenue exceeding 10 million rupees in the previous tax reporting period and an annual utility bill above 800,000 rupees starting in July 2016. Registration thresholds were somewhat lower prior to July 2016.

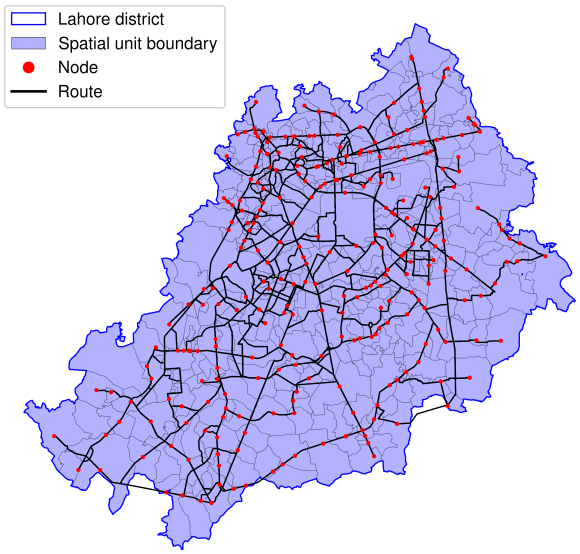
5 Model estimation

We use the administrative boundaries, road network and traffic data described in Section 4 to construct a graph that corresponds to the locations and edges in the model described in Section 3. Locations generally correspond to mauza boundaries, though in instances where no major road (up to and including tertiary roads) intersects a mauza, we aggregate it with the closest mauza that contains a major road. This yields a total of 287 locations.

In order to construct edges in the network, we project the centroid of each of these locations onto the nearest point on the road network[10], and calculate the pairwise fastest path between these points. This process uses the Dijkstra algorithm and assigns travel speeds to be the mean speed on each road type from the truck location data described in Section 4.4. We extract from this matrix only edges connecting adjacent locations.[11]

Figure 4 illustrates the location boundaries, the nearest points to their centroids on the road network, and the fastest routes connecting adjacent locations.

Figure 4 Map of spatial units and shortest paths



This figure shows the spatial unit boundaries and routes connecting each pair of adjacent locations. Red dots represent the nearest points on the road network from centroid of each spatial unit. Shortest routes are calculated using Dijkstra's algorithm.

The following subsections describe the empirical counterparts to the key inputs needed to calibrate the baseline model and run counterfactuals on this graph.

[10] In doing so, we prioritize the primary, then secondary, then tertiary road network in selecting the nearest edge.

[11] We include edges connecting spatial units that share a boundary and those that have at most one spatial unit separating them, in order to avoid artificial 'breaks' in the network where routes must pass through non-neighboring units to reach neighboring units.

5.1 Commuter and goods traffic

Commuter flows between each pair of locations are estimated by using cellphone GPS signals to identify home and workplace locations for each device following Kreindler and Miyauchi (2023). We assign a device's 'home' location to the geocoordinate that appears most frequently in the signals it emits between 9pm and 5am, and 'work' locations to the geocoordinates that appear most frequently between 10am and 3pm for each device-day excluding weekends. Home-work pairs that share the same geocoordinate are dropped.

While we restrict each device to have a single home location[12], it is often the case that a device has multiple work locations. We weigh each unique home-work pair by its share of all home-work pair observations at the device-day level originating from the device's primary home location. These are then aggregated into a location-level origin-destination commuting matrix. By multiplying the matrix of commuting probabilities $(\omega_{ji})_{j,i}$ by the vector of residents by location, L_i^0, we obtain the vector of the mass of workers in each location, $m_i^{\sigma_w} L_i = \sum_j \omega_{ij} L_j^0$.

To obtain the matrix of commuter traffic Ξ^l (goods traffic Ξ^y), we project GPS signals from cellphones (trucks) onto the road network; calculate the shortest path between each sequential pair of projected signals; and use this to calculate the number of commuters (trucks) that traverse at least a part of each edge on each day. These counts are de-meaned by projecting them onto day and edge fixed effects, where edge fixed effects can be interpreted as average day-cellphone (day-truck) counts for each edge.

Commuter traffic flows Ξ^l are calculated by multiplying these counts by a scaling factor chosen to match the average annual daily traffic (AADT) counts from the traffic surveys. For goods traffic, we choose the scaling factor for the truck-day count to match average AADT truck counts. Traffic counts are converted into nominal values by assuming that the value of the average truck shipment is equal to average per-capita annual income in urban Pakistan (PKR 267,597, from Pakistan Bureau of Statistics, 2020). This assumption follows Garg, Ghosh and Tan (2023), who document that the average shipment value in the Indian state of Karnataka over 2018-19 closely matches the state's average per-capita annual income.[13]

5.2 Source apportionment matrices

Source apportionment matrices $\mathbf{M}_{ij}^{y}, \mathbf{M}_{ij}^{\Xi_y}$ and $\mathbf{M}_{ij}^{\Xi_l}$ capture the share of pollution concentrations in each location i that can be attributed to industrial emissions, goods traffic, or commuter traffic, respectively, from location or edge j. We calculate these in three steps.

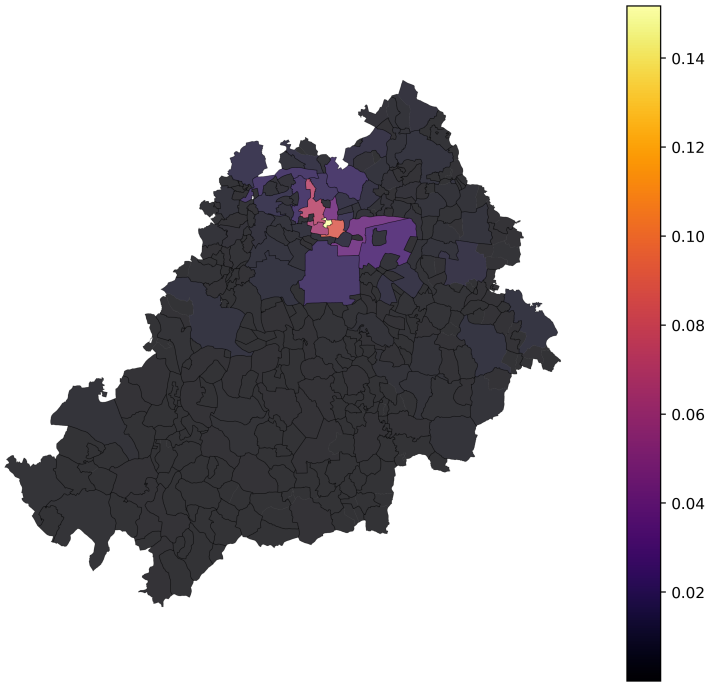
First, the pollution transport matrix γ_{ij} is estimated using the ADMS model at a spatial resolution of 500m×500m, averaging concentrations across a year. For this estimation, we use meteorological data for 2018 and model the locations used in the analysis as homogeneous emission sources with equal emission rates. Figure 5 shows an example of the output of γ_{ij} for a given

[12] Any home-work pairs associated with 'home' locations other than the primary home location are removed.

[13] Average shipment value of 202,291 INR, relative to per-capita annual income of 205,245 INR (data from the Indian Ministry of Statistics and Programme Implementation).

location, displaying the concentrations attributed to emissions from neighboring areas. Second, the Emissions Database for Global Atmospheric Research (EDGAR) (see, for example, Crippa et al. (2020)), which provides a gridded top-down sectoral emissions inventory at 0.1°×0.1° resolution, is used to estimate μ_y. Finally, generating the source apportionment matrices from these inputs also requires the number of firms in each location, which we obtain from the firm location data based on VAT records.

Figure 5 Example of a row γ_{ij}



This figure shows an example of γ_{ij} for a given row i. These can be interpreted as the proportion of concentrations in mauza i which are attributable to mauza j. These are calculated using the source apportionment option of the ADMS model with mauzas modelled as areas with homogenous emissions, receptors placed in a 500m×500m grid and using 2018 meteorological data. γ_{ij} are calculated as the PM2.5 concentration from mauza j as a share of concentration from all sources, averaged across all receptors in i.

5.3 Calibrated model parameters

The trade elasticity is calibrated to $\theta = 6.83$, the congestion elasticity in traffic to $\lambda = 0.092$, and the agglomeration parameter to $\alpha = -0.12$, following Allen and Arkolakis (2022).

The parameter ν governs the relative congestion effect of goods traffic versus commuter traffic. We calibrate this parameter using the ratio between the average passenger car equivalents (PCE) for vehicles used for commuting relative to those used to trade goods (TRB, 2010). Traffic shares for each vehicle type are taken from traffic surveys in (JICA, 2010) and PCE values from Farooq and Akram (2018) (see Appendix Table A.1). This yields a value of $\nu = 5.01$.

The elasticities σ_h and σ_w capture the percentage change in utility and earnings, respectively, with a percentage change in pollution concentrations. We calibrate these parameters using estimates

based on particulate matter exposure in other middle income countries.

Greenstone et al. (2025) study particulate pollution in urban India, and value the gain in life-years from reduced exposure using estimates from Ebenstein et al. (2017) that a lifetime increase in particulate matter by 1 $\mu g/m^3$ reduces life expectancy in China by 0.064 years, together with average life expectancy of 70 years and value of a statistical life-year of \$9,500 in India. In our setting, the median particulate matter concentration is $90\mu g/m^3$, such that a 1% reduction on average corresponds to \$7.8 using these values. Given average annual income in urban Pakistan of PKR 267,597 (USD 2193) in 2018-2019 (Pakistan Bureau of Statistics, 2020; Exchange-rates.org, 2018), this yields $\sigma_h = 0.36$.[14]

The percentage change in earnings with a percentage change in concentrations, σ_w, is calibrated based on evidence from Adhvaryu, Kala and Nyshadham (2022) in India. This study finds that a $16\mu g/m^3$ increase in fine particulate matter exposure reduces productivity by 0.8 percentage points, from mean productivity levels of 49%. Based on this, we use a central value of $\sigma_w = 0.09$.

6 Counterfactual simulations

The model is used to estimate the benefits and costs of major realized and counterfactual road investment projects in Lahore, accounting for both market integration gains and the local costs of air pollution externalities.

The first counterfactual exercise estimates the impacts of the Lahore ring road by simulating the effects of its removal. The Lahore ring road is an orbital highway surrounding the central city, shown in Figure 6, which connects to major national highways. Construction began in 2006 (Punjab Ring Road Authority (2018)), with opening of the Northern Loop in 2012, two sections of the Southern Loop in 2017, and a third section of the Southern Loop in 2024 (Directorate General Public Relations Punjab (2024)); the Western Loop remains in the planning phase.

The ring road carries an estimated 400,000-500,000 vehicles daily, primarily comprising heavy transport and commuter vehicles using the route to bypass urban congestion (Punjab Ring Road Authority (2018)).[15] This has resulted in significant shifts in the location of traffic flows throughout the city. An example is shown in Figure 7, which shows traffic flows in the GPS truck location data described in Section 4.4 before and after the opening of the southern section of the ring road connecting Kamahan to Lake City in December 2017. A clear change in the location of truck signals can be seen as trucks start to traverse the ring road route along the bottom and right edges of the panels.

The central estimates suggest that removing the Northern and Southern loops of the Lahore

[14] This calculation uses an estimate of the value of a statistical life-year from Nair et al. (2021). Other estimates of marginal willingness to pay for clean air in India based on observed purchases of defensive technologies yield lower estimates, likely driven by information and behavioral frictions in demand for clean air. In line with the latter estimates, we consider the sensitivity of results to assuming a marginal willingness to pay of \$1.14 for a 10 $\mu g/m^3$ reduction in annual $PM_{2.5}$ concentrations, following Baylis et al. (2024).

[15] Access is restricted to high-performance vehicles and excludes pedestrians, animal-driven carriages, bicycles, and low-capacity motorcycles.

Figure 6 Lahore ring road

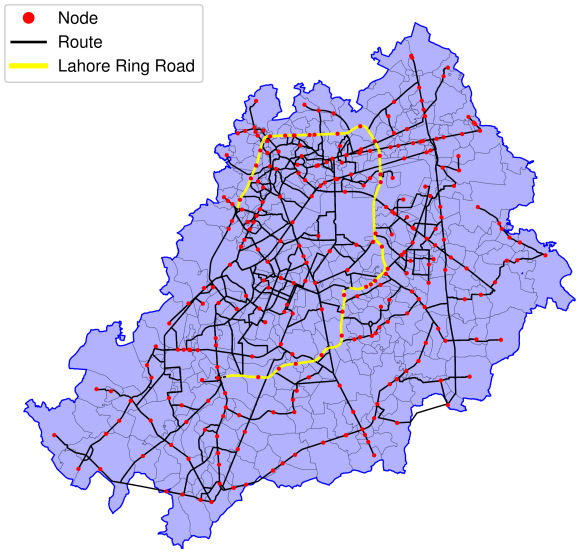


Figure 7 GPS signals before and after opening of the southern section of the Lahore ring road



(a) 21 December 2017 (b) 28 February 2018

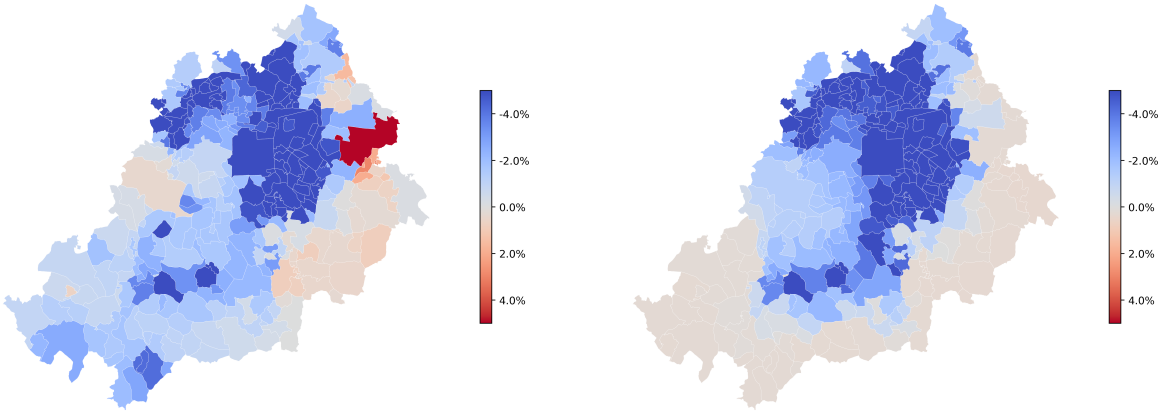
The figure shows the location of GPS signals from truck location data before (panel (a)) and after (panel (b)) the opening of the southern seciton of the Lahore ring road on 22 December 2017.

ring road[16] would result in an aggregate welfare loss of 3.80%, broadly in line with estimates of the welfare impacts from other major urban transport infrastructure improvements in developing country cities (e.g. Tsivanidis (2022); Zárate (2022)). There is substantial spatial heterogeneity in the distribution of losses from removal of the ring road across spatial units, as shown in Figure 8(a). Welfare losses are intuitively most pronounced in areas directly neighboring the ring road route, which suffer from large reductions in their goods and commuter market access with its removal.

The reduction in aggregate welfare from removal of the ring road reflects the net impact of increases in trade and commuting costs across the city, as well as changes in damages from traffic-induced air pollution. To quantify the importance of the latter channel, we resimulate the model shutting down pollution adjustments.

[16]'Removal' of the ring road is implemented in the model estimation as a 200% increase in transportation costs along the edges in the road network corresponding to the ring road.

Figure 8 Welfare changes from removing the Lahore ring road



(a) Overall changes

(b) Excluding pollution

The results of this exercise reveal that changes in vehicle-induced air pollution brought about by the ring road significantly alter both the aggregate and distributional welfare gains that it confers. While changes in trade and commuting costs alone would have implied welfare losses of 4.30% from removing the ring road, the induced reductions in pollution damages from concentrating traffic in less densely-populated areas reduce these losses to 3.80% (a 12% reduction in losses). As shown in Figure 8(b), abstracting from the effects on pollution also results in a shift in the spatial distribution of welfare gains from removing the ring road.

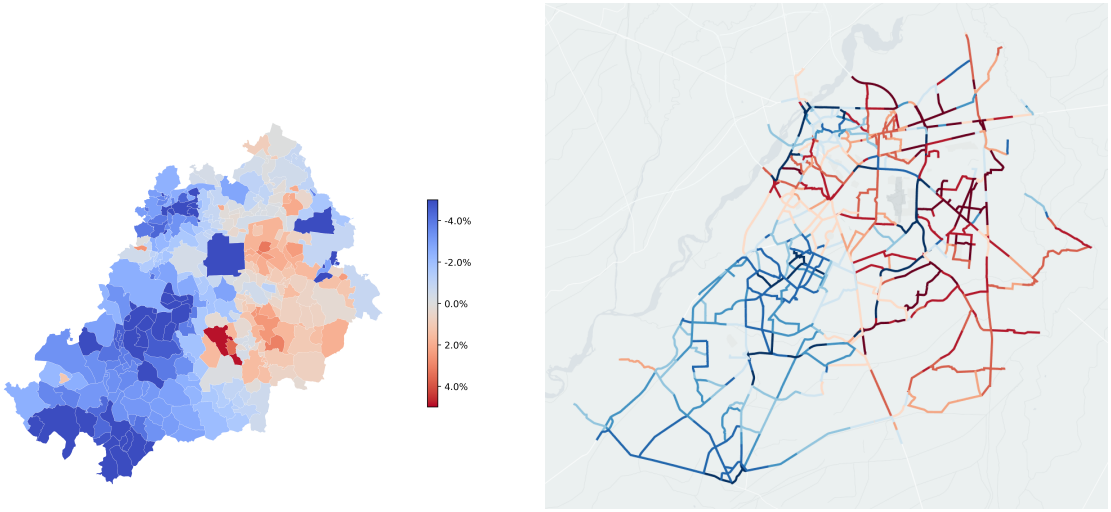
Figure 9 maps the spatial distribution of changes in pollution concentrations (left-hand panel) and commuter traffic (right-hand panel) induced by removal of the ring road. Pollution increases are especially pronounced in the east of Lahore, while reductions are strongest in the west and south. The pattern of changes in commuting traffic reveals that this is driven by significant substitution of commuter traffic away from the ring road route to neighboring road edges in the east of the city, where higher levels of congestion drive large increases in ambient pollution concentrations.

The spatial heterogeneity in impacts across spatial units shown in Figure 8(b) makes clear that accounting for pollution damages changes the distributional, as well as aggregate, impacts of the ring road substantially. Trade and commuting cost changes alone would suggest positive impacts of the ring road removal in 63 spatial units (22% of the total spatial units). In 43 locations, accounting for pollution changes our assessment of whether removal of the ring road results in an overall increase or decrease in local welfare.

The results suggest that abstracting from the local pollution impacts of this important new road investment would lead to a quantitatively important over-estimate of the welfare gains associated with its construction. In this case, planners would also draw erroneous conclusions about its distributional impacts, wrongly concluding that there are benefits in some regions that actually experience substantial welfare losses from the new infrastructure.

Finally, we conduct a counterfactual simulation exercise which aims to estimate the *aggregate*

Figure 9 Changes in pollution concentrations and traffic from removing the Lahore ring road

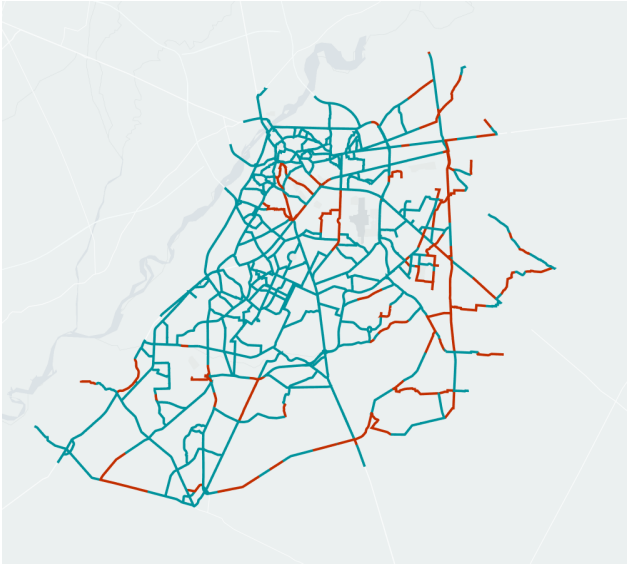


(a) Changes in pollution concentrations

(b) Changes in commuter traffic

(i.e. population-weighted) change in welfare $\bar{W}_i$ when each road segment in Lahore's road network is removed (again implemented as a 200% increase in transportation costs). Figure 10 colors the roads whose removal would result in an aggregate welfare decrease in green, and roads where aggregate welfare would increase in red. The map shows that roads for which a shutdown would result in a net welfare increase (which is the case for 110 out of 623 roads) are concentrated on the outskirts of the city. This is consistent with peripheral roads being associated with less pronounced trade and commuting gains, since they are less crucial to connecting outlying areas to the center, but still having negative impacts on local residents via increased particulate concentration levels. This analysis highlights that, once the local externalities associated with vehicular pollution are accounted for, not every road segment necessarily yields a net welfare improvement, and a utilitarian planner may therefore derive gains from shutting down some road segments.

Figure 10 Net welfare gains from road removal



The figure shows the location of roads, colored in red, where aggregate (i.e. population-weighted) welfare $\bar{W}_i$ would increase when the road is shut down (simulated as a 200% increase in trade costs). All other roads are colored in green.

7 Conclusions

Rapid advances in our understanding of the magnitude and damages from local air pollution has contributed to large-scale government programs to control pollution in low and middle income countries, including China's 2014 'War on Pollution' and India's 2019 'National Clean Air Programme' (Greenstone et al., 2021; Ganguly, Selvaraj and Guttikunda, 2020). While pollution control measures often focus on command-and-control regulation of industrial emissions, the results of this paper suggest that transport infrastructure investments – a cornerstone of development policy in many developing economies – may also have first order impacts on air pollution exposure.

Using a quantitative urban equilibrium model integrated with a granular atmospheric model of air pollution dispersion, we consider how far accounting for such effects alters the estimated welfare gains from major transportation infrastructure investments in Lahore, Pakistan. We find that incorporating the local air pollution impacts of major road investments has an important bearing on both the estimated aggregate welfare gains from these investments, and assessments of which regions stand to gain or lose. The results point to the importance of understanding and accounting for local externality impacts, alongside trade and commuting cost gains, in evaluating and assessing optimal placement of spatial policy investments.

References

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2022. "Management and shocks to worker productivity." Journal of Political Economy, 130(1): 1–47.

Air Quality Expert Group, Department for Environment, Food and Rural Affairs. 2004. "Nitrogen Dioxide in the United Kingdom."

Allen, Treb, and Costas Arkolakis. 2022. "The welfare effects of transportation infrastructure improvements." The Review of Economic Studies, 89(6): 2911–2957.

Almagro, Milena, Felipe Barbieri, Juan Camilo Castillo, Nathaniel G Hickok, and Tobias Salz. 2024. "Optimal urban transportation policy: Evidence from chicago." National Bureau of Economic Research.

Bakht, Masoud, and Muhammad Omair Ahmad. 2012. "Traffic Counts on M-9 and National Highway N-05." National Transport Research Center, Ministry of Communications, Government of Pakistan NTRC-295.

Balboni, Clare A, and Joseph S Shapiro. 2025. "Spatial Environmental Economics." National Bureau of Economic Research.

Balboni, Clare, Johannes Boehm, and Mazhar Waseem. 2023. "Firm adaptation and production networks: Structural evidence from extreme weather events in Pakistan."

Bassi, Vittorio, Matthew E Kahn, Nancy Lozano Gracia, Tommaso Porzio, and Jeanne Sorin. 2022. "Jobs in the Smog: Firm Location and Workers' Exposure to Pollution in African Cities." National Bureau of Economic Research.

Baylis, Patrick, Michael Greenstone, Kenneth Lee, and Harshil Sahai. 2024. "Is the Demand for Clean Air Too Low? Experimental Evidence from Delhi."

Bieser, J., A. Aulinger, V. Matthias, M. Quante, and H.A.C. Denier van der Gon. 2011. "Vertical emission profiles for Europe based on plume rise calculations." Environmental Pollution, 159(10): 2935–2946. Nitrogen Deposition, Critical Loads and Biodiversity.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis. 2020. "Cities in the developing world." Annual Review of Economics, 12(1): 273–297.

Byun, Daewon, and Kenneth L. Schere. 2006. "Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System." Applied Mechanics Reviews, 59(2): 51–77.

Carruthers, D. J., H. A. Edmunds, C. A. McHugh, and R. J. Singles. 1998. "Development of Adms-Urban and Comparison with Data for Urban Areas in the UK." Air Pollution Modeling and Its Application XII, , ed. Sven-Erik Gryning and Nadine Chaumerliac, 467–475. Boston, MA:Springer US.

Chay, Kenneth Y, and Michael Greenstone. 2005. "Does air quality matter? Evidence from the housing market." Journal of political Economy, 113(2): 376–424.

Choi, Jungsoon, Montserrat Fuentes, and Brian J. Reich. 2009. "Spatial–temporal association between fine particulate matter and daily mortality." Computational Statistics and Data Analysis, 53(8): 2989–3000.

Crippa, Monica, Efisio Solazzo, Ganlin Huang, Diego Guizzardi, Ernest Koffi, Marilena Muntean, Christian Schieberle, Rainer Friedrich, and Greet Janssens-Maenhout. 2020. "High resolution temporal profiles in the Emissions Database for Global Atmospheric Research." Scientific data, 7(1): 121.

Cristea, Anca, David Hummels, Laura Puzzello, and Misak Avetisyan. 2013. "Trade and the greenhouse gas emissions from international freight transport." Journal of environmental economics and management, 65(1): 153–173.

Currie, Janet, and Reed Walker. 2011. "Traffic congestion and infant health: Evidence from E-ZPass." American Economic Journal: Applied Economics, 3(1): 65–90.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. "Global rebalancing with gravity: measuring the burden of adjustment." IMF Economic Review, 55(3): 511–540.

Denier, Hugo van der Gon, Carlijn Hendriks, Arjo Segers Jeroen Kuenen, and Antoon Visschedijk. 2011. "Description of current temporal emission patterns and sensitivity of predicted AQ for temporal emission patterns." TNO report, EU FP7 MACC deliverable report D_D-EMIS_1.3.

Directorate General Public Relations Punjab. 2024. "Lahore Ring Road Southern Loop 3 Project opens to traffic." https://dgpr.punjab.gov.pk/node/24210, Accessed: 2025-03-10.

Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, and Maigeng Zhou. 2017. "New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy." Proceedings of the National Academy of Sciences, 114(39): 10384–10389.

EMP/EEA. 2019. "EMEP/EEA air pollutant emission inventory guidebook - Technical guidance to prepare national emission inventories." EEA Report, , (13).

Exchange-rates.org. 2018. "US Dollar (USD) To Pakistan Rupee (PKR) Exchange Rate History for 2018." https://www.exchange-rates.org/exchange-rate-history/usd-pkr-2018.

Fann, Neal, Amy Lamson, Susan Anenberg, Karen Wesson, David Risley, and Bryan Hubbell. 2011. "Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone." Risk analysis : an official publication of the Society for Risk Analysis, 32: 81–95.

Farooq, Danish, and Tayyab Akram. 2018. "Traffic Flow Analysis and Solutions to Ease Traffic Flow at Unsignalized Taxila Intersection." Periodica Polytechnica Transportation Engineering, 46.

Freeman, Richard, Wenquan Liang, Ran Song, and Christopher Timmins. 2019. "Willingness to pay for clean air in China." Journal of Environmental Economics and Management, 94: 188–216.

Ganguly, Tanushree, Kurinji L Selvaraj, and Sarath K Guttikunda. 2020. "National Clean Air Programme (NCAP) for Indian cities: Review and outlook of clean air action plans." Atmospheric Environment: X, 8: 100096.

Garg, Shresth, Pulak Ghosh, and Brandon Joel Tan. 2023. "Within firm supply chains: Evidence from India." Journal of International Economics, 144: 103793.

Govt of Punjab. 2014. "Traffic Summary 2014, Provincial & District Roads in Punjab." Government of the Punjab, Communication & Works Department.

Greenstone, Michael, and Q Fan. 2019. "Pakistan's Air Pollution Challenge & Potential for Longer Lives." Air Quality Life Index (AQLI), February.

Greenstone, Michael, Guojun He, Shanjun Li, and Eric Yongchen Zou. 2021. "China's war on pollution: Evidence from the first 5 years." Review of Environmental Economics and Policy, 15(2): 281–299.

Greenstone, Michael, Rohini Pande, Nicholas Ryan, and Anant Sudarshan. 2025. "Can pollution markets work in developing countries? Experimental evidence from India." The Quarterly Journal of Economics, qjaf009.

Gulzar, Ahmed, Shahbaz Latif Mirza, and M.S. Qureshi. 2013. "Diligence Checks on NHA Toll Plazas M-9." National Transport Research Center, Ministry of Communications, Government of Pakistan NTRC-329.

Hall, James, Jian Zhong, Sue Jowett, Andrea Mazzeo, G. Neil Thomas, John R. Bryson, Steve Dewar, Nadia Inglis, Mark Wolstencroft, Catherine Muller, William James Bloss, Roy M. Harrison, and Suzanne E. Bartington. 2024. "Regional impact assessment of air quality improvement: The air quality lifecourse assessment tool (AQ-LAT) for the West Midlands combined authority (WMCA) area." Environmental Pollution, 123871.

Heblich, Stephan, Alex Trew, and Yanos Zylberberg. 2021. "East-side story: Historical pollution and persistent neighborhood sorting." Journal of Political Economy, 129(5): 1508–1552.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm. 2020. "The making of the modern metropolis: evidence from London." The Quarterly Journal of Economics, 135(4): 2059–2133.

Hoffmann, Bridget, and Juan Pablo Rud. 2024. "The unequal effects of pollution on labor supply." Econometrica, 92(4): 1063–1096.

Holub, Felix, and Beate Thies. 2023. "Air quality, high-skilled worker productivity and adaptation: Evidence from GitHub." Working Paper.

Hood, C., I. MacKenzie, J. Stocker, K. Johnson, D. Carruthers, M. Vieno, and R. Doherty. 2018. "Air quality simulations for London using a coupled regional-to-local modelling system." Atmospheric Chemistry and Physics, 18(15): 11221–11245.

Jensen, Steen Solvang, Matthias Ketzel, Thomas Becker, Jesper Christensen, Jørgen Brandt, Marlene Plejdrup, Morten Winther, Ole-Kenneth Nielsen, Ole Hertel, and Thomas Ellermann. 2017. "High resolution multi-scale air quality modelling for all streets in Denmark." Transportation Research Part D: Transport and Environment, 52: 322–339.

JICA. 2010. "The Project For Lahore Urban Transport Master Plan in the Islamic Republic of Pakistan." Japan International Cooperation Agency for the Transport Department, Government of the Punjab.

Kahn, Matthew E. 2007. Green cities: urban growth and the environment. Rowman & Littlefield.

Knittel, Christopher R, Douglas L Miller, and Nicholas J Sanders. 2016. "Caution, drivers! Children present: Traffic, pollution, and infant health." Review of Economics and Statistics, 98(2): 350–366.

Kreindler, Gabriel. 2024. "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications." Econometrica, 92(4): 1233–1268.

Kreindler, Gabriel E, and Yuhei Miyauchi. 2023. "Measuring commuting and economic activity inside cities with cell phone records." Review of Economics and Statistics, 105(4): 899–909.

Leroutier, Marion, and Hélene Ollivier. 2022. "The cost of air pollution for workers and firms." mimeo. 13th Toulouse Conference on The Economics of Energy and Climate.

Meta. 2021. "Data for Good at Meta: High Resolution Population Density Maps." Dataset covering Pakistan and India available at https://data.humdata.org/dataset/pakistan-india_all-files-high-resolution-population-density-maps#.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, migration, and local employment elasticities." American Economic Review, 108(12): 3855–3890.

Nagpure, Ajay Singh, Bhola Ram Gurjar, Vivek Kumar, and Prashant Kumar. 2016. "Estimation of exhaust and non-exhaust gaseous, particulate matter and air toxics emissions from on-road vehicles in Delhi." Atmospheric environment, 127: 118–124.

Nair, Moorthy, Hemant Bherwani, Shahid Mirza, Saima Anjum, and Rakesh Kumar. 2021. "Valuing burden of premature mortality attributable to air pollution in major million-plus non-attainment cities of India." Scientific reports, 11(1): 22771.

NTRC. 2017. "Pilot Traffic Count Survey on N-5 (Peshawar to Karachi)." School of Civil & Environment Engineering, National University of Science and Technology for the National Transport Research Center, Ministry of Communications, Government of Pakistan NTRC-313.

NTRC. 2020. "Traffic Summary 2020 Phase I-II." National Transport Research Center, Ministry of Communications, Government of Pakistan.

Oguge, Otienoh, Joshua Nyamondo, Noah Adera, Lydia Okolla, Beldine Okoth, Stephen Anyango, Augustine Afulo, Abera Kumie, Jonathan Samet, and Kiros Berhane. 2024. "Fine particulate matter air pollution and health implications for Nairobi, Kenya." Environmental Epidemiology, 8(3): e307.

OpenStreetMap. 2022. "HOT OpenStreetMap Export Dataset." https://export.hotosm.org/en/v3/exports/dfc79247-171d-402a-9816-f21786f46a4b/.

Pakistan Bureau of Statistics. 2020. "Household Integrated Economic Survey 2018-19." https://www.pbs.gov.pk/sites/default/files//pslm/publications/hies2018-19/TABLE_09.pdf.

Peng, Wei, Junnan Yang, Xi Lu, and Denise L. Mauzerall. 2018. "Potential co-benefits of electrification for air quality, health, and CO2 mitigation in 2030 China." Applied Energy, 218(C): 511–519.

Punjab Ring Road Authority. 2018. "Background." https://prra.punjab.gov.pk/background, Accessed: 2025-03-10.

Quadrant. 2024. "Mobile Location Data." https://www.quadrant.io/mobile-location-data.

Redding, Stephen J, and Esteban Rossi-Hansberg. 2017. "Quantitative spatial economics." Annual Review of Economics, 9(1): 21–58.

Redding, Stephen J, and Matthew A Turner. 2015. "Transportation costs and the spatial organization of economic activity." Handbook of regional and urban economics, 5: 1339–1398.

San José, Roberto, Juan Luis Pérez, Libia Pérez, and Rosa Maria Gonzalez Barras. 2018. "Effects of climate change on the health of citizens modelling urban weather and air pollution." Energy, 165(PA): 53–62.

Sarmiento, Luis. 2022. "Air pollution and the productivity of high-skill labor: evidence from court hearings." The Scandinavian Journal of Economics, 124(1): 301–332.

Shabbir, Asma. 2024. "Environmental Degradation and Smog: Assessing the Governance Mechanism in Lahore." Journal of Climate and Community Development, 3(2): 232–243.

Shah, Izhar Hussain, and Muhammad Zeeshan. 2016. "Estimation of light duty vehicle emissions in Islamabad and climate co-benefits of improved emission standards implementation." Atmospheric Environment, 127: 236–243.

Shapiro, Joseph S. 2016. "Trade costs, CO2, and the environment." American Economic Journal: Economic Policy, 8(4): 220–254.

Simpson, D., A. Benedictow, H. Berge, R. Bergström, L. D. Emberson, H. Fagerli, C. R. Flechard, G. D. Hayman, M. Gauss, J. E. Jonson, M. E. Jenkin, A. Nyíri, C. Richter, V. S. Semeena, S. Tsyro, J.-P. Tuovinen, Á. Valdebenito, and P. Wind. 2012. "The EMEP MSC-W chemical transport model technical description." Atmospheric Chemistry and Physics, 12(16): 7825–7865.

Taimoor, Aqeel, Zafar Iqbal Khan, Javaid Rabbani, and Fahd Nawaz. 2023. "Article Effect of Fuel Quality, Vehicle Maintenance and Advanced Emission Control Technology on Pakistan In -Use Light Vehicle Emissions." Journal of King Abdulaziz University: Engineering Sciences, 33(1): 95–112.

Thunis, Philippe, Alain Clappier, Leonor Tarrasón, Cornelis Cuvelier, A Monteiro, Enrico Pisoni, Joost Wesseling, CA Belis, G Pirovano, Stijn Janssen, et al. 2019. "Source apportionment to support air quality planning: Strengths and weaknesses of existing approaches." Environment International, 130: 104825.

TRB. 2010. "The Highway Capacity Manual, 6th edition: A guide for multimodal mobility analysis." TR News, 86: 14–18.

Tsang, Stephanie, and Muhammad Adil. 2022. "Pakistan Automotive Legislation: Driving Emissions Up or Slowing Them Down?" RSIL Law Review, 6.

Tsivanidis, Nick. 2022. "Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio." Unpublished manuscript, 18.

Urban Unit. 2021. "Digital Mouza Boundaries." https://urbanunit.gov.pk/Download/publications/Files/11/2021/Digital%20Mouza%20Boundries.pdf.

Usman, Muhammad, Hafsa Yasin, Haroon Rashid, and Abdul Nasir. 2017. "Quantification of CO2 emissions from vehicles and possible remedial strategies in Faisalabad City." Earth Sciences Pakistan (ESP), 1(1): 15–18.

Verma, Rajat, Shagun Mittal, Zengxiang Lei, Xiaowei Chen, and Satish V Ukkusuri. 2024. "Comparison of home detection algorithms using smartphone GPS data." EPJ Data Science, 13(1): 6.

Zárate, Román D. 2022. "Spatial misallocation, informality, and transit improvements: Evidence from mexico city."

A Additional Tables and Figures

Table A.1 Vehicle shares and Passenger car equivalent (PCE) values

Vehicle	Traffic shares (%)	PCE
Motorcycle	48.06	0.5
Car	25.66	1
LGV	9.23	3
Bus	2.90	3.5
RHGV 2 axles	6.55	4
RHGV 3 axles	2.14	5
AHGV 3/4/5/6 axles	5.46	6

Table A.2 Traffic shares are calculated from the Lahore Urban Transport Master Plan JICA (2010) averaged across all survey locations. PCE values are taken from the Highway Capacity Manual as reported in Farooq and Akram (2018).

B Data Appendix

B.1 Ground-level air quality sensors data

Ground-level data on air quality was collected from the the following sources:

1. The Pakistan Air Quality Initiative provided hourly data on $PM_{2.5}$ concentrations from 10 sensors located primarily in Lahore from 2016 to 2019. This initiative shares air quality data from private low-cost air quality monitors.
2. The Urban Unit provided data on $PM_{2.5}$ at two sensors in Lahore, located at the Aljazari Water and Sanitation Academy and Forman Christian College, and additional data on PM_{10}, NO_2, CO and NO_2 at the first of these locations. Pollutants are reported hourly for many days of the year between 2019 and 2023.
3. Hourly data on $PM_{2.5}$ recorded at two sensors at the American consulates in Islamabad and Lahore is obtained for most days from 2019 to 2023.
4. We collected daily data from 100 sensors across 12 cities in Pakistan from publicly available Air Quality Monitoring Reports of Pakistan's Environmental Protection Agency. The location of monitors was inferred from location descriptions in these reports and may therefore be imprecise.

   The location of all monitors in our sample is shown in Figure B.1. Table B.1 summarizes the number of observations from each source in our data.

B.2 Traffic survey data

Data on traffic counts is sourced from the following surveys:

Figure B.1 Map of air quality sensor locations

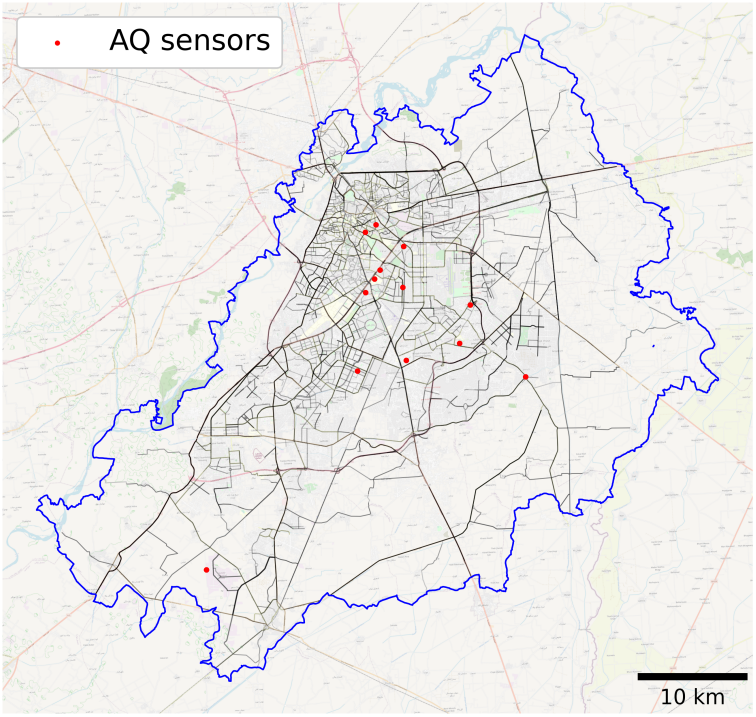


Table B.1 Pollutant detection days by data source

Data sources	N° of sensors	2016	2017	2018	2019	2020	2021	2022	2023
Urban Unit (AWSA)	1	0	0	0	0	0	219	361	158
Urban Unit (FCC)	1	0	0	0	43	313	303	289	200
US consulate (Islamabad)	1	0	0	0	251	366	300	345	274
US consulate (Lahore)	1	0	0	0	237	366	297	346	260
Pakistan Air Quality Initiative	10	3	137	365	7	0	0	0	0
Pakistan EPA	100	0	48	321	363	365	364	365	365

Table B.2 Sensor information

Location	Data source	Lat	Lon	Start	End	N	Distance to: Road	Main Road	High gps road
Abu Bakar Block	PAQI	31.512	74.323	2018-09-01	2019-01-07	2885	102.8	270.0	102.8
Beaconhouse Canal Campus	PAQI	31.455	74.359	2018-11-30	2019-01-07	700	396.1	396.1	396.1
Beaconhouse Liberty Campus	PAQI	31.514	74.359	2018-11-30	2019-01-07	916	37.5	37.5	37.5
DHA Phase 5 (Block F)	PAQI	31.466	74.410	2016-12-29	2019-01-07	8552	237.7	237.7	237.7
FCC Syed Maratib Ali Road	PAQI	31.529	74.338	2018-01-11	2018-07-03	1166	137.7	137.7	137.7
Fida Hussain House	PAQI	31.436	74.471	2018-03-06	2019-01-07	3843	4.5	4.5	4.5
Punjab Assembly	PAQI	31.561	74.325	2018-12-13	2019-01-07	604	45.9	146.6	146.6
Lahore Upper Mall	PAQI	31.548	74.361	2017-11-01	2019-01-07	9899	133.7	246.3	246.3
Ghazi Road Interchange	PAQI	31.497	74.422	2018-01-11	2019-01-07	5678	10.0	14.7	14.7
Sundar Industrial Estate	PAQI	31.292	74.159	2018-12-07	2019-01-07	700	605.4		
Aljazari Water and Sanitation Academy	Urban Unit	31.448	74.312	2021-05-11	2023-07-03	18172	68.9	275.6	275.6
Forman Christian College	Urban Unit	31.522	74.332	2019-11-18	2023-10-02	122227	208.1	208.1	208.1
US consulate	AirNow	31.560	74.336	2019-05-09	2023-10-01	35758	36.0	36.0	36.0

1. Traffic count data from the Lahore Urban Transport Master Plan 2010 conducted by the Japan International Cooperation Agency for the Government of Punjab (JICA (2010)). This study presents detailed traffic counts for each vehicle category at each station. For three stations, hourly records of traffic counts are provided, while for the remaining 13 stations,

traffic counts were conducted from 6:30 to 23:30.[17]

2. A study on "Traffic counts on M-9 and national highway N-5" in 2012 by the Pakistan National Transport Research Centre. (Bakht and Ahmad (2012)) This study presents detailed traffic counts for each vehicle category at each station, recorded hourly.
3. A study on "Diligence Checks at NHA Toll Plazas M-9 (Preliminary Evaluation)" in 2013 by the Pakistan National Transport Research Centre (Gulzar, Mirza and Qureshi (2013)). This study provides detailed traffic counts for each vehicle category at each station, recorded hourly.
4. Summary of traffic data from 2014 across provincial and district roads in Punjab, directed by the Government of the Punjab Communication and Works department (Govt of Punjab (2014)). This summary provides average daily traffic counts for each vehicle category at each station.
5. Data from the "Pilot Traffic Count Survey N-5 (Peshawar to Karachi)", conducted in 2017 by the Pakistan National Transport Research Centre (NTRC (2017)). This study provides comprehensive traffic counts for each vehicle category at each station, recorded hourly.
6. Phase I and Phase II summaries of the permanent traffic count in 2020 by the Pakistan National Transport Research Centre (NTRC (2020)). This study provides monthly traffic counts for each vehicle category at each station.

Figure B.2 shows the locations of traffic counts from these sources. Table B.3 shows the number of road edges for which traffic survey data from these sources was linked, as well as the average observed Annual Average Daily Traffic (AADT) for each source. The majority of observations were obtained from the 2014 Punjab provincial roads traffic surveys, which cover non-urban roads and as a result display the lowest traffic counts. Our preferred specifications exclude this source given that characteristics of vehicles traveling on non-urban roads may not provide a good approximation to those on urban roads; we also examine the robustness of results to including results from all surveys described in Table B.3.

B.3 Atmospheric Dispersion Modeling System (ADMS-Urban)

The ADMS-Urban model uses a gaussian plume dispersion approximation to calculate the concentration of pollutants at different locations and heights. Estimates are based on the emission rates of pollution sources, local meteorological conditions, chemistry mechanisms, deposition from precipitation, complex terrain and resuspension from vehicle traffic. The model can estimate how emissions from road sources contribute to pollutant concentrations at a very high spatial resolution

[17]These traffic counts were scaled to cover the whole day by analysing traffic patterns on nearby urban roads; this implied that daily traffic counts were increased by 9.3% for these observations.

Table B.3 Traffic surveys descriptive statistics

	Year	N	Avg AADT
1. Lahore Urban Transport Master Plan	2010	16	51,201.4
2. Traffic counts on M-9 and national highway N-5	2012	6	14,914.8
3. Diligence Checks at NHA Toll Plazas M-9	2013	2	14,179.5
4. Punjab provincial roads Traffic surveys	2014	632	5,910.2
5. Pilot Traffic Count Survey N-5 (Peshawar to Karachi)	2017	5	24,644.6
6. Permanent traffic counts (Phase I and II)	2020	10	13,463.3
Total		671	7,347.5

Note: N denotes number of road segments surveyed in each data source. Avg AADT denotes the Annual Average Daily Traffic counts across all type of vehicles averaged across road segments.

Vehicle (%)	All surveys	No provincial roads	Lahore only
Motorcycle	48.54	43.03	48.06
Car	24.70	28.01	25.66
LGV	9.45	9.19	9.23
Bus	2.90	2.90	2.90
RHGV 2 axles	6.77	7.52	6.55
RHGV 3 axles	2.15	3.53	2.14
AHGV 3/4 axles	4.84	4.70	4.87
AHGV 5 axles	0.31	0.47	0.28
AHGV 6 axles	0.35	0.64	0.30

Table B.4 Vehicle Shares Across Surveys

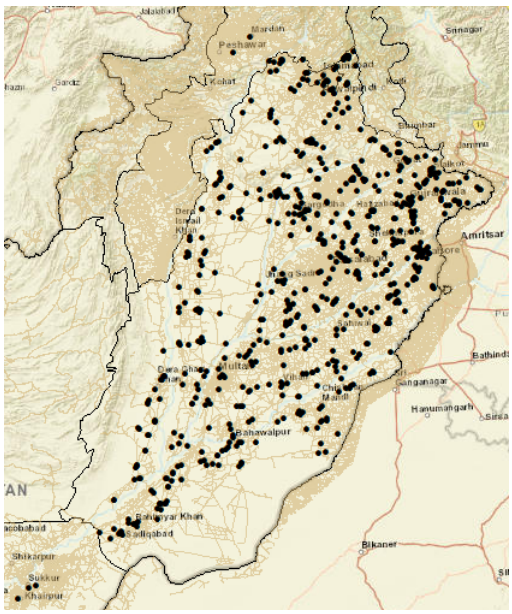


Figure B.2 Map of traffic survey stations in the area of interest

(meters) near each source at hourly intervals.[18]

[18]The model is less well suited for receptors far from the source (e.g. more than 50 km) given its use of simplified chemical mechanisms and spatially homogeneous meteorological data.

The regional model is a grid-based chemical transport model (see Simpson et al. (2012); Byun and Schere (2006)) which incorporates mechanisms that act over long spatial and temporal scales, such as long-range transport of particulates and the accumulation of concentrations in low wind speed conditions. Combining the ADMS-Urban and regional models yields detailed estimates of pollutant concentration gradients near roads, alongside lower resolution estimates with an accurate representation of transport and chemistry over large spatial and temporal scales.

We consider Punjab province as the domain of interest for the model, defining an area of approximately 750 kilometers by 800 kilometers between the coordinate points (27°N, 68°E) and (36°N, 75°E). We use the UTM 42N projection where the region of interest covers a total area of 747km by 825km and is divided into a grid of 5976 rectangular cells of dimensions 9km by 11km. The size of the cells reflect the resolution of the meteorological data (0.1°x0.1°). Each input dataset is projected to this grid.

B.3.1 Model inputs

Traffic emissions data

Emissions from each road source in the model are captured by an emission rate measured in g/km/s. This emission rate is calculated using estimates of traffic volumes in vehicles per kilometer per second (veh/km/s) together with emission factors, yielding emissions of each pollutant from the transit of a certain vehicle in grams per vehicle per second (g/veh/s).

The road network used was obtained from OpenStreetMap, filtered to include only motorway, trunk, primary, secondary, and tertiary roads.[19] In order to include the road network in the atmospheric dispersion model, the road geometry was simplified using the Douglas-Pecker algorithm with maximum tolerance of 50 meters to reduce the number of vertices to less than 50; roads which still had more vertices than the threshold after the simplifications were split into multiple segments. Road widths are approximated using 3 meters times the number of lanes. When the number of lanes was unreported this was assumed to be three lanes per direction for motorways and primary roads and two for other roads, based on a manual inspection of satellite images.

Traffic volumes on each road segment were estimated using a statistical model based on data from Pakistani traffic surveys, extended to unsurveyed roads using GPS signals from trucks and population density. This gives an estimate of the annual average daily traffic (AADTs) for all road segments for each of the nine vehicle categories considered: motorcycles, cars, buses, light goods vehicles, rigid heavy goods vehicles (RHGV) with two axles, RHGVs with three axles, articulated heavy goods vehicles (AHGV) with either three or four axles, AHGVs with five axles and AHGVs with 6 axles or above.

Emission factors for each vehicle depend on type and weight of the vehicle, fuel used, engine characteristics, road characteristics, and the speed of traffic. We differentiate between exhaust emissions and non-exhaust emissions including road wear, tyre wear, brake wear and resuspension.

[19] Link roads are also included in the network. Unclassified roads are considered as tertiary roads. Road segments with a length less than 10 meters are excluded.

Exhaust emission factors were obtained from the COPERT v5.3 database (EMP/EEA, 2019).[20] Information on vehicle characteristics, such as Euro standards, fuel used, and weight, was used to map each vehicle category in the COPERT database to a representative vehicle type in the Punjabi fleet.[21]. The average emission factor for each vehicle type was then calculated as the weighted average proportional to the prevalence of each vehicle category in the fleet. Vehicle fleet characteristics for the Punjab were obtained from consultants at Pakistan Urban Unit, as well as Shah and Zeeshan (2016); Usman et al. (2017) and Taimoor et al. (2023). Table B.5 shows euro standards and fuel characteristics which were used to calculate emission factors for each vehicle type.

Non-exhaust emission factors are also assumed to be equal to the baseline values in the COPERT database. Alternative emission factors are also considered where $PM_{2.5}$ resuspension is calibrated to Pakistani roads using measurements from Nagpure et al. (2016) since the COPERT database assumes zero resuspension for $PM_{2.5}$ which may not be applicable to Pakistani roads where road dust is an important factor.

Table B.6 shows the average emission factors used for each vehicle category and pollutant. Since these vary depending on the speed of traffic, the factors presented are calculated using a weighted average of speeds across road segments.

Table B.5 Characteristics of Vehicle fleet in the Punjab

	Motorcycle	Car	LGVs	Bus	HGV
Euro standards					
Pre-Euro 1	78%	49%	69%	100%	100%
Euro 1	0%	0%	0%	0%	0%
Euro 2	22%	34%	21%	0%	0%
Euro 3	0%	0%	0%	0%	0%
Euro 4	0%	17%	10%	0%	0%
Fuel					
Diesel	0%	7%	16%	100%	100%
LPG	0%	7%	75%	0%	0%
Petrol	100%	86%	9%	0%	0%

The time varying structure of traffic emissions is captured in fixed time-scaling factors applied to the emission rates, which vary based on variation in hour of the day, day of the week and monthly traffic captured in traffic survey data.[22] These time-varying factors, shown in Figure B.3, map AADTs to hourly-level traffic volumes for all road segments.

[20]The emission factors for LGV vehicles using LPG fuel were not available in the COPERT database, and so were calculated applying a scaling factor to the matching diesel LGV vehicle.

[21]The emission standards vehicles can reach will be limited by the available fuel. Poor adoption of higher quality fuel in Pakistan has slowed down emission reductions even as technology is improving (Tsang and Adil, 2022).

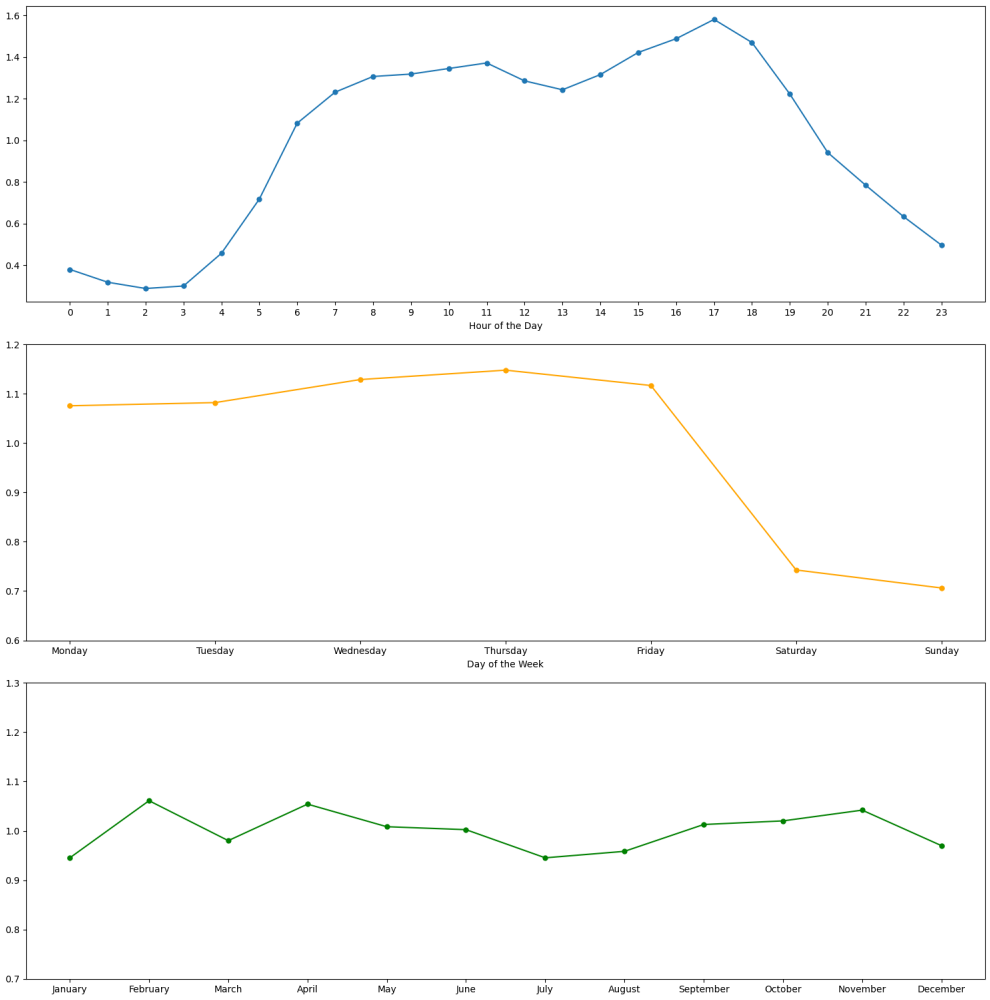
[22]Hour of the day and day of the week variation was obtained from the Pilot Traffic Count Survey on N-5 (Peshawar to Karachi) from the National Transport Research Centre of Pakistan, while monthly variation factors were obtained from data on monthly emissions from transport in Pakistan from the EDGAR emission inventory.

Table B.6 Average emission factors

Vehicle	$PM_{2.5}$ *(g/veh/km)*	PM_{10} *(g/veh/km)*	NO_2 *(g/veh/km)*	CO *(g/veh/km)*
Motorcycle	0.0018	0.0021	0.0002	0.8066
Car	0.0013	0.0021	0.0026	0.3367
LGV	0.0027	0.0037	0.0038	0.1107
Bus	0.0387	0.0491	0.0836	0.2543
RHGV 2 axles	0.0246	0.0349	0.0445	0.1637
RHGV 3 axles	0.0320	0.0423	0.0828	0.1725
AHGV 3/4 axles	0.0317	0.0424	0.0808	0.1829
AHGV 5 axles	0.0379	0.0486	0.1006	0.2042
AHGV 6 axles	0.0407	0.0514	0.1120	0.2144

Note: Emission factors calculated using COPERTv5.3 using Pakistani fleet information. Average EFs are weighted by the frequency of the speed across road segments. Euro standards ta

Figure B.3 Time scaling factors for traffic emissions



Emissions inventory

Emissions from sources other than traffic are included in the model as grid-sources, i.e. their emissions are not assigned to a specific location, but are instead assumed to be homogeneously distributed within cells of the 9kmx11km grid across Punjab. While this does not capture fine details of concentration gradients around a specific source, it accounts for large scale impacts of these emissions. We use data from the Emissions Database for Global Atmospheric Research (EDGAR), version 6.1, which provides a monthly and sector-specific emissions inventory at a resolution of 0.1°x0.1°. We use the 2018 inventory for $PM_{2.5}$, PM_{10}, NOx, CO and SO_2, and map the dataset to the model grid using bilinear interpolation.

The EDGAR inventory contains monthly data on emissions from 23 sectors. In order to specify hourly emission rates, each sector was assigned a temporal profile using time-scaling factors based on day of the week and hour of the day variation from Denier et al. (2011) after mapping each EDGAR sector to its corresponding SNAP sector. Gridded emissions from road transport are calculated using our bottom-up estimation of road emissions by aggregating all road segments' emissions within a cell; for areas without major roads, we revert to EDGAR gridded emissions. The vertical distribution of emissions was calculated by assigning each cell's emission rate proportionally to 8 possible altitude values ranging from 10m above sea level to 1134m, based on scaling factors adjusted by CERC based on Bieser et al. (2011).

Meteorological data

Meteorological data was taken from the ECWMF ERA5 reanalysis dataset, which provides hourly data based on observations from weather stations, satellite data and modelled data at a 0.1°x0.1° resolution. We use horizontal wind speed and direction at 10 m above ground level, near-surface air temperature (1-2 m), planetary boundary layer height, total cloud cover, and precipitation flux. This dataset is mapped to the grid of 9km by 11km cells across Punjab described above.

Background concentrations data

We use reanalysis data from the EAC4 model from the Copernicus Atmosperic Monitoring System (CAMS) to obtain an estimate of chemical background concentrations. This dataset provides information on $PM_{2.5}$, PM_{10}, NOx, NO_2, CO, SO_2, and O_3 concentrations at 0.75°x0.75° resolution at 3 hour intervals. We interpolate the data bilinearly to obtain an hourly dataset for each cell in the grid at 9kmx11km resolution. The reanalysis combines observations from ground-level sensors and satellite data with modelled data to provide a globally complete and consistent dataset. This dataset does not capture the detailed spatial distribution of pollution at a high resolution which may be caused by local sources, such as roads, but estimates background pollution caused by large scale mechanisms and long-range transport.

A comparison between the concentrations of $PM_{2.5}$ from this dataset and those recorded at ground-level sensors across Punjab is shown in Figure B.4. This suggests that the EAC4 data predicts high concentrations in the summer compared to what is observed on the ground. To adjust for this difference, the background concentration data is rescaled using weekly weights which

are calculated as the average ratio between CAMS and sensor data across all sensors in a 5-week interval centered on each week. Weights are smoothed using a 5-hour moving average to avoid discontinuities at the end of each week. These weights are used to scale $PM_{2.5}$ concentrations across the area of interest. The resultant scaled CAMS EAC4 data is shown in Figure B.4.

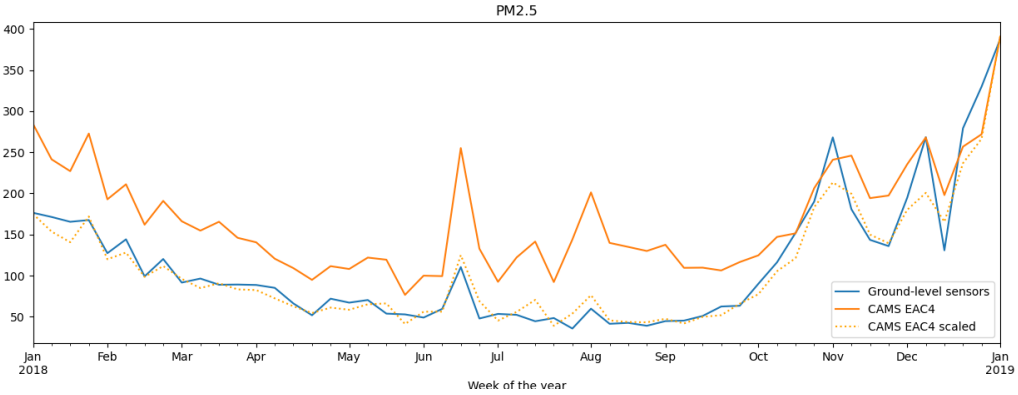


Figure B.4 Background $PM_{2.5}$ concentrations compared to ground-level sensors data

C Mathematical Appendix

C.1 Concentration share matrices

The matrices $\mathbf{M}^{y}_{ij}$, $\mathbf{M}^{\Xi_y}_{ie}$, and $\mathbf{M}^{\Xi_l}_{ie}$ capture the share of concentrations in a location i that come from a particular emissions type (industrial, goods traffic, or commuter traffic; in the superscript) in an origin (j for industrial emissions; edge e for traffic emissions).

$$\mathbf{M}^{y}_{ij} = \frac{\gamma_{ij}\mu_y y_j}{\sum_{j\in I}\gamma_{ij}\mu_y y_j + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_y}\Xi^y_e + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_l}\Xi^l_e}$$

$$\mathbf{M}^{\Xi_y}_{ie} = \frac{\gamma_{ie}\mu_{\Xi_y}\Xi^y_e}{\sum_{j\in I}\gamma_{ij}\mu_y y_j + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_y}\Xi^y_e + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_l}\Xi^l_e}$$

$$\mathbf{M}^{\Xi_l}_{ie} = \frac{\gamma_{ie}\mu_{\Xi_l}\Xi^l_e}{\sum_{j\in I}\gamma_{ij}\mu_y y_j + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_y}\Xi^y_e + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_l}\Xi^l_e}$$

C.2 Recovering commuter traffic from the commuting flow matrix

This section discusses how to recover the commuter traffic matrix Ξ^{ℓ} from the commuting flows matrix $(L_{ji})_{i,j}$ where L_{ji} is the number of residents of i that commute to j. Allen and Arkolakis (2022) discuss explicitly the inverse of this operation (in their Section 5.1).

Let $C = (c_{ji})_{i,j}$ where

$$c_{ji} = \frac{L_{ji}}{\tilde{L}_j L^0_i}$$

and $\tilde{L}_j = m^{\sigma_w} L_j$ is the number of workers that commute to j. Further, let

$$A = C^{-1} - M_1$$

where M_1 is a diagonal matrix that has i-th diagonal element

$$\frac{1}{2}\left(L_i + L_i^0\right).$$

For all the off-diagonal elements $j \neq i$ we have that

$$\Xi_{ji}^l = A_{ji},$$

the diagonal elements of the traffic matrix are

$$\Xi_{ii}^l = A_{ii} + \frac{1}{2}\left(\sum_{j \neq i} A_{ij} + \sum_{j \neq i} A_{ji}\right).$$