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@upf.edu

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

March 26, 2026

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 Allan Hsiao and Ivan Rudik for helpful comments and 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 urban Lahore using fine-resolution 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. This allows us to build on evidence that changes in road infrastructure can alter both traffic and pollution elsewhere (Bou Sleiman, 2023) by modeling general equilibrium adjustments in economic activity, trade and commuting traffic, and pollution concentrations directly.

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

[1] This contrasts with studies of how international transportation of goods may influence emissions of global pollutants such as carbon dioxide (Cristea et al., 2013; Shapiro, 2016). Bou Sleiman and Bureau (2022) also focuses on carbon emissions, developing a theoretical model to consider when decarbonization imperatives might justify removal of express roads.

road. The aggregate estimated welfare loss from removing the ring road is 3.78%, with especially pronounced losses in areas of the city directly surrounding the ring road route where market access declines markedly. These and downwind areas, however, also benefit from an associated reduction in local pollution concentrations, which offset some of the losses in market access; as a result some locations even experience net welfare increases.

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 4.35%, making clear that ignoring the impacts of local air pollution would result in a quantitatively sizable over-estimate of the welfare gains from the ring road. Accounting for pollution also leads to significant changes in the distributional impacts of the ring road: for 24% 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.

We consider how far this assessment may be altered by improved compliance with emissions standards across Pakistan's vehicle fleet. While Euro-2 emissions standards were introduced in 2009 for petrol vehicles (2012 for diesel vehicles), enforcement remains weak, with compliance at only 21-34% across different vehicle types. We perform a further policy counterfactual that simulates removal of the Lahore ring road with vehicular emissions reduced to reflect full compliance with the Euro-2 emissions standard. In this case, aggregate welfare losses are increased from 3.78% (the previous estimate under prevailing emissions standards) to 3.93%. While this is intuitively closer to the estimated welfare impacts of the ring road in the scenario without pollution, even with a substantially cleaner vehicle fleet, accounting for pollution still reduces the estimated gains from the ring road by 10%.

Finally, we simulate the net welfare impact of restricting traffic on each individual road segment throughout the city. This exercise suggests that closure would yield positive net welfare gains for around 4% of road edges in Lahore once the local pollution externalities associated with vehicle traffic are accounted for. This exercise demonstrates that the benefits from transportation infrastructure investments need not always outweigh their costs, and that careful cost-benefit analyses should account for the externality effects of investments in evaluating the impacts of such investments.

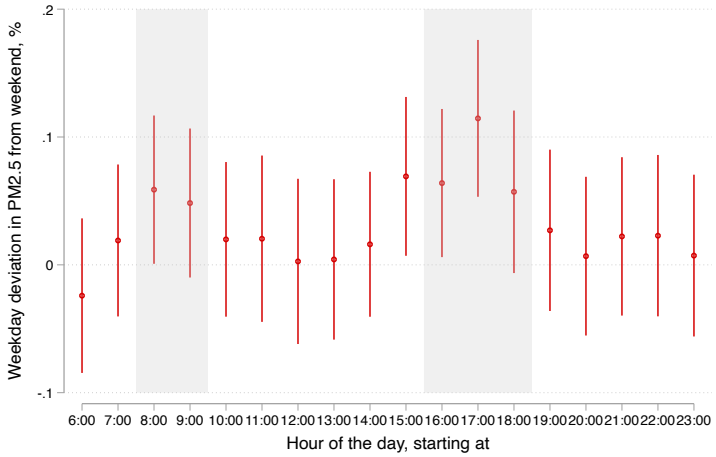
The rest of the paper is structured as follows. Section 2 presents motivating facts relating to vehicle emissions and local pollutant concentrations in Lahore, Pakistan. Section 3 presents the theoretical framework that we use to study the aggregate and distributional implications of transportation infrastructure 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 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.

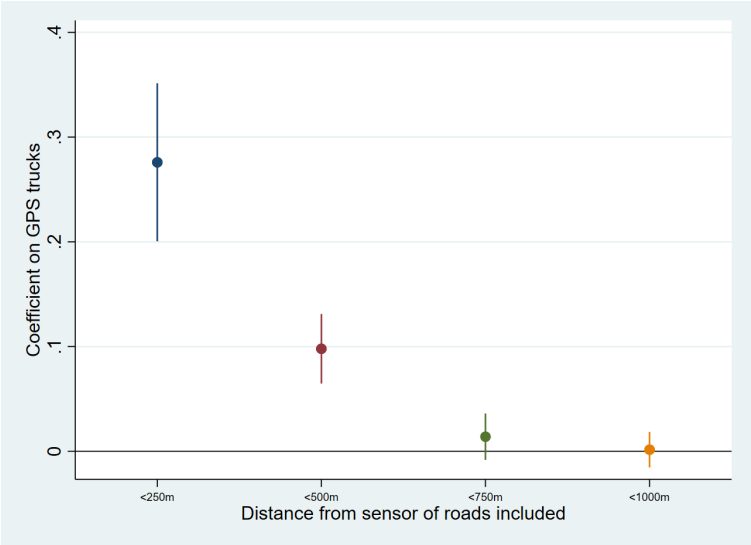
Figure 1 $PM_{2.5}$ concentrations and rush hour periods in Lahore



Notes: 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 of 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.

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 vehicles 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.5).

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



Notes: Estimated coefficient on GPS truck signals from separate regression of $PM_{2.5}$ concentrations on GPS truck signals, weekend and Ramadan dummies, and meteorological controls, where the distance within which roads are included in the traffic variable changes across regressions (eg. <250m only includes GPS signals from roads within 250m of each sensor).

Our counterfactual simulations depend crucially on how changes in traffic patterns influence local pollutant concentrations across the city. Atmospheric processes complicate the relationship 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 $0.25° \times 0.25°$. 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

Notes: 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:

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

[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 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 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:

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

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:

[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: $t_{ii} > 1$.

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 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 α:

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

Second, congestion externalities arise because the deterministic cost of crossing an edge e increases 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 (Chandra and Sikdar, 2000; Sharma and Biswas, 2021). 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. We assume that there is a background rate of pollution $\bar{m}_j$ from other sources, such as domestic emissions and agricultural burning. As a result, concentrations satisfy the relationship

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

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. NO_x and PM from traffic decrease with distance from a source due to dispersion and deposition, as well as chemical conversion of NO_x 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,

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

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

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 $\mathbf{M}_{ij}^y$, $\mathbf{M}_{ij}^{\Xi^y}$, $\mathbf{M}_{ij}^{\Xi^{\ell}}$, and $\bar{\mathbf{M}}_{\mathbf{i}}$, which capture the fraction of pollution concentrations in each location that derive from industrial, goods traffic, and commuter traffic emissions from other locations, as well as aggregate emissions from other sources[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

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

determines the change in the transportation cost component $\widehat{\tilde{t}}$ on the road segments in question.

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 Residential population

The spatial distribution of residential population comes from the Global Human Settlement Layer (Schiavina, Freire and MacManus, 2023), which provides a raster dataset of population density estimates at 100-meter resolution. We use the 2020 population layer.

4.2 Output

To generate a spatially disaggregated measure of economic output within Lahore, we combine satellite-based nighttime lights data with official GDP statistics. The VIIRS Day–Night Band Version 2.1 annual composite for 2018 (Elvidge et al., 2021) provides cloud- and stray-light–filtered radiance values at a native resolution of 15 arc seconds; we increase the spatial resolution of this data using bilinear interpolation. The units are calibrated such that the sum of luminosity within Lahore equals the district-level GDP, estimated based on Pakistan's national GDP for fiscal year 2017-18 (Pakistan Bureau of Statistics, 2025) together with the estimated 11.5% share attributed to Lahore district (Tahir, 2017).

4.3 Road network

We obtain the network of major roads (motorways, trunks, primary, secondary, and tertiary roads and their associated links) across Lahore from OpenStreetMap (2022). To simplify the network, we merge parallel lanes of multi-lane roads and remove minor geometrical details, such as roundabouts and small connecting junctions. Continuous road segments are connected, and split only at road intersections points. The resulting 2,561 road segments form the "edges" in the model.

4.4 Mobile phone signals

Information on commuting flows within Lahore is based on anonymized mobile location data from Quadrant (2024). The underlying GPS dataset covers the entirety of Pakistan for the year 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] Appendix B.1 describes the raw data and sample selection in detail.

4.5 Truck signals

We obtain data on the movement of trucks along Lahore's road network in 2018 from GPS trackers installed in 4,939 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. Signals are projected onto the closest road edge within 10 meters of the observation coordinates, dropping observations that are further from the road network. This yields 18 million edge-day-vehicle observations across Lahore during 2018.

4.6 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 this from six surveys conducted by Government Ministries and international agencies between 2010 and 2020, described at Appendix B.3. 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.

4.7 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 (Carruthers et al., 1998; Hood et al., 2018), which 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).

The model 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, described in detail in Appendix B.4. Traffic-related emissions from

[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.

each road in Lahore are estimated by combining traffic volumes (derived from the traffic surveys, truck signals and population data described above) with pollutant-specific emission factors from the COPERT v5.3 database (EMP/EEA, 2019). Emissions from non-traffic sources come from the Emissions Database for Global Atmospheric Research (EDGAR); hourly meteorological data from the ECWMF ERA5 reanalysis dataset; and background chemical concentrations from the Copernicus Atmospheric Monitoring System (CAMS).[9]

4.8 Pollution readings from ground monitors

Air quality data was obtained from 114 ground-level sensors across Punjab from 2016 to 2023. Of these, 13 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.

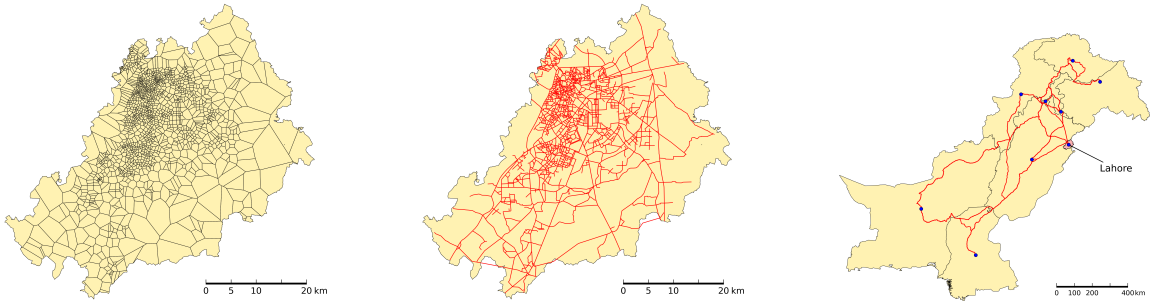
5 Model estimation

To estimate the model described in Section 3, we construct a graph that corresponds to the locations and edges in the model. The roads described in Section 4.3 form the model's (bidirectional) edges, whose endpoints and intersections form the network's vertices. Around each vertex, we construct a Voronoi cell, which partitions the spatial plane such that every point within a cell is closest to its corresponding node. This yields 1,655 Voronoi cells, which serve as the locations in the model. Figures 3a and 3b illustrate the location boundaries and edges that constitute the graph. In robustness specifications which also model goods trade with locations outside Lahore, we construct a national road network and assign average 2018 travel speeds for each road type based on the truck GPS data described in Section 4.5, computing least-cost paths between provinces (Figure 3c) 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.

[9]Total traffic-related emissions from EDGAR can be combined with data on the number of vehicles in use from OICA (2025) to yield average emissions per vehicle in Pakistan and the US in 2020. This suggest that the Pakistani fleet is 5.3 times more polluting than the US fleet in terms of $PM_{2.5}$, and 4.6 times more polluting in terms of NO_x.

Figure 3 Map of location boundaries and road network



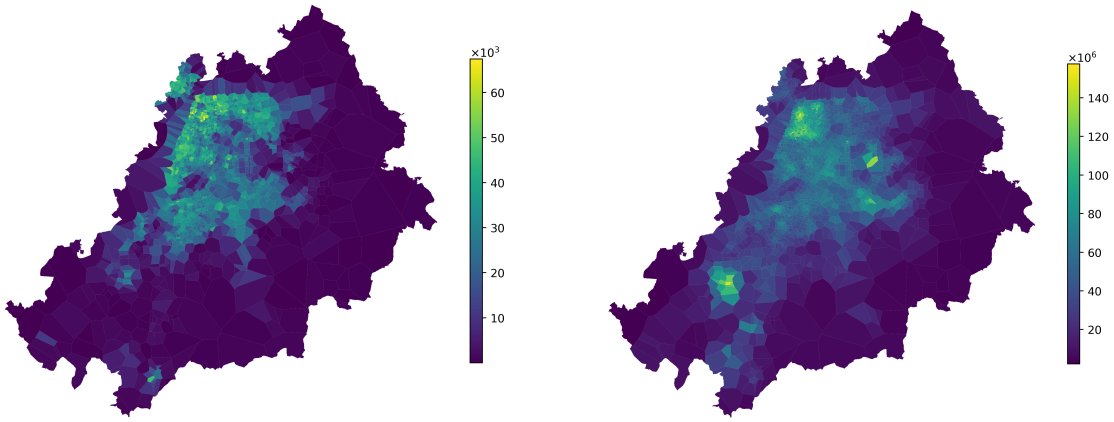
(a) Lahore locations (b) Lahore road network (c) Locations outside Lahore

Notes: The left panel shows the model locations, defined as Voronoi cells constructed around the vertices of the simplified road network. Vertices correspond to all unique road segment endpoints and intersections. The middle panel shows the edges of the graph, which are obtained by simplifying the network of major roads in Lahore. The right panel shows the locations outside Lahore and edges connecting these locations.

5.1 Location-level population and output

Residential population and output levels for each Voronoi cell are constructed by aggregating the data described in Sections 4.1 and 4.2 respectively to this level. The implied distribution of population and output density are shown in Figure 4, with summary statistics for all location- and edge-level variables in Table A.2.

Figure 4 Population and output density



(a) Population density (/km^2) (b) Output per area (USD/km^2)

Notes: The left panel shows the population density per km^2 in each location. The right panel shows the location-level output per km^2 derived nighttime lights and GDP estimate.

5.2 Bilateral commuter flows

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.

While we restrict each device to have a single home location,[10] 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, $m_i^{\sigma_w} L_i = \sum_j \omega_{ij} L_j^0$.

5.3 Commuter and goods traffic

To obtain the matrix of commuter traffic Ξ^l (goods traffic Ξ^y), we project GPS signals from cellphones (trucks) onto the road network. For each consecutive pair of signals, we compute the shortest path between the two projected locations and identify the set of edges traversed. Observations are then aggregated to obtain edge-day-level counts of cellphones (trucks), and average daily traffic for each edge computed by dividing total annual counts by 365. For edge observations with no recorded GPS activity, we assign average daily traffic equal to the first percentile of the corresponding edge traffic distribution. Since the road network is undirected, we distribute traffic equally across both directions of travel by assigning half of the edge-level average daily traffic to each direction $i \to j$ and $j \to i$. As a robustness check, we also construct asymmetric traffic matrices by separately counting GPS signals traversing each edge in each direction of travel.

We scale the GPS signals captured in our data to total vehicle counts using the ratio of average annual daily traffic (AADT) counts based on GPS signals (in both directions) to AADT counts from traffic surveys for edges where these are available.[11] This exercise is conducted separately for commuter and goods traffic and yields estimated AADT counts shown in Figure 5.

This procedure yields commuter traffic directly, under the assumption that each commuter vehicle carries one passenger. The structure of the model requires, however, that goods traffic must be scaled to be interpretable as a nominal value, in order to calculate the coefficient matrices $\Xi^y_{ji}/(Y_i + \sum_k \Xi^y_{ki})$ and $\Xi^y_{ij}/(Y_i + \sum_k \Xi^y_{ik})$ in equations (19) and (20).[12] Given that data on shipment

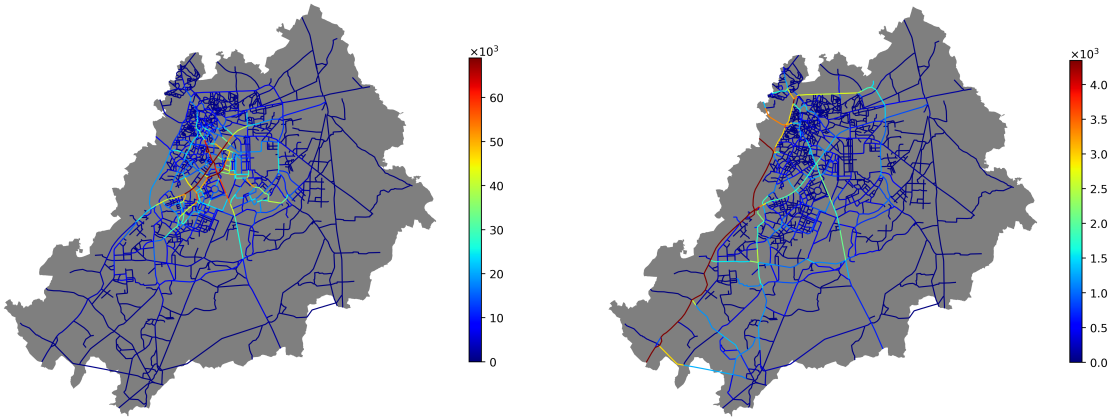
[10] Any home-work pairs associated with 'home' locations other than the primary home location are removed. See Appendix B.1.2 for more details and validation exercises.

[11] The estimated AADTs are broadly consistent with external data. For instance, Mehmood et al. (2023) report a traffic survey conducted by the Traffic Engineering and Transport Planning Agency (not used in the calibration of our scaling factors) which found a total daily traffic volume of 80,635 vehicles at the junction of Wahdat Road and Ferozepur Road, close to our estimate of 78,214 vehicles per day on an edge connected to this junction.

[12] This scaling factor is quantitatively important, because the coefficient matrices govern the degree of openness of each location, and therefore the potential for infrastructure improvements to lead to welfare gains (Arkolakis, Costinot and Rodríguez-Clare, 2012).

values is not available, we use the structure of the model, together with firm-to-firm transactions data from VAT records (Balboni, Boehm and Waseem, 2023), to estimate (and target in our calibration) the average degree of trade openness of our model locations. In particular, goods traffic on an edge $l \to k$ can be decomposed into traffic from all source-destination pairs (i, j), $\Xi^y_{kl} = \sum_{i,j} \pi^{kl}_{ji} X_{ji}$, where X_{ji} is total sales from i to j and $\pi^{kl}_{ji} = \tau^{\theta}_{ji} \tau^{-\theta}_{li} t^{-\theta}_{kl} \tau^{-\theta}_{jl}$ is the average number of times a seller from i to j will cross the edge (k, l). We observe the matrix X_{ji} as the sum of all sales transactions from i to j in the firm-to-firm transactions data. Calibrating t_{kl} as one plus the travel time from l to k yields expressions for π^{kl}_{ji} and therefore model-consistent traffic values Ξ^y_{kl}.[13] We use this to calibrate the average value of a GPS truck shipment such that $Y_i/(Y_i + \sum_k \Xi^k_{ki})$ matches the empirical share for the average location in Lahore that reports positive sales in the firm-to-firm transactions data, which is 15.7%.

Figure 5 Edge-level traffic (Ξ^l and Ξ^y)



(a) Commuter (b) Goods

Notes: This figure shows maps of commuter (left) and goods (right) annual average daily traffic (AADT) derived from cellphone and truck GPS records. Each edge is colored according to the sum of rescaled AADT across both directions.

5.4 Source apportionment matrices

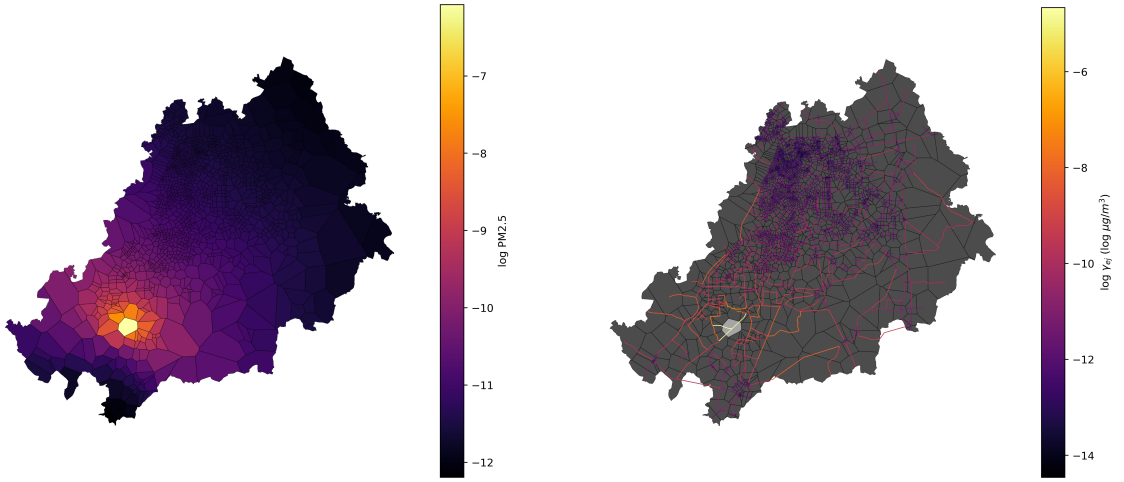
Source apportionment matrices $\mathbf{M}^y_{ij}, \mathbf{M}^{\Xi_y}_{ij}$ and $\mathbf{M}^{\Xi_l}_{ij}$ capture the share of pollution concentrations in each location i that can be attributed to industrial emissions, goods traffic, or commuter traffic, respectively, from location or edge j. We calculate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

First, the pollution transport matrices γ_{ij} and γ_{ej} are estimated using the ADMS Urban model. This models the short-range transport of pollution from an emission source to a concentration

[13] The travel time between locations in minutes is calculated using the following procedure. First, truck GPS pings are projected onto the road network, and the shortest path between each pair of consecutive pings computed, together with travel distance and time. Average travel speeds on each edge are then estimated using observed truck speeds. Finally, an origin–destination cost matrix is constructed based on the shortest path between all location pairs using Dijkstra's algorithm.

receptor. 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. The concentration receptors are spread-out in a non-regular grid spaced each 500m from the edges.[14] We use the source apportionment tool in ADMS to determine the additional concentration from emissions of $1g/s$ from each source at each receptor, then average the output from all receptors within each cell using population weights. Details on the setting of the model are discussed in Appendix B.5. Figure 6 shows an example of the output of γ_{ji} and γ_{je} for a given location, displaying the concentrations attributed to emissions from neighboring areas and neighbouring road edges respectively.

Figure 6 Example of a row γ_{ji} and γ_{je}



(a) γ_{ji} (b) γ_{je}

Notes: This figure shows an example of γ_{ji} and γ_{je} for a cell j. These can be interpreted as the marginal increase of pollutant concentrations in cell j which are attributable to one unit of industry emissions in cell i or one unit of traffic emissions on edge e. These are calculated using the source apportionment option of the ADMS model with cells modelled as areas with homogeneous emissions, receptors placed at intervals of 500m from edges and using 2018 meteorological data. γ_{ij},γ_{ej} are calculated as the average concentration in each cell j across all receptors in j.

Second, we use estimates of the share of $PM_{2.5}$ concentrations attributable to traffic and industry to calibrate $\mu_y, \mu_{\Xi_y}, \mu_{\Xi_l}$. Emissions from transport are assumed to account for a total of 24% of population weighted $PM_{2.5}$ concentrations, and industrial emissions 16%, in line with estimates from World Bank (2025) for greater Lahore[15]. We assume that the emission factors for traffic

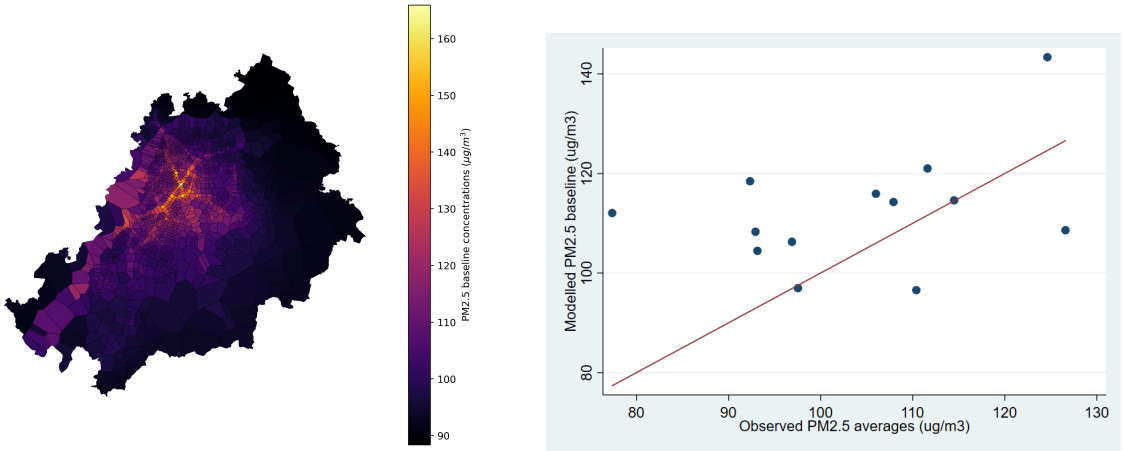
[14] We use a non-regular grid to avoid receptors being placed at different distances from edges based on orientation, since distance to the edge significantly affects the concentration from road emissions. Additional receptors are also placed at 5 and 50 meters from the road edges to better capture the pollution gradient near roads.

[15] While estimates vary across sources, these shares are broadly in line with other estimates in Lahore (Ilyas and Nissar, 2023, EDGAR, 2025) and other South Asian cities (Jalan and Dholakia, 2019; CSIR - National Environmental Engineering Research Institute and Indian Institute of Technology Bombay, 2023; Systemiq and Bandung Institute of Technology, 2025). This calibration is also consistent with the reduced-form estimates that $PM_{2.5}$ concentrations on weekdays rise by around 10% during rush-hours relative to weekends. Alternative calibrations that match $PM_{2.5}$ concentration patterns across Lahore less well are also considered.

vary by type of vehicle and speed in line with data from the COPERTv5.3 database. This dataset determines the emission factors of a AADT unit of commuting traffic relative to goods traffic, as well as the emission factors of traffic at different speeds.[16] Details are provided in Appendix B.5.

Third, the background concentration $\bar{m}$ is then calibrated to match the observed median concentration at the US Lahore consulate (the only sensor for which we observe a full year of observations before the Covid-19 pandemic) of 85.1 $\mu g/m^3$. The background concentration reflects emissions that originate from external or unmodeled sources, such as dust storms and agricultural burning, and is assumed to be homogeneous over the area of interest.

Figure 7 Out-of-sample validation of pollution model



(a) Modeled concentration

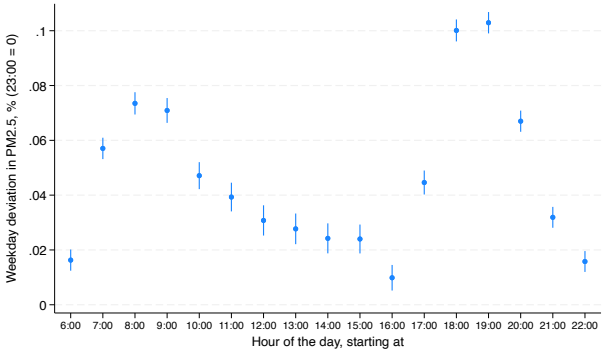
(b) Model vs. observed

Notes: The left panel illustrates the spatial distribution of modeled yearly average $PM_{2.5}$ concentration ($\mu g/m^3$) across cells from industrial production, traffic and background concentrations. The right panel shows the scatter plot of cross-sectional distribution of $PM_{2.5}$ concentrations modeled at baseline against concentration measured at ground-level sensors.

Figure 7a maps modeled $PM_{2.5}$ concentrations from production, traffic and background concentrations. Concentration magnitudes are in line with observed averages, with a sensible spatial distribution where production centers around the central business district and major roads are visible. Figure 7b demonstrates a positive correlation between average $PM_{2.5}$ concentrations measured at the 13 ground station monitors for which we have data available (which are not used in calibration of the model) and modeled concentrations at these locations. Summary statistics for the pollution source apportionment matrices are included in Appendix Table A.3: these reveal that on average close to half of the concentration shares attributed to industrial or traffic pollution in each location are from cells or edges within 1km. Appendix Figure A.1 shows the average marginal increase in $PM_{2.5}$ concentrations from an average level of production (traffic) across cells (edges) grouped by distance, and demonstrates rapid spatial decay in each case, with effects nearing zero

[16] Traffic is assumed to move at different speeds depending on the road edge it traverses: on road edges where traffic is faster, fuel consumption is more efficient and so emission factors tend to be lower. The traffic speed for each road edge is estimated using the trucks' GPS data.

Figure 8 Simulated diurnal profile of weekday PM2.5 concentrations



by 5km.

A more demanding validation exercise considers how well the modeled data matches our motivating descriptive that pollution concentrations during the morning and evening rush hours are elevated by 5-10%. To examine this, we enter average weekday traffic volumes observed in the traffic surveys into the ADMS model and simulate the resultant percentage deviations in $PM_{2.5}$ concentrations at the same 13 pollution monitoring stations that were used in the regressions underlying Figure 1. Figure 8 shows that the model predicts concentration increases during rush hours of between 5% and 10%, which match closely the estimates from the ground-level monitors.

5.5 Calibrated model parameters

The trade elasticity is calibrated to $\theta = 6.83$, the congestion elasticity in traffic to $\lambda = 0.071$, 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 (Nair et al., 2021). 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 $\sigma_h = 0.36$. This is at the lower end of the 0.34-0.42 range

for the estimated elasticity of willingness to pay with respect to air quality based on household location decisions in the USA from 1990-2000 in Bayer, Keohane and Timmins (2009).[17]

The elasticity of earnings with respect to concentrations, σ_w, is calibrated based on an estimated value of $\sigma_w = 0.052$ in India from Adhvaryu, Kala and Nyshadham (2022). We consider the robustness of results to a wide range of alternative values estimated in other settings, ranging from $\sigma_w = 0.0023$ to $\sigma_w = 0.44$, in Section 6.1.2.

6 Counterfactual simulations

The model is used to estimate the welfare impacts of major realized and counterfactual road investment projects in Lahore, accounting for both market integration gains and the local costs of air pollution externalities. We first consider how accounting for local pollution impacts alters welfare assessments of the Lahore ring road project, a major recent infrastructure project in the city. A second set of counterfactual simulations then considers whether there are road segments in the network for which pollution externalities are sufficiently strong as to result in the road having a net *negative* impact on welfare.

6.1 Does pollution change the welfare assessment of road projects?

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 9, 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).[18] This has resulted in significant shifts in the location of traffic flows throughout the city. An example is shown in Figure 10, which shows traffic flows in the GPS truck location data described in Section 4.5 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.

[17] 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 much lower value of $\sigma_h = 0.0047$ based on an estimated marginal willingness to pay of \$1.14 for a 10 $\mu g/m^3$ reduction in annual $PM_{2.5}$ concentrations in Baylis et al. (2024). We also consider robustness to a higher elasticity of $\sigma_h = 1.16$ in line with estimates from China in Freeman et al. (2019).

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

Figure 9 Lahore ring road

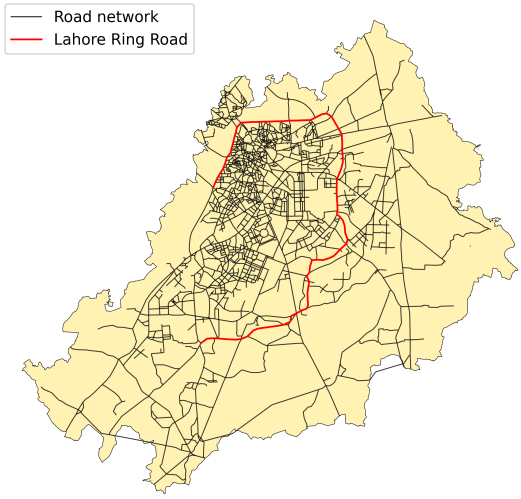
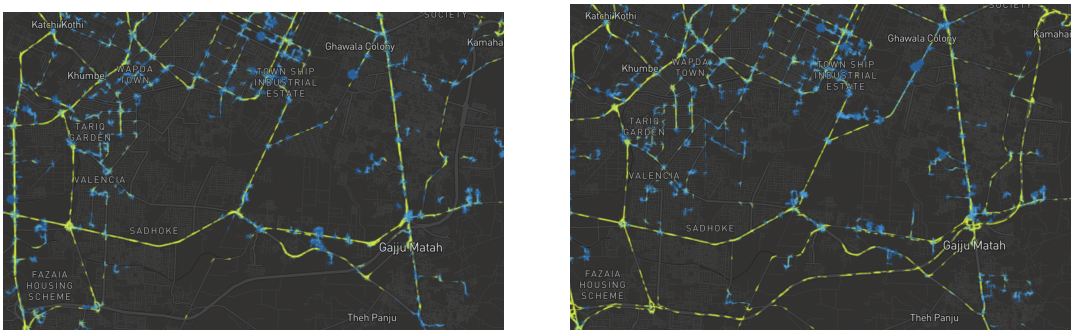


Figure 10 GPS signals before and after opening of the southern section of the Lahore ring road



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

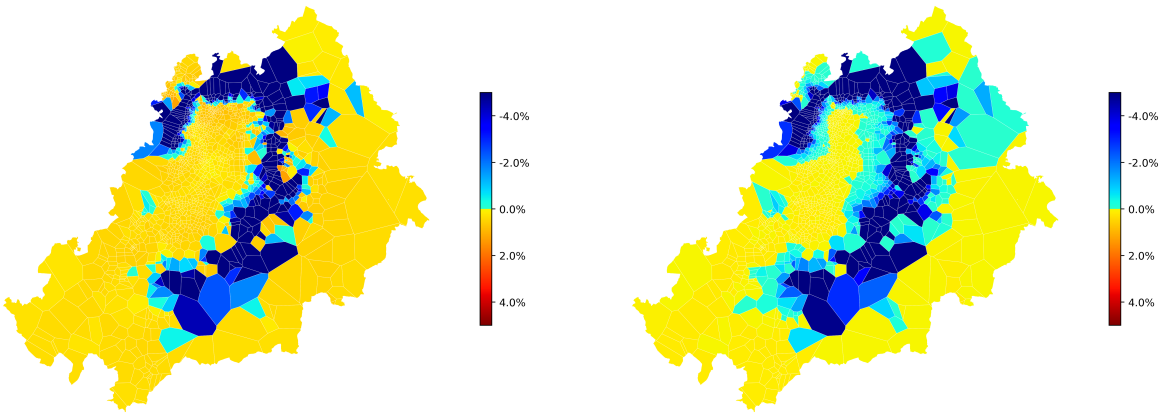
Notes: The figure shows the location of GPS signals from truck location data before (panel (a)) and after (panel (b)) the opening of the southern section of the Lahore ring road on 22 December 2017.

Estimation of the full model suggests that removing the Northern and Southern loops of the Lahore ring road[19] would result in an aggregate welfare loss of 3.78%, broadly in line with estimates of the welfare impacts from other major urban transport infrastructure improvements in developing country cities (e.g., Tsivanidis, 2026; 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 11a. Welfare losses from removing the ring road (shown in blue) 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. There are, however, regions that experience gains, for instance in eastern areas of the city where removal of the ring road lowers incoming air pollution.

[19]'Removal' of roads is implemented in the model estimation as a 200% increase in transportation costs along corresponding road edges.

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 pollution channel in influencing the welfare impacts of the road, we solve the model (again in changes) but shut down any role for pollution affecting welfare or economic activity ($\sigma_h = \sigma_w = 0$).

Figure 11 Welfare changes from removing the Lahore ring road



(a) Overall changes

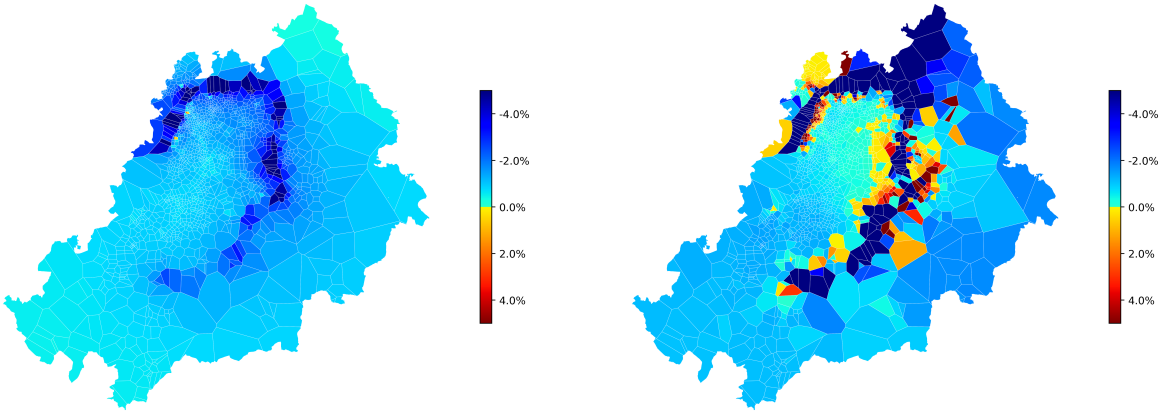
(b) Excluding pollution

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 population-weighted average welfare losses of 4.35% from removing the ring road, the induced changes in pollution damages reduce these losses to 3.78%, a 13% reduction in losses. Abstracting from pollution impacts also results in a shift in the spatial distribution of welfare changes from removing the ring road: as shown in Figure 11b, welfare losses from removal of the ring road are in this case much more concentrated along the ring road route. Trade and commuting cost changes alone would suggest negative welfare impacts from removing the ring road in 53% of locations, while in the scenario accounting for pollution impacts this share falls to 29%. In 24% of locations, accounting for pollution changes our assessment of whether removal of the ring road results in a net increase or decrease in local welfare.

Studying the changes in pollution concentrations, production patterns and traffic induced by removal of the ring road helps to understand the spatial incidence of gains and losses. Reductions in concentrations (Figure 12a) are especially pronounced around the ring road and in downwind eastern areas. Locations on the ring road see reductions in industrial output (Figure 12b) due to worsening access to workers; however, adjacent areas see output increases driven by workers substituting nearby and productivity improvements induced by lower pollution concentrations. Changes in commuter and goods traffic are shown in Figures 13a and 13b respectively. While goods traffic declines everywhere, but most especially close to the ring road route, there is significant

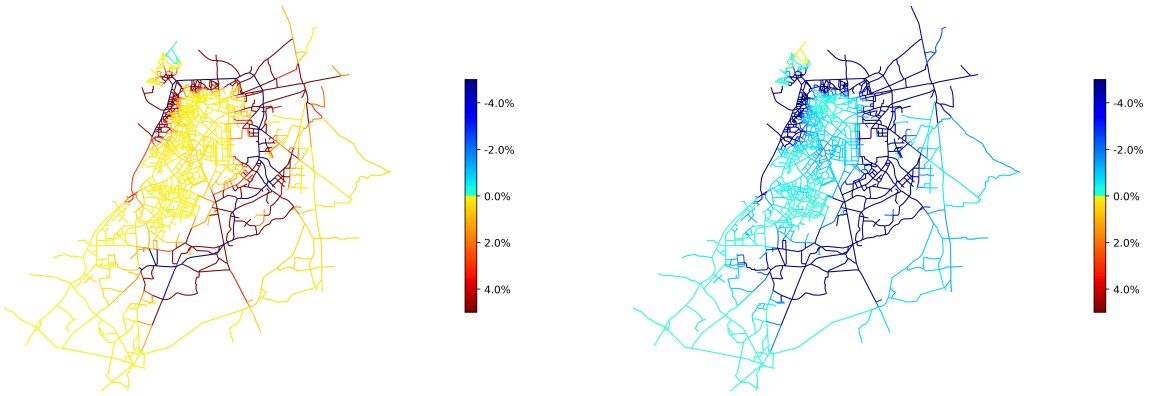
substitution of commuter traffic away from the ring road route to neighboring road edges.

Figure 12 Changes in pollution and production from removing the Lahore ring road



(a) Pollution concentrations (b) Production

Figure 13 Changes in commuter and goods traffic from removing the Lahore ring road



(a) Commuter traffic (b) Goods traffic

Taken together, 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 on traffic, production, pollution and welfare, wrongly concluding that there are benefits in some regions that actually experience substantial welfare losses from the new infrastructure.

6.1.1 Impacts with improved enforcement of vehicle emissions standards

The results in this section suggest that accounting for pollution substantially attenuates the estimated gains from road investments in this setting. Given that the vehicle fleet in Pakistan is characterized by low compliance with emissions standards that are routinely enforced in other countries, this may not be the case in contexts with a cleaner vehicle fleet. We investigate this by re-simulating the removal of the Lahore ring road incorporating improvements in compliance with vehicle emissions standards.

Euro-2 vehicle emission standards[20] were introduced in Pakistan in 2009 for petrol vehicles and in 2012 for diesel vehicles. While all newly manufactured vehicles were subject to this regulation, older vehicles with less efficient combustion systems remained on the road (Shah and Zeeshan, 2016). In 2020, the standard was raised to Euro-5, but again with weak enforcement (Tsang and Adil, 2022). Euro-2 compliance remains low across all vehicle types in Punjab's fleet: only 22% of motorcycles, 34% of cars, and 21% of LGVs meet Euro-2 standards (see Appendix Table B.6). In response to severe air pollution, national and local governments (especially in Punjab) are making efforts to enforce environmental standards, and vehicle emissions testing is ongoing in Lahore as of June 2025. The counterfactual we implement lowers emissions factors to those that would prevail were the entire vehicle fleet to comply with the Euro-2 standard.[21]

The results of this counterfactual simulation reveal that enforcement of Euro-2 standards would alter the welfare evaluation of removing the ring road. Crucially, however, accounting for pollution still has a substantial influence on assessment of gains from the ring road even with stricter environmental compliance. The estimated aggregate welfare loss from removing the ring road in this case is 3.93%, 10% lower than estimated losses when abstracting from the impacts of pollution. The spatial heterogeneity of welfare impacts, shown in Appendix Figure A.2, is very similar to that in the baseline scenario shown in Figure 11a. While this difference is smaller than the 13% reduction in losses in the scenario with prevailing environmental standards, this suggests that welfare estimates of road improvement projects may still be substantially altered by accounting for pollution impacts even with much more stringent environmental enforcement.

6.1.2 Robustness specifications

We consider the robustness of the central findings in this section to alternative modeling assumptions and parameterizations, summarized in Table 2.

Endogenous pollution assumptions. The central simulations calibrate the population-weighted share of pollution concentrations from transport to be 24% and the share from industrial emissions to be 16%, with the remaining 60% attributed to fixed background sources. We consider the robustness of the results to assuming that these endogenous pollution sources captured by the model

[20] Khan and Yasmin (2014) document that Pakistan's emission standards (Pak-2) for passenger cars specify g/km limits for CO, HC+NO_x, and PM that are almost identical to Euro-2 standards for the same vehicle categories. The only difference is a slightly higher HC+NO_x limit for diesel vehicles under Pak-2 (0.7, compared to 0.6 under Euro-2).

[21] Under this counterfactual, emission factors are reduced on average by 65% for goods carrying vehicles and by 72% for passenger-carrying vehicles relative to our baseline specification.

account for a smaller share of overall emissions, by reducing the share of emissions from transport and industry to 10% each, with background sources accounting for 80%. In this counterfactual scenario, the implied welfare loss from shutting the ring road is larger than in the central scenario, but this remains substantially smaller than the 4.35% welfare loss from shutting the ring road when abstracting from pollution impacts.

Elasticity of household utility with respect to pollution. As discussed in Section 5.5, we use a central estimate of $\sigma_h = 0.36$ for the elasticity of utility with respect to pollution concentrations. We consider the robustness of results to alternative parameterizations at the high and low end of available estimates in the literature, in particular $\sigma_h = 0.0047$ based on observed purchases of defensive technologies in India from Baylis et al. (2024), and $\sigma_h = 1.16$ based on estimates using a model of household residential location decisions in China from Freeman et al. (2019). Qualitative conclusions remain robust. Estimated welfare losses from shutting the ring road are intuitively closer to the case with no pollution with a lower elasticity of utility with respect to pollution, and further when this elasticity is higher.

Elasticity of worker productivity with respect to pollution. The central estimate used for the elasticity of productivity with respect to pollution concentrations is $\sigma_w = 0.052$, based on evidence from Adhvaryu, Kala and Nyshadham (2022) in India. Estimates of this parameter in China (Chang et al., 2019; Fu, Viard and Zhang, 2021; He, Liu and Salvo, 2019) and the USA (Chang et al., 2016) span a wide range of values; we show the robustness of our results to the minimum ($\sigma_w = 0.0023$ from Chang et al. (2019)) and maximum ($\sigma_w = 0.44$ from Fu, Viard and Zhang (2021)) values from these studies.

Goods trade outside Lahore. The central simulations consider commuting, goods trade and pollution transport across regions within the city of Lahore. While commuting and pollution transport are typically highly localized, a significant share of goods trade may occur with locations outside Lahore: for instance, the VAT transactions data in Balboni, Boehm and Waseem (2023) reveal that 61% of sales from firms in Lahore are to firms in locations outside the city. We re-estimate the model including an additional eight locations to capture Pakistan's provinces, where Lahore is removed from the data aggregates of Punjab province. We construct an analogous transport network for the entirety of Pakistan using the same methodology described in Section 4.3 for Lahore, but applied to the nationwide major road network. Each province is represented by its centroid, projected onto the nearest edge in the road network. We then compute least-cost routes between all province pairs using Dijkstra's algorithm, where edge costs are defined by travel times based on road-type-specific speeds estimated from the truck GPS data described in Section 4.5. Finally, we estimate goods traffic between provinces using the truck GPS data and the same procedure and scaling factor as used for locations within Lahore. Welfare losses accruing to Lahore's residents from removal of the ring road display the same pattern as in the central estimates: in this case, the estimated welfare losses from removing the ring road are 35% lower when accounting for pollution (relative to 13% lower in the central scenario).

Trade and congestion elasticities. The trade and congestion elasticities in the central scenario

are calibrated based on those used in the urban commuting model of Allen and Arkolakis (2022). We test the sensitivity of results to instead using the elasticities from the economic geography model in Allen and Arkolakis (2022), as well as to alternative parameterizations investigated in robustness specifications in the same paper. The results remain qualitatively robust.

Smoothing zeros in the commuting matrix. The bilateral commuting matrix in the central scenario described in Section 5.2 features a significant share of origin-destination pairs which have zero commuters. We test the sensitivity of results if we replace these zeros with a small value.[22] Very similar results obtain in this case.

Exposure to pollution while commuting. The baseline model incorporates damages from pollution exposure as a result of pollution at the location of residence reducing utility and pollution at the workplace reducing productivity. We extend the model to incorporate additional utility costs from pollution experienced on workers' commutes, as described in Appendix C.3. This implementation relies on two new inputs. First, we calculate the time taken to traverse each edge in the road network dividing edge length by edge-specific average traffic speeds derived from truck GPS data. We then use this to calculate the time individuals spend on the commute versus at home. Second, we calibrate the parameter capturing relative exposure to pollution while commuting compared to while being at home to $\delta = 1.2$.[23] Results are qualitatively robust in this scenario, with an intuitive increase in the negative welfare impacts of pollution from the ring road when accounting for this additional channel for damages.

[22] To construct this small value, for each origin cell, we compute the difference between the origin cell population and the number of commuters observed in the raw cellphone data. We then identify the origin cell with the smallest such difference and divide this value by the number of destinations with zero observed commuters from that origin. This defines a uniform small value, which we use to replace all zero entries in the commuting matrix. We proportionally rescale the non-zero entries in each origin so that total commuters equal the origin cell population.

[23] This calibration is based on evidence from Dhaka that paired indoor-outdoor particulate pollution measures show ratios close to 1 (Saha et al., 2025), and that public buses and cars with open windows experience concentrations 20% higher than urban background locations (Saha et al., 2024).

Table 2 Robustness tests: Counterfactual removing ring road

	Robustness test	Welfare loss (pollution)	Welfare loss (no pollution)
(1)	Central scenario	3.78%	4.35%
(2)	Conservative endogenous pollution	3.91%	4.35%
(3)	$\sigma_h = 0.0047$	4.31%	4.35%
(4)	$\sigma_h = 1.16$	2.61%	4.35%
(5)	$\sigma_w = 0.0023$	3.84%	4.35%
(6)	$\sigma_w = 0.44$	3.31%	4.35%
(7)	Including trade outside Lahore	2.29%	3.50%
(8)	$\theta = 8,\ \lambda = 0.092$	3.16%	3.61%
(9)	$\theta = 4$	6.28%	6.92%
(10)	No zeros in commuting matrix	3.78%	4.35%
(11)	On-route pollution exposure	2.28%	4.35%

6.2 Are there roads that have a net negative impact on welfare?

The results in the previous subsections suggest that accounting for pollution attenuates estimates of welfare gains from the ring road. This raises the possibility that, for other roads, the welfare costs of negative pollution externalities may be sufficiently strong as to entirely offset the gains from market access improvements. To investigate this, we conduct a counterfactual simulation exercise which aims to estimate the aggregate population-weighted change in welfare when each road segment in Lahore's road network is removed.

Figure 14 colors the roads whose removal would result in an aggregate welfare decrease in blue, and roads where aggregate welfare would increase in red. These simulations suggest that 4% of road edges have *negative* aggregate welfare impacts once the impact they have on pollution is accounted for. Many of these edges are peripheral roads on the outskirts of the city: these are associated with less pronounced trade and commuting gains than roads connecting outlying areas to the center, but still having negative impacts on local residents via increased pollution concentrations. Some smaller roads in the densely populated urban center, for which alternative substitute routes are typically available, are also associated with net welfare losses.

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.

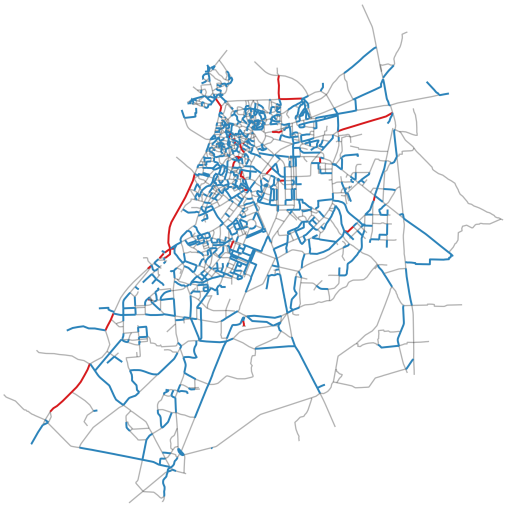


Figure 14 Net welfare gains from road removal

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 100-fold increase in transportation costs). All other roads are colored in green.

7 Conclusions

Rapid advances in our understanding of the magnitude and damages from local air pollution have 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.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New trade models, same old gains?" The American Economic Review, 102(1): 94–130.

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, and Joseph S. Shapiro. 2025. "Spatial Environmental Economics." Handbook of Regional and Urban Economics, 6: 585–652.

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.

Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins. 2009. "Migration and hedonic valuation: The case of air quality." Journal of Environmental Economics and Management, 58(1): 1–14.

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

Bou Sleiman, Léa. 2023. "Displacing congestion: Evidence from Paris." CEPREMAP.

Bou Sleiman, Léa, and Dominique Bureau. 2022. "When should urban roads be removed to lower carbon emissions?"

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

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.

Chandra, Satish, and PK Sikdar. 2000. "Factors affecting PCU in mixed traffic situations on urban roads." Road and Transport Research, 9(3): 40–50.

Chang, Tom, Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2016. "Particulate pollution and the productivity of pear packers." American Economic Journal: Economic Policy, 8(3): 141–169.

Chang, Tom Y, Joshua Graff Zivin, Tal Gross, and Matthew Neidell. 2019. "The effect of pollution on worker productivity: evidence from call center workers in China." American Economic Journal: Applied Economics, 11(1): 151–172.

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.

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.

CSIR - National Environmental Engineering Research Institute and Indian Institute of Technology Bombay. 2023. "Final Report: Air Quality Assessment, Emission Inventory & Source Apportionment Study for Mumbai." CSIR - National Environmental Engineering Research Institute and Indian Institute of Technology Bombay.

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.

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.

EDGAR. 2025. "EDGAR - Emissions Database for Global Atmospheric Research." https://edgar.jrc.ec.europa.eu/dataset_ap81.

Elvidge, Christopher D, Mikhail Zhizhin, Tilottama Ghosh, Feng-Chi Hsu, and Jay Taneja. 2021. "Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019." Remote Sensing, 13(5): 922.

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.

Fu, Shihe, V Brian Viard, and Peng Zhang. 2021. "Air pollution and manufacturing firm productivity: Nationwide estimates for China." The Economic Journal, 131(640): 3241–3273.

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.

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.

He, Jiaxiu, Haoming Liu, and Alberto Salvo. 2019. "Severe air pollution and labor productivity: Evidence from industrial towns in China." American Economic Journal: Applied Economics, 11(1): 173–201.

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, Christina, Ian MacKenzie, Jenny Stocker, Kate Johnson, David Carruthers, Massimo Vieno, and Ruth Doherty. 2018. "Air quality simulations for London using a coupled regional-to-local modelling system." Atmospheric Chemistry and Physics, 18(15): 11221–11245.

Ilyas, Hassan, and Hira Nissar. 2023. "Sectoral emissions inventory of Lahore." Urban Unit.

Jalan, Ishita, and Hem H. Dholakia. 2019. "What is Polluting Delhi's Air? Understanding Uncertainties in Emissions Inventories." Council on Energy, Environment and Water.

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. Brookings Institution Press.

Khan, Muhammad Imran, and Tabassum Yasmin. 2014. "Development of natural gas as a vehicular fuel in Pakistan: Issues and prospects." Journal of Natural Gas Science and Engineering, 17: 99–109.

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.

Mehmood, Zahid, Ali Hasnain, Muhammad Luqman, Sohaib Muhammad, Narayan Babu Dhital, Arooba John, Maryam Iqbal, Amna Ejaz, Matiba Tufail, Harma, et al. 2023. "Assessment of Air Pollution Tolerance and Physicochemical Alterations of Alstonia scholaris (L.) R. Br. along Roadsides of Lahore, Pakistan." Aerosol and Air Quality Research, 23(7): 230038.

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.

OICA. 2025. "Vehicles in Use: World Vehicles in Use by Country/Region and Type (Passengers Cars, Commercial Vehicles, All Vehicles)." https://www.oica.net/category/vehicles-in-use/.

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.

Pakistan Bureau of Statistics. 2025. "Gross Domestic Product of Pakistan (at current basic prices of 2015-16)." https://www.pbs.gov.pk/wp-content/uploads/2020/07/Table_4-3.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.

Saha, Provat K, Ahsan Habib, Dipika R Prapti, Talha Jubair, Abu U Zarrah, Chowdhury A Hossain, Sheikh M Rahman, Abdus Salam, Md Aynul Bari, and Julian D Marshall. 2025. "Characterizing indoor-outdoor PM2. 5 concentrations using low-cost sensor measurements in residential homes in Dhaka, Bangladesh." Atmospheric Environment, 342: 120945.

Saha, Provat K, Shaikh Ashik-Un-Noor, Allen L Robinson, and Albert A Presto. 2024. "In-vehicle ultrafine and fine particulate matter exposures during commuting in a South Asian megacity: Dhaka, Bangladesh." Atmospheric Environment, 321: 120340.

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.

Schiavina, Marcello, Sergio Freire, and Kytt MacManus. 2023. "GHS-POP R2023A-GHS population grid multitemporal (1975–2030)." Joint Research Centre Data Catalogue.

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.

Sharma, Manjul, and Subhadip Biswas. 2021. "Estimation of Passenger Car Unit on urban roads: A literature review." International Journal of Transportation Science and Technology, 10(3): 283–298.

Systemiq and Bandung Institute of Technology. 2025. "Better air, better Indonesia: The Economic and Political Case for Urgent and Coordinated Action for Jakarta's Clean Air." Systemiq and Bandung Institute of Technology.

Tahir, Nadia. 2017. "The size and growth of the economy of Lahore." Directorate of Research and Development.

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. 2026. "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio." American Economic Review, 116: 418–463.

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.

World Bank. 2025. "Punjab Clean Air Program (P508222): Program Information Document (PIDPA00180)." World Bank, Washington, DC.

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

A Additional Tables and Figures

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

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

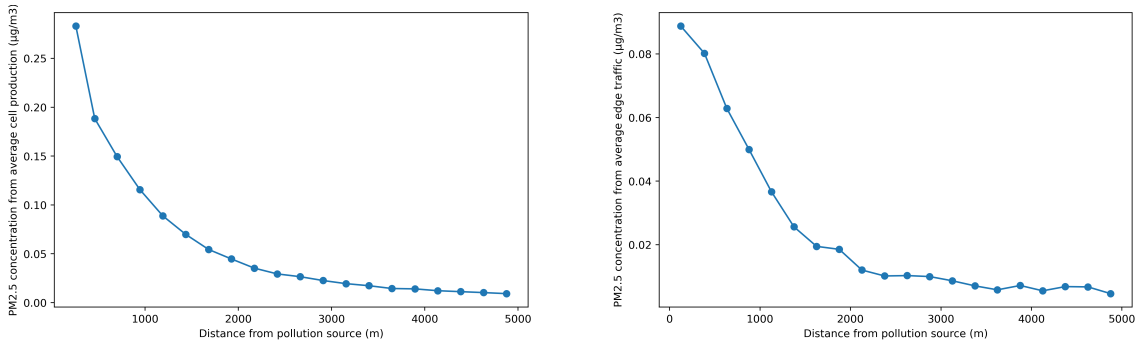
Notes: Traffic shares are calculated from the Lahore Urban Transport Master Plan JICA (2010) averaged across all survey locations. PCE values are taken from the Highway Capacity Manual as reported in Farooq and Akram (2018).

Table A.2 Summary statistics

	Mean	SD	Min	Max
Voronoi cell level ($N = 1,655$)				
Population	7,546.08	8,003.12	1.13	88,967.37
Output (million USD)	22.92	27.49	0.34	245.03
Share commuting outside	0.37	0.13	0.00	0.94
Average commuting distance (km)	2.71	1.73	0.02	15.68
Edge level ($N = 2,561$)				
Average daily commuters	7,488.72	12,671.19	5.09	97,073.96
Average daily trucks	368.62	939.76	0.20	9,528.22

Notes: First panel reports summary statistics of key input variables at the location level. Average commuting distance is calculated based on the distance between location centroids, while within-location commuting distance is approximated by $\frac{2}{3}\sqrt{\text{area}/\pi}$. The second panel reports summary statistics of traffic volume at the edge level, with traffic volumes aggregated across both directions for each edge.

Figure A.1 Rapid pollution decay with distance from source



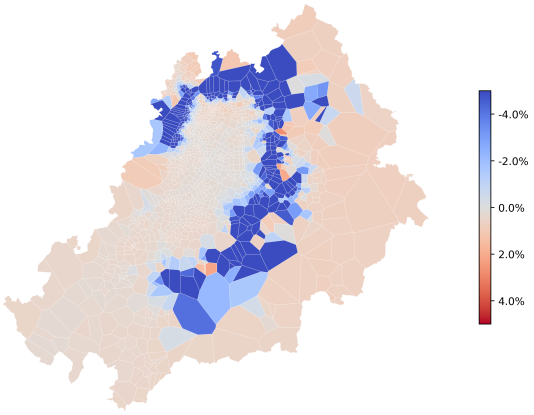
(a) Production (b) Traffic

Notes: This figure shows the average marginal increase in concentration ($\mu g/m^3$) at location j from an average level of production/traffic ($\gamma_{ij} \times \mu$) across cells/edges grouped by distance. Distance from the source is calculated from the centroid of the cell to the boundary of each cell/edge. For production these are calculated multiplying each respective γ_{je} by μ_y and the average level of production across cells $\overline{y}$, for traffic each γ_{je} is multiplied by a weighted average of $\mu_{\Xi_l} \times \overline{\Xi}_y$ and $\mu_{\Xi_l} \times \overline{\Xi}_l$ based on traffic shares.

Table A.3 Summary statistics: pollution source apportionment matrices

	Mean	SD	Min	Max
Cell-to-cell pollution ($N = 1,655 \times 1,655$)				
Share of industrial pollution: $\sum_i M_{ji}^y$				
From own cell j	3.23%	1.65%	0.05%	12.72%
From cells within 1km	8.16%	3.38%	0.95%	19.73%
From cells within 10km	14.75%	4.66%	2.64%	26.22%
From all cells	18.01%	4.99%	4.99%	28.84%
Edge-to-cell pollution ($N = 1,655 \times 2,561$)				
Share of traffic pollution: $\sum_i M_{ji}^{\overline{\Xi}_y} + M_{ji}^{\Xi_l}$				
From edges in j	6.60%	6.26%	0.00%	30.07%
From edges within 1km	10.59%	6.81%	0.01%	31.32%
From edges within 10km	18.44%	7.23%	0.05%	34.42%
From all edges	21.99%	4.99%	11.16%	35.01%

Figure A.2 Welfare changes from removing the Lahore ring road (Euro 2 standards)



B Data Appendix

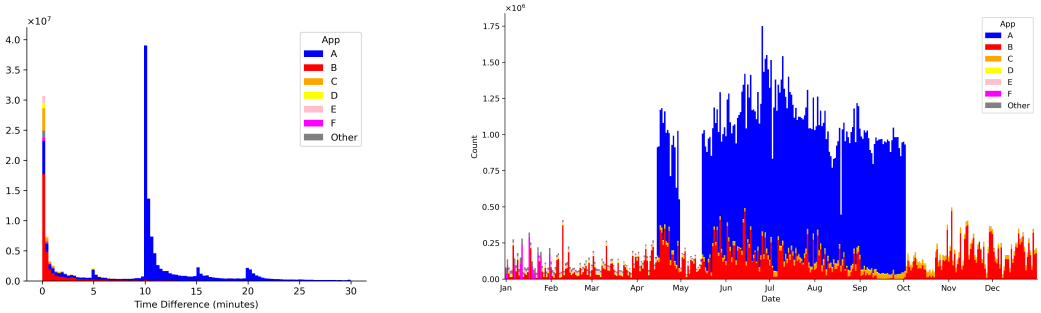
B.1 Mobile location data

This appendix describes the cellphone GPS data used to estimate commuting flows and commuter traffic in Lahore, and documents the sample construction, cleaning, and validation procedures. The truck GPS data used to estimate goods traffic are drawn from the same commercial data provider and are described in Balboni, Boehm and Waseem (2023); where our processing of the cellphone data follows similar steps, we note this explicitly below.

B.1.1 Raw data and sample construction

The raw GPS dataset is obtained under license from a commercial data provider (Quadrant, 2024) and covers all of Pakistan over the calendar year 2019. Each observation records a unique device identifier, a timestamp, a pair of geocoordinates (longitude and latitude), and a measure of horizontal accuracy. The full national dataset contains approximately 1.19 billion signals from 16.6 million unique devices. Signals originate from more than 400 distinct app IDs, though 93 per cent of observations come from just two applications. These apps differ systematically in signal frequency: one generates signals at approximately 10-minute intervals, while others transmit far more frequently (Figure B.1a). The volume of signals also varies substantially over the year (Figure B.1b), reflecting the fact that certain apps entered the data mid-sample, introducing pronounced seasonal heterogeneity in temporal coverage.

Figure B.1 Signal frequency and temporal coverage by app



(a) Time gap distribution

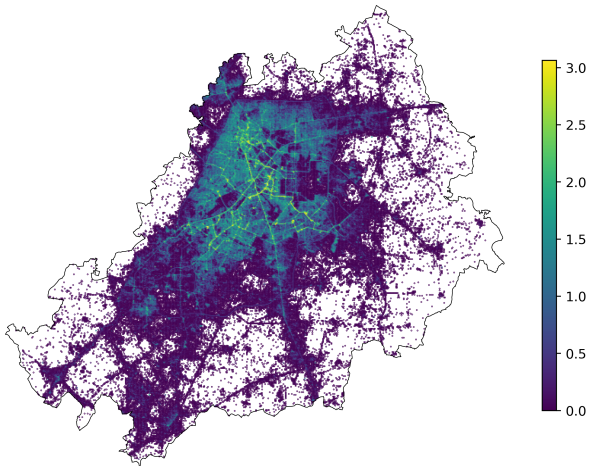
(b) Signal volume over time

Notes: The left panel shows the distribution of the time gap between consecutive GPS signals for a given device, illustrating substantial heterogeneity in signal frequency. The right panel shows the total volume of GPS signals per day over the sample period, illustrating pronounced variation in coverage across the year.

Sample construction. We construct the working sample for Lahore through three sequential steps. First, we restrict to signals recorded within Lahore district and drop observations with a reported horizontal accuracy exceeding 1,000 meters. Second, we retain only devices with at least 50 observations over the sample period, ensuring sufficiently rich location histories for home

and work inference. Third, we remove erroneous observations where a device appears to travel an implausibly large distance in a short period of time; following Verma et al. (2024), we exclude signals with a segment speed exceeding 50 m/s or an acceleration outside the range $[-10, 10]$ m/s^2, computed from consecutive signal pairs for each device. The resulting cleaned sample comprises 198,143,576 observations from 387,835 unique devices. Figure B.2 shows the spatial distribution of signals across Lahore.

Figure B.2 Spatial distribution of cellphone GPS signals in Lahore



Notes: The map shows the spatial distribution of cellphone GPS signals at a 0.001°×0.001° resolution. Color indicates average number of unique devices recorded per day in log, based on the cleaned sample of approximately 198 million GPS signals from 387,835 devices within Lahore.

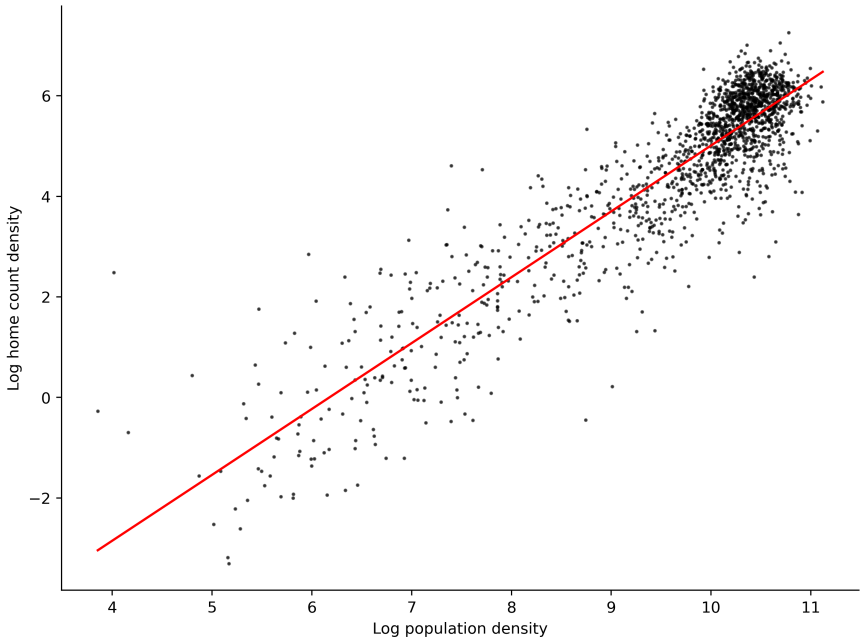
B.1.2 Home and work location inference

Following Kreindler and Miyauchi (2023), each device is assigned a single primary home location and one or more work locations based on the temporal pattern of its signals. Home signals are defined as those recorded between 9 pm and 5 am; work signals are those recorded between 10 am and 3 pm on weekdays. A device's home (work) location is the most frequently visited geocoordinate during the respective time window over the full sample period. This device-level assignment yields 116,336 device-home-work location pair observations for the Lahore sample, of which 45% have distinct home and work locations. Devices with multiple work locations are accommodated by weighting each unique home-work pair by its share of all home-work observations at the device-day level originating from the device's primary home; these weighted pairs are then aggregated into a location-level origin-destination commuting matrix.

Validation: residential population. We assess the plausibility of the inferred home locations by comparing their spatial distribution with residential population data at the location level. For each location, we compute home-location density as the number of device-level home assignments

divided by the area of the location in km^2, and compare this with population density from Schiavina, Freire and MacManus (2023). Home-location density and population density follow a log-linear relationship with slope 1.31 and $R^2 = 0.82$ (Figure B.3), indicating that the GPS-derived home locations track the spatial structure of the population closely, though with some over-representation of more densely populated areas.

Figure B.3 Population density and home density



Notes: Each dot represents a location (Voronoi cell). The fitted line corresponds to an OLS regression of log home density on log population density (per km^2). Population density is from the GHS population layer (Schiavina, Freire and MacManus, 2023); home density is constructed from device-level home locations inferred from cellphone GPS signals.

Validation: gravity regression. We further validate the commuting matrix by estimating a standard gravity equation for bilateral commuting flows, aggregated to the location level ($N = 1655$, yielding up to 2,739,025 origin-destination pairs). The estimating equation is

$$\ln C_{ij} = \mu_i + \psi_j - \beta \ln d_{ij}, \tag{27}$$

where C_{ij} is the bilateral commuting flow between residence location i and workplace location j, μ_i and ψ_j are origin and destination fixed effects, and d_{ij} is bilateral distance between the centroids of locations i and j. Equation 27 is estimated by Poisson pseudo-maximum likelihood (PPML), excluding self-pairs and locations with no identified home or work locations. Table B.1 reports results using both geodesic distance and shortest-path road distance along Lahore's major road network. Both specifications yield negative, precisely estimated distance coefficients, consistent with strong spatial decay in commuting flows. we also estimate a non-parametric version of Equa-

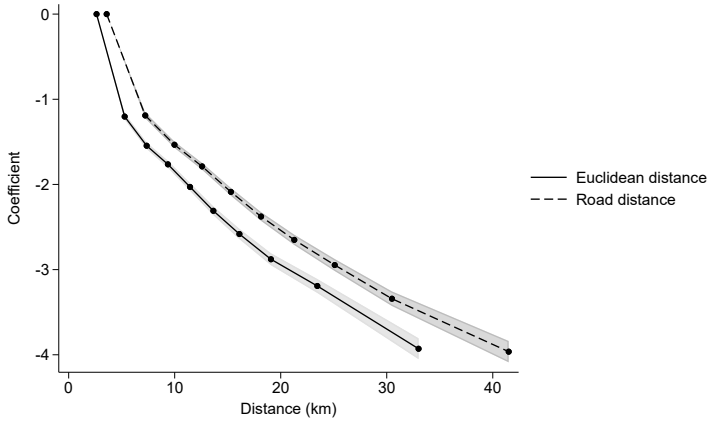
tion 27, replacing $\ln d_{ij}$ with a set of indicators for deciles of log bilateral distance. Figure B.4 plots the resulting coefficients against distance. The decay in commuting flows with distance is smooth and strictly decreasing for both distance measures.

Table B.1 Gravity equation estimates

	Commuter flows	
	(1)	(2)
Log distance (Euclidean, km)	-1.376***	
	(0.00715)	
Log distance (Road network, km)		-1.241***
		(0.00623)
Observations	2,485,390	2,485,390
Pseudo R^2	0.30	0.30

Notes: Standard errors clustered at the origin-destination level in parentheses. Dependent variable is number of cellphone commuters from i to j. All models include origin and destination fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure B.4 Gravity equation: non-parametric estimates



Notes: Each point represents the coefficient on an indicator for a decile of log bilateral distance, estimated from equation 27 with origin and destination fixed effects but replacing $\ln d_{ij}$ with distance decile dummies. Shaded areas indicate 95% confidence intervals. Euclidean and road distances are shown separately.

B.1.3 Commuter traffic

To construct edge-level traffic counts, we transform the cleaned GPS dataset into consecutive signal pairs at the device level and project these onto Lahore's road network.

Signal-pair construction and filtering. Starting from the cleaned Lahore Signal dataset, we transform the data to the consecutive signal-pair level for each device. A signal pair is retained if it satisfies all of the following conditions: (i) both signals are within 10 meters of the nearest road segment; (ii) the device is not stationary, i.e., the two signals are not co-located; (iii) the elapsed time between signals does not exceed 30 minutes; and (iv) the Euclidean distance between the two signals does not exceed 20 km.

Shortest-path routing. For each retained signal pair, each signal is first projected onto the nearest road segment. We then compute the shortest path between the two projected locations using Dijkstra's algorithm weighted by edge length, yielding the road distance and the ordered sequence of edges traversed. For signal pairs whose projected start and end points lie on the same edge, the traversed distance is simply the sub-segment between the two projected points. For pairs on different edges, the start and end edges are temporarily split at the respective projection points and re-inserted into a copy of the graph before the shortest path is computed. For each traversed edge we also record the direction of travel — forward (1) or reverse (0) relative to the edge's geometry orientation — by comparing the travel direction $u \to v$ against the stored start and end coordinates of the edge geometry.

Aggregation. The signal-pair-level route data are aggregated to the edge-direction level by counting the number of distinct devices traversing each directed edge across the sample period. Average daily traffic for each directed edge is then obtained by dividing the total count by 365. Edge-direction observations with no recorded GPS activity are assigned a traffic count equal to the first percentile of the corresponding edge-direction traffic distribution.

B.2 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 by the air quality sensor at the American consulate in Lahore is obtained for most days from 2019 to 2023.

   The location of all air-quality monitors in our sample is shown in Figure B.5. Table B.2 summarizes the number of observations from each source in our data.

Figure B.5 Map of air quality sensor locations

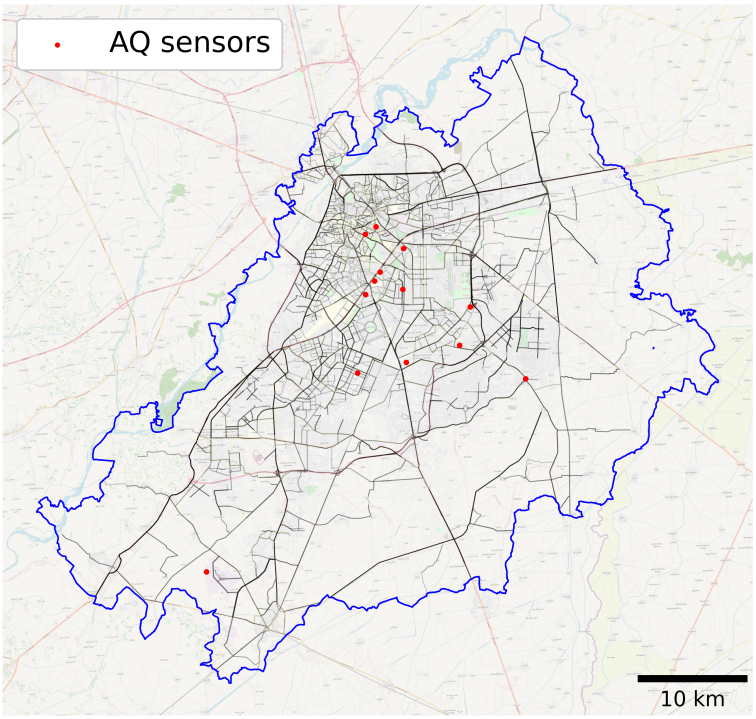


Table B.2 Pollutant detection days by data source

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 (Lahore)	1	0	0	0	237	366	297	346	260
Pakistan Air Quality Initiative	10	3	137	365	7	0	0	0	0 height

Table B.3 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

B.3 Traffic survey data

Data on traffic counts is sourced from the following surveys:

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.[24]

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.6 shows the locations of traffic counts from these sources. Table B.4 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.4.

B.4 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

[24]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.4 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

Notes: 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.5 Vehicle Shares Across Surveys

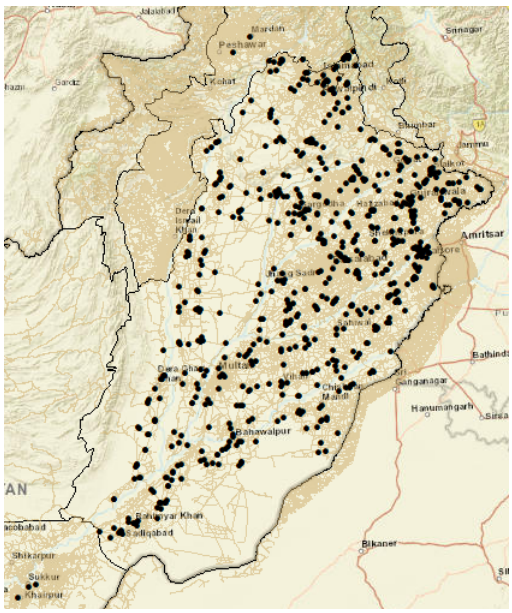


Figure B.6 Map of traffic survey stations across Punjab

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 (meters) near each source at hourly intervals.[25]

The ADMS-Urban model calculates pollutant concentrations from each emission source explicitly included in the model. This allows to perform source-apportionment exercises: the pollution concentration at each location can be decomposed into the contribution of each source in the model and a background concentration. This corresponds exactly to the concentration shares that our model requires to perform the counterfactual simulations, equations (28) to (31) in Appendix C.1.

B.4.1 Model inputs

Pollution sensors location

The ADMS-Urban model computes pollutant concentrations at a predefined set of receptor (sensor) locations supplied as inputs. To accurately represent spatial variation in exposure, particularly in proximity to major roads, we employ a non-regular grid of sensor locations. This approach is designed to capture the steep concentration gradients near roads and to avoid biases from a regular grid where sensors may be placed disproportionately close to roads in some locations relative to others based on the alignment of the grid.

To better characterize near-road gradients, receptors are placed on the roads and at 500 m intervals extending outward from the roads. We place additional receptors at fixed distances of 5 m, and 50 m from the road edge to better capture concentrations near roads. For a given distance from the road we place sensors every 500 meters. Each sensor is assumed to be at a height of 2 meters. This configuration ensures a fine spatial resolution where concentration changes are most pronounced, while maintaining coverage at broader distances.

Cell-level pollution concentrations are computed as weighted averages of the modeled concentrations at each sensor within a cell. The weights are calculated based on the number of GPS cellphone pings across Lahore. Specifically, cellphone pings are first aggregated to a $0.001^\circ \times 0.001^\circ$ grid. Then for each receptor the weight is defined as the log number of cellphone pings at the nearest grid point, normalized by the total log pings within the corresponding cell. For cells which end up with no sensors from this grid, we place supplementary sensors at their centroids. We also add additional sensors at the location of the 13 sensors for which we have data shown in Table B.3. This results in a total of 22506 sensors across Lahore. Figure B.7 shows the spatial distribution of these receptor locations across Lahore.

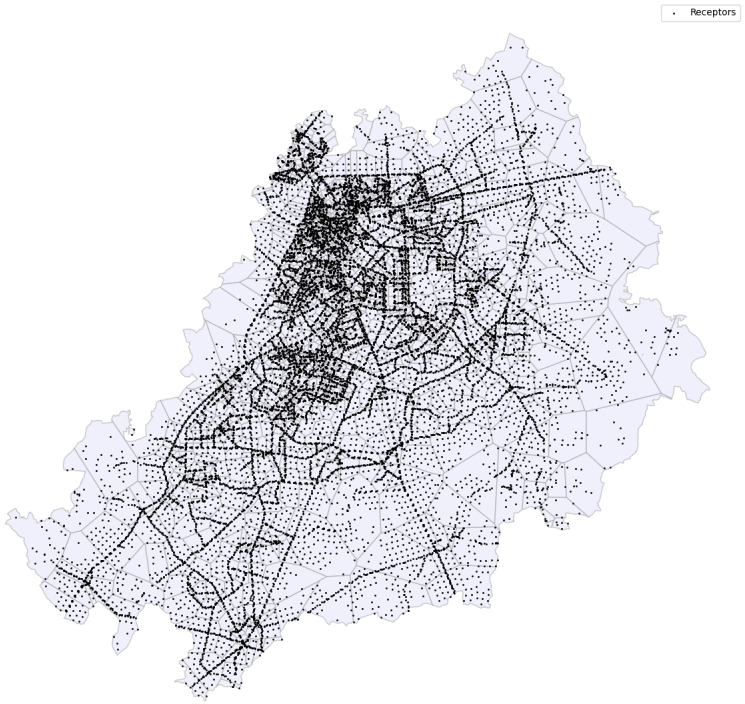
Edge-level pollution concentrations are calculated similarly but only using sensors placed on road edges.

Traffic emissions

Emissions from each road source in the model are captured by an emission rate measured in g/km/s. Emissions in the model are measured in g/s, thus to estimate the marginal concentration from one

[25] 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.

Figure B.7 Locations of PM2.5 concentration receptors included in the model



unit of emissions, i.e. γ_{je}, we set the emission rate for each road edge as one over the length of the edge in km.

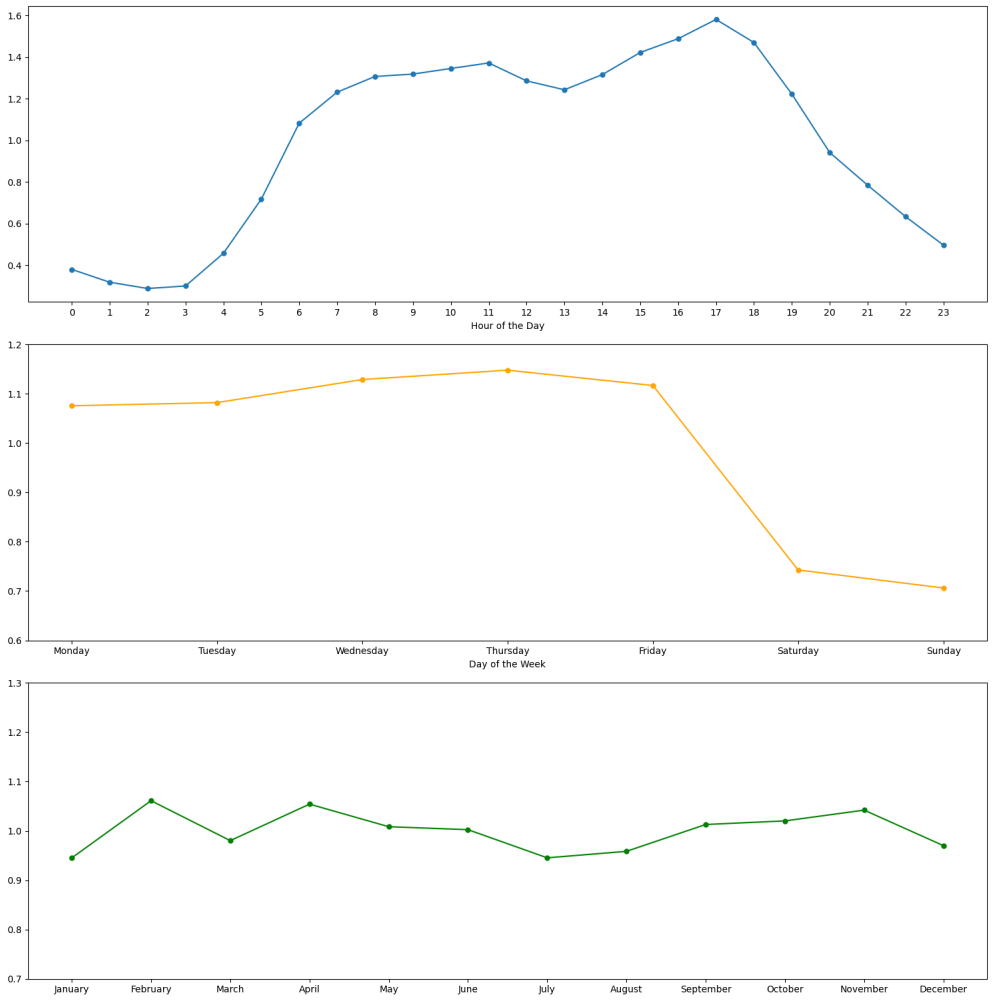
The time varying structure of traffic emissions is then captured by rescaling the traffic rate using fixed factors, which vary based on variation in hour of the day, day of the week and monthly traffic captured in traffic survey data.[26] These time-varying factors, shown in Figure B.8, map AADTs to hourly-level traffic volumes for all road segments. These time-scaling factors multiply the average traffic rate to create more realistic hourly traffic while maintaining daily traffic rates unchanged. Capturing the time variation of traffic emissions is important as PM2.5 concentration are sensitive to changes in meteorological conditions between night and day.

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

[26]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.

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. Each road is assumed to be 6 meters wide.

Figure B.8 Time scaling factors for traffic emissions



Emission factors

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. We consider nine vehicle categories: motorcycles, cars, and buses - which contribute to commuting traffic - 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 - which contribute to goods traffic. Using traffic survey data we calculate the share of each vehicle as a fraction of total commuting/goods AADT, eg. the traffic share for motorcycles would be the AADT for motorocycles divided by the total AADT of commuting vehicles: motorcycles, cars and buses. The share weights are shown in Table A.1.

We use these traffic shares for commuting/goods traffic to obtain average emission factors for one AADT unit of commuting/goods traffic.

Exhaust emission factors were obtained from the COPERT v5.3 database (EMP/EEA, 2019).[27] 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.[28]. 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.6 shows euro standards and fuel characteristics which were used to calculate emission factors for each vehicle type.

Non-exhaust emission factors such as tire, road and brake wear are assumed to be equal to the baseline values in the COPERT database. 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.

We also construct a set of counterfactual emission factors based on the scenario where Euro2 emission standards were perfectly enforced. These are constructed by obtaining the reduction in emission factor for each type of vehicle from the pre-Euro to Euro2 emission standards. The reduction in the emission factor for a representative AADT unit of commuting/goods traffic is then calculated by a weighted average of each type of vehicle within each category based on the traffic shares shown in Table A.1. This results in an average reduction in emission factors by 64% for commuting and 71% for goods traffic.

Figure B.9 shows the emission factors for each type of traffic at different speeds, relative to the emission factor of unit of commuting traffic traveling at 20 km/h. This profile of emissions is then rescaled in the calibration to match the share of population weighted $PM_2.5$ concentration.

Industrial emissions

The model also includes emissions arising from economic activity. This reflects all emissions from the production of output within a given cell. We include this type of emissions in ADMS-Urban as volume sources. Each cell generates homogeneous emissions over the area it covers up to a height of 50 meters. Total emissions, calculated in grams per second, are assumed to be proportional to the output of each cell. Using total emissions we can then calculate the emission rate for each volume source in $g/s/m^3$ - the required input for ADMS-Urban - by dividing total emissions by the volume of each source.

Instead of explicitly modelling the emission rate of one unit of output, we set each cell with total emissions of $1g/s$. Then we rescale the obtained γ_{ij} to match the the relative contribution of

[27]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.

[28]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).

Table B.6 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%

Table B.7 Emission factors comparison

Vehicle	$PM_{2.5}$ emission factor (g/km)		
	Baseline	*No resuspension*	*Euro 2*
Motorcycle	0.03681	0.03411	0.00780
Car	0.03217	0.02537	0.01412
LGV	0.05794	0.05114	0.01634
Bus	0.69748	0.65138	0.20785
RHGV 2 axles	0.47089	0.42479	0.16717
RHGV 3 axles	0.60046	0.55436	0.21316
AHGV 3/4 axles	0.59849	0.55239	0.21246
AHGV 5 axles	0.70625	0.66015	0.25072
AHGV 6 axles	0.76012	0.71402	0.26984

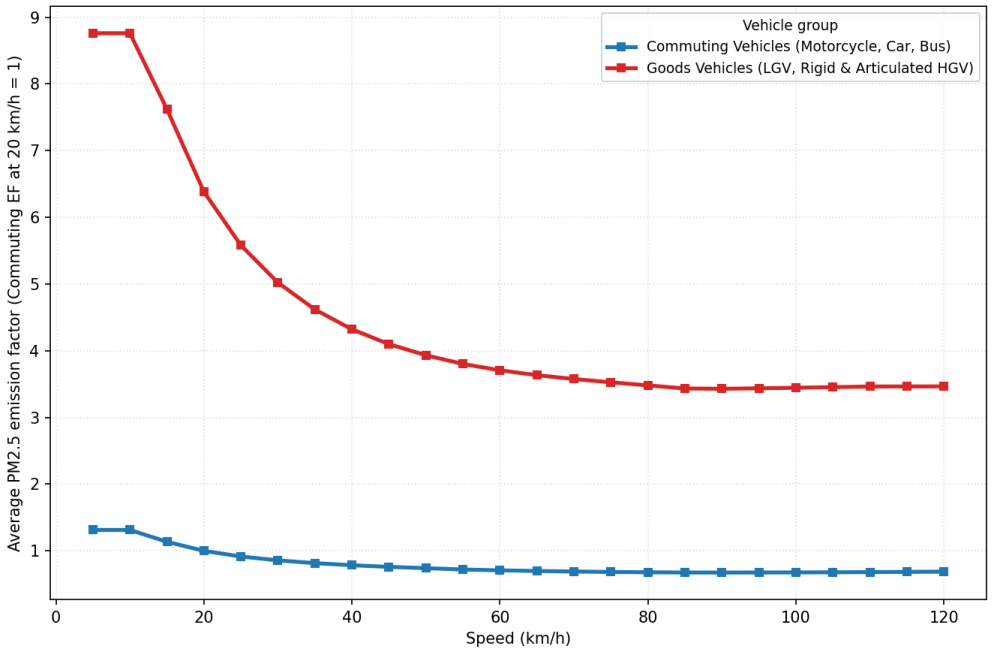
Notes: Emission factors from COPERT v5.3 with Pakistani fleet information. Baseline emission factors include estimates for each vehicle type from Nagpure et al. (2016). Euro 2 emission factors rescale baseline values using the average reduction from pre-Euro to Euro 2 in the COPERT database based on the vehicle fleet characteristics in the Punjab. Emission factors are all shown for a speed of 30km/h.

industrial emissions and traffic emissions to PM2.5 concentrations (see Calibration Appendix).

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. ADMS-Urban takes as an input homogeneous meteorological conditions over the area of interest. Meteorological data for Lahore is obtained by mapping the gridded data to a 9x11 km grid in the UTM 42N projection and taking meteorological data for the point closest to the centre of Lahore.

Figure B.9 Relative emission factors by type of traffic at different speeds



Notes: Emission factors are taken from COPERTv5.3 database adjusted for Pakistani fleet characteristics. Emission factors for each type of traffic are calculated averaging vehicle EFs using traffic share weights. Emission factors decline with speed as fuel consumption become more efficient. Units are rescaled such that the EF from commuting at 20 km/h is equal to 1.

B.5 Calibrating parameters

Calibrating ν. The parameter ν reflects how many more vehicles are required to transport one unit of goods vs commuters, measured in passenger car equivalents (PCE). This captures the relative contribution to traffic congestion of one unit of AADT from goods traffic relative to commuting traffic. We calibrate it by matching the average share of PCEs in the model to the observed share of PCEs from goods traffic observed in the traffic surveys. The PCE share from goods traffic is calculated by multiplying the traffic volume of each vehicle by its PCE weight, then calculating the ratio of total PCEs from goods vehicles over the total PCEs from all vehicles for each road in the traffic surveys. The traffic volumes of each types of vehicles are taken from the traffic surveys (see Data appendix), while the PCEs for each type of vehicle is taken from the Highway Capacity Manual as reported in Farooq and Akram (2018) and shown in A.1. This results in an average PCE share from goods traffic of 58%. The parameter ν is then calibrated to match this to the average PCE share across all edges in the model:

$$\sum_e \frac{\nu \Xi_e^y}{\nu \Xi_e^y + \Xi_e^l} = 0.58$$

This yields a value of $\nu = 16.01$, which implies that on average one vehicle used for goods trade corresponds to around 16 times more PCE units than one for commuting.

Calibrating the γ_{ij}, γ_{ei}: The parameters γ_{ij} denote the determine the strength of the pollution externality across cells. These parameters capture the role of meteorology and geography in the dispersion of pollution in a given location to other neighbouring areas. Each γ_{ji} determines the marginal impact of one unit of output emissions in cell i, measured in g/s, on PM2.5 concentration in cell j, measured in $\mu g/m^3$.

Similarly, the parameters γ_{je} denote the determine the strength of the pollution externality from traffic emissions on edge e to pollution concentrations in cell j. Finally, we also compute $\gamma_{e'e}$ as the impact of emissions on edge e on pollution concentrations on edge e'.

These parameters are computed using ADMS-Urban. We input emission rates to each pollution source which correspond to 1 g/s total emissions, and obtain the γ_{ji}, γ_{je},$\gamma_{e'e}$ as the simulated PM2.5 concentration from each source measured in $\mu g/m^3$.

Calibrating emission factors $\mu_y, \mu_{\Xi_y}, \mu_{\Xi_l}$: The emission factor parameters determine the PM2.5 emissions in g/s of one unit of output y and one unit of goods/commuting traffic Ξ_y, Ξ_l respectively.

We calibrate these parameters to match three moments: 1) the population weighted share of $PM_2.5$ concentrations from traffic, 2) the population weighted share of $PM_2.5$ concentrations from industrial sources, amd 3) the relative emission factors between goods and commuting vehicles.

The first two moments are calibrated to be in line with estimates from World Bank (2025) for greater Lahore. Emissions from transport are assumed to account for a total of 24% of population weighted $PM_{2.5}$ concentrations while industrial emissions 16%. This leads to the following conditions:

$$\begin{aligned} \frac{\sum_{j\in I}\omega_j(\sum_e \gamma_{je}\mu_{\Xi_y}\Xi_e^y + \gamma_{je}\mu_{\Xi_l}\Xi_e^l)}{\sum_{j\in I}\omega_j m_j} &= 0.24 \\ \frac{\sum_{j\in I}\omega_j(\sum_{i\in I}\gamma_{ji}\mu_y Y_i)}{\sum_{j\in I}\omega_j m_j} &= 0.16 \end{aligned}$$

The relative emission factors for each type of traffic and for different traffic speeds are calibrated to match the profile obtained from the COPERTv5.3 database. We do this by redefining emission factors relative to the emission factor of a unit of commuting traffic travelling at 20 km/h, denoted by $\tilde{\mu}_{\Xi_l}$. Then we can rewrite the emission factors as $\mu^e_{\Xi_l} = \chi^l_e\tilde{\mu}_{\Xi_l}$ and $\mu^e_{\Xi_y} = \chi^y_e\tilde{\mu}_{\Xi_l}$ where the relative scaling factors χ^l_e, χ^y_e depend on the traffic speed of the edge and the type of traffic. These scaling factors are taken from the COPERT database and they are shown in Figure B.9. Then we

can simply rewrite the moment conditions as:

$$\begin{aligned}\tilde{\mu}_{\Xi_l} &= 0.24 \times \frac{\sum_{j\in I}\omega_j m_j}{\sum_{j\in I}\omega_j(\sum_e \gamma_{je}\chi_e^y \Xi_e^y + \gamma_{je}\chi_e^l \Xi_e^l)} \\ \mu_y &= 0.16 \times \frac{\sum_{j\in I}\omega_j m_j}{\sum_{j\in I}\omega_j(\sum_{i\in I}\gamma_{ji} Y_i)}\end{aligned}$$

Calibrating background concentration $\bar{m}$: The background concentration $\bar{m}$ reflects $PM_{2.5}$ concentrations from external or unmodelled sources and it is assumed to be homogeneous over the area of interest.

This is calibrated to match the observed median $PM_{2.5}$ concentration at the US consulate in Lahore in 2019 which is 85.1 $\mu g/m^3$. This gives an additional moment to match which also depends on the calibration of the emission factors. Denoting as k the cell containing the US consulate this can be expressed as:

$$m_k = \sum_{i\in I}\gamma_{ki}\mu_y Y_i + \sum_e \gamma_{ke}\mu_{\Xi_y}\Xi_e^y + \gamma_{ke}\mu_{\Xi_l}\Xi_e^l + \bar{m} = 85.1$$

Discussion of alternative calibrations

Bottom-up calibration using emission factors: Using emission factors from the COPERTv5.3 database and traffic share weights, we can directly obtain estimates for the μ_{Ξ_l}, μ_{Ξ_y} emission factors from one representative AADT unit of commuting and goods traffic.

This method relies on both accurate estimates of the characteristics of the Pakistani vehicle fleet, the comparability of COPERT emission factors when applied to Pakistani vehicles, and accurate estimates of traffic volumes in AADT units.

Top-down calibration matching total emissions from traffic: An alternative calibration is also considered where emission factors are calibrated to match the reported total $PM_{2.5}$ emissions from traffic. The Urban Unit Lahore emission inventory in 2023 estimates total $PM_{2.5}$ emission from traffic to amount to 397 tonnes per year.

In the model total emissions are calculated multiplying traffic volumes Ξ_l, Ξ_y by the relevant emission factors μ_{Ξ_y}, μ_{Ξ_l} and summing across all road edges. This gives the following calibration:

$$\sum_e \mu_{\Xi_y}\Xi_e^y + \mu_{\Xi_l}\Xi_e^l = 397\, tonnes/year$$

These calibration strategies rely on accurate measures of $PM_2.5$ emissions which may be hard to obtain. These methods predict a population weighted share of concentration from traffic between 3% to 5%. When comparing the simulated concentrations with reduced-form results, these calibration strategies give unrealistic cross-sectional distribution of concentrations across Lahore and much smaller $PM_{2.5}$ spikes during rush-hours.

Conservative traffic emissions: Emissions included in the model only capture those arising from traffic from major roads, whereas in reality some emissions may be coming from minor and

residential roads. Thus, modelled emissions instead of capturing 24% of concentration as in the baseline calibration, would capture a lower share of total concentration. To estimate the share of unmodelled traffic emissions, we first calculate the share of GPS pings from cellphone and trucks that are on roads which are not included in the model - these are 47% and 9% respectively. Then we calculate the total share of unmodelled emissions by assuming that these reflect the share of emissions on residential and minor roads. Thus, using these factors yields that 76.8% of total emissions are captured in the model, whereas the remaining 24.2% should be included in background concentrations. This implies that traffic in the model should capture 14.7% of concentrations rather than 24%. This gives an alternative calibration strategy with more conservative emission factors for traffic.

C Mathematical Appendix

C.1 Concentration share matrices

The matrices $\mathbf{M}_{ij}^{y}$, $\mathbf{M}_{ie}^{\Xi_y}$, and $\mathbf{M}_{ie}^{\Xi_l}$ 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).

$$\mathbf{M}_{ij}^{y} = \frac{\gamma_{ij}\mu_y Y_j}{\sum_{j\in I}\gamma_{ij}\mu_y Y_j + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_y}\Xi_e^y + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_l}\Xi_e^l + \bar{m}_i} \tag{28}$$

$$\mathbf{M}_{ie}^{\Xi_y} = \frac{\gamma_{ie}\mu_{\Xi_y}\Xi_e^y}{\sum_{j\in I}\gamma_{ij}\mu_y Y_j + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_y}\Xi_e^y + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_l}\Xi_e^l + \bar{m}_i} \tag{29}$$

$$\mathbf{M}_{ie}^{\Xi_l} = \frac{\gamma_{ie}\mu_{\Xi_l}\Xi_e^l}{\sum_{j\in I}\gamma_{ij}\mu_y Y_j + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_y}\Xi_e^y + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_l}\Xi_e^l + \bar{m}_i} \tag{30}$$

The matrix $\bar{\mathbf{M}}_\mathbf{i}$ captures the baseline share of emissions from other external sources, such as domestic emissions and agricultural burning:

$$\bar{\mathbf{M}}_i \equiv \frac{\bar{m}_i}{\sum_{j\in I}\gamma_{ij}\mu_y Y_j + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_y}\Xi_e^y + \sum_{e\in E}\gamma_{ie}\mu_{\Xi_l}\Xi_e^l + \bar{m}_i} \tag{31}$$

Hence it needs to be the case that for every j:

$$\sum_i \mathbf{M}_{ji}^{y} + \sum_e \mathbf{M}_{je}^{\Xi_y} + \sum_e \mathbf{M}_{je}^{\Xi_l} + \bar{\mathbf{M}}_j = 1$$

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

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

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

$$\Xi^l_{ji} = A_{ji},$$

the diagonal elements of the traffic matrix are

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

C.3 Exposure of households to pollution while commuting

In the scenario where households experience utility losses from pollution experienced on their commute, households have utility

$$V_{ji,r} = [m_i^{\sigma_h}] \times \frac{m_j^{\sigma_w} \times w_j}{p_j} \times \left[\prod_{l=1}^{K(l)} t_{r_{l-1},r_l}\right]^{-1} \times \eta_{ji,r}$$

where

$$t_{kl} = \bar{t}_{kl} \times \Xi_{kl}^{\lambda} \times (m_{kl})^{\sigma_{c,kl}}$$

and σ_c has the interpretation

$$\sigma_{c,kl} = \frac{\partial \log t_{kl}}{\partial \log \text{conc}_{kl}}$$

All other derivations are identical to the baseline model, with the caveat that transportation costs now include both congestion effects and utility costs from exposure to pollutants. We then have

$$\begin{aligned}
t^c_{kl} &= \bar{t}_{kl}\left(\nu_t \Xi^y_{kl} + \Xi^l_{kl}\right)^{\lambda} (m_{kl})^{\sigma_{c,kl}} \\
&= \bar{t}_{kl} t_{kl}^{-\lambda\theta}\left(\nu \times p_l^{-\theta}\Pi_k^{-\theta} + \left(\Pi^l_k\right)^{-\theta}(\Theta_l)^{-\theta}\right)^{\lambda} (m_{kl})^{\sigma_{c,kl}} \\
t^c_{kl} &= \bar{t}_{kl}^{\frac{1}{1+\lambda\theta}}\left(\nu \times p_l^{-\theta}\Pi_k^{-\theta} + \left(\Pi^l_k\right)^{-\theta}(\Theta_l)^{-\theta}\right)^{\frac{\lambda}{1+\lambda\theta}} (m_{kl})^{\frac{\sigma_{c,kl}}{1+\lambda\theta}} \\
(t^c_{kl})^{-\theta} &= \bar{t}_{kl}^{-\frac{\theta}{1+\lambda\theta}}\left(\nu \times p_l^{-\theta}\Pi_k^{-\theta} + \left(\Pi^l_k\right)^{-\theta}(\Theta_l)^{-\theta}\right)^{-\frac{\lambda\theta}{1+\lambda\theta}} (m_{kl})^{-\frac{\theta\sigma_{c,kl}}{1+\lambda\theta}}
\end{aligned}$$

and, expressed in changes,

$$(\hat{t}^c_{kl})^{-\theta} = \hat{\bar{t}}_{kl}^{-\frac{\theta}{1+\lambda\theta}}\left(\frac{\nu_t \Xi^y_{kl}}{\nu_t \Xi^y_{kl} + \Xi^l_{kl}}\hat{p}_l^{-\theta}\hat{\Pi}_k^{-\theta} + \left(\frac{\Xi^l_{kl}}{\nu_t \Xi^y_{kl} + \Xi^l_{kl}}\right)\left(\hat{\Pi}^l_k\right)^{-\theta}\left(\hat{\Theta}_l\right)^{-\theta}\right)^{-\frac{\lambda\theta}{1+\lambda\theta}} (\hat{m}_{kl})^{-\frac{\theta\sigma_{c,kl}}{1+\lambda\theta}}$$

Exposure to pollution during the commute have pure utility costs only. As a result, $\sigma_{c,kl}$ is calibrated in relation to σ_h: if a household spends $\mathfrak{t}_{kl}$ units of time crossing an edge, and η_h units of time at home, and δ is the relative exposure to pollution while commuting compared to while being at

home, then

$$\sigma_{c,kl} = \delta \frac{\mathsf{t}_{kl}}{\eta_h} \sigma_h.$$